DOCUMENT RESUME

ED 285 345 EC 200 282

AUTHOR Shaver, James P.; And Others
TITLE The Modification of Attitudes Toward Persons with Handicaps: A Comprehensive Integrative Review of Research. Final Report.
INSTITUTION Utah State Univ., Logan. Bureau of Educational Research.
SPONS AGENCY Office of Special Education and Rehabilitative Services (ED), Washington, DC.
PUB DATE 28 Feb 87
GRANT G008530210
NOTE 585p.; Research in Education of the Handicapped Program.
PUB TYPE Information Analyses (070) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF03/PC24 Plus Postage.
DESCRIPTORS *Attitude Change; Attitude Measures; Desensitization; *Disabilities; Effect Size; *Exceptional Child Research; Literature Reviews; Meta Analysis; Positive Reinforcement; Reliability; *Research Design; Research Methodology; *Research Reports; Validity

ABSTRACT

This final report analyzes results of a comprehensive, integrative review of 273 primary research studies on modifying attitudes toward disabled persons. Attitude modification approaches identified included providing information, direct contact with disabled persons, vicarious experiences related to having a disability, systematic desensitization, positive reinforcement, and combinations of these approaches. The primary analytic focus was on mean effect sizes: 644 effect sizes, obtained from 200 reports, constituted the main data set. Outcomes were not related to three basic indicators of study quality (general treatment validity, general internal validity, and adequacy of test validity). Methodological quality was not high, treatment effects were moderate and heterogeneous, and treatment techniques were heterogeneous in characteristics and outcomes. The studies differed on such characteristics as type of comparison, time of posttest, and type of dependent measure. Alternative and supplementary data explorations did not produce results markedly different from the analyses of individual effect sizes. Among the areas of needed improvement in research methodology are definition of the construct of "attitude toward persons with disabilities," selection of assessment techniques that yield reliable and valid scores, attention to treatment validity (including more careful reporting of procedures and design elements), and the development of programmatic, replication-oriented research. (Author/JW)

FINAL REPORT

PROJECT NO. 023CH50160

GRANT NO. G008530210

THE MODIFICATION OF ATTITUDES TOWARD PERSONS WITH HANDICAPS: A COMPREHENSIVE INTEGRATIVE REVIEW OF RESEARCH

JAMES P. SHAVER

CHARLES K. CURTIS

JOSEPH JESUNATHADAS

CAROL J. STRONG

SUBMITTED TO

U. S. DEPARTMENT OF EDUCATION

OFFICE OF SPECIAL EDUCATION AND REHABILITATIVE SERVICES

RESEARCH IN EDUCATION OF THE HANDICAPPED PROGRAM

FINAL REPORT
Project No. 023CH50160
Grant No. G008530210

THE MODIFICATION OF ATTITUDES TOWARD PERSONS WITH HANDICAPS: A COMPREHENSIVE INTEGRATIVE REVIEW OF RESEARCH

James P. Shaver   Charles K. Curtis

Joseph Jesunathadas   Carol J. Strong

Bureau of Research Services
Utah State University
Logan, Utah 84322-2800

February 28, 1987

Submitted to
U. S. DEPARTMENT OF EDUCATION
OFFICE OF SPECIAL EDUCATION AND REHABILITATIVE SERVICES
RESEARCH IN EDUCATION OF THE HANDICAPPED PROGRAM

The findings, conclusions, opinions, and recommendations in this report are those of the authors and do not necessarily reflect the views of the funding agency.

EXECUTIVE SUMMARY

The Modification of Attitudes Toward Persons With Handicaps: A Comprehensive Integrative Review of Research

James P. Shaver, Charles K. Curtis,
Joseph Jesunathadas, and Carol J. Strong

Changing Negative Attitudes Has Been A Research Interest

Legislation and judicial decisions are bringing handicapped persons into the mainstream of educational, social, and economic life in this society. Nevertheless, negative attitudes toward persons with disabilities continue to be detrimental to their potential to live dignified, productive lives and to contribute to society. A major research interest has been how to modify the negative attitudes and thereby mitigate the effects on persons with disabilities.

Prior Reviews Have Not Been Comprehensive or Quantitative

Despite the availability of much larger numbers of reports of research on modifying attitudes toward disabled persons, prior reviews of that research have typically cited only 32 to 36 studies. Moreover, an analysis of seven full reviews and eight brief reviews indicated that they suffered from a number of shortcomings. For example, methods of locating reports and criteria for including research reports in the review were not reported; there was a lack of systematic data collection and analysis as a basis for conclusions about the effectiveness of methods of attitude change; and, in almost all of the reviews, the moderating effects of other variables--variations in treatments, other study characteristics, and sample characteristics--were not addressed.

As a result of the lack of comprehensiveness and the narrative approach of the prior reviews, it could not be discerned whether the conclusions generally drawn--i.e., that the effects of attitude modification techniques were negligible and that both negative and positive results were often produced--reflected inadequacies in the review procedures or were accurate depictions of the research literature.

This Review Was Comprehensive and Quantitative

This integrative review was based on all reports of relevant research that could be located through computer and hand searches of commonly used indexes, and in the reference lists for prior reviews and primary research reports. A coding instrument was used to collect data on sample characteristics, the attitude modification techniques investigated and their characteristics, design and instrumentation, including indicators of methodological quality, and effect sizes. The basic effect size was Delta (symbolized as $\underline{D}$): the mean of the experimental group minus the mean of the control group, divided by the control group standard deviation ($\underline{D} = \bar{X}_E - \bar{X}_C/SD_C$). Correlation coefficients were also used, and variance ratios--the posttest variance of the experimental group divided by the posttest variance of the comparison group--were computed for exploratory analyses.

644 Effect Sizes from 200 Studies Were the Main Data Set

The number of primary research studies coded for the integrative review was 273 (based on 303 reports, some of which were about the same research). Two hundred studies involving treatment versus control, treatment versus placebo, or single-group, pre-posttest comparisons for which data were available to compute effect sizes made up the population for the main analyses of data for the review. Supplementary analyses were conducted on effect sizes from 15 studies in which alternative treatments, rather than treatment versus the absence of treatment, were compared, and from nine reports of the effects of mainstreaming students with disabilities in regular classrooms. In addition, the results from 55 studies for which statistical significance, but not the information to compute effect sizes, was reported, were also coded for a supplementary analysis to determine whether that population of studies differed from our main population of 200 studies.

All told, 644 effect sizes, obtained from 200 reports, constituted our main data set. In addition, 61 effect sizes from Treatment A vs. Treatment B comparisons, 18 effect sizes from mainstreaming studies, and 182 results for which effect size information was missing were obtained for analysis.

Inferential Statistics Were Not Used

Because the search for reports was comprehensive, including efforts to obtain information for computing effect sizes when it was not available in the report, the studies upon which data were collected were considered a population rather than a sample. Consequently, inferential statistics were not used in analyses.

The primary analytic focus was on mean effect sizes. Standard deviations were also reported, along with Eta^2s as appropriate to indicate the proportion of the variance in Ds associated with the particular categories included in the analyses. Data were organized in one-way and two-way tables, and differences were scrutinized using as the criterion for triviality any difference between mean Ds that was less than .12. Pearson product-moment correlations were also run on some of the data.

Outcomes Were Not Related to Study Quality

Initial analyses indicated that outcomes were not related to three basic indicators of study quality--general treatment validity, general internal validity, and adequacy of test validity. That lack of relationship was at least in part because few studies were rated high (on a three-point scale) on any of the indicators of quality.

Analyses also indicated that, as has been the case in other quantitative integrative reviews, the mean _D_ for journal articles was higher than that for other types of research reports. By the same token, the mean _D_ for single-group, pre-posttest comparisons was higher than that for treatment versus control and treatment versus placebo comparisons. However, there was no indication that either type of report or type of comparison was unevenly distributed among the various attitude modification techniques which had been investigated.

Methodological Quality Was Not High

The predominantly low to moderate ratings of quality reflected a variety of methodological weaknesses in the studies, including failures to verify the implementation of the independent variable, low ratings on actual treatment implementation, low ratings on various indicators of treatment and internal validity, and a lack of information in many cases to determine whether or not there were threats to treatment or internal validity. There also was a lack of replications. Perhaps the most important conclusion drawn from the integrative review was that the methodological quality of research in this field is not high, and future efforts must address both the more adequate design of individual studies and the need for programs of research which include the replication of studies to determine the reliability and generalizable of results.

Despite the inability to determine whether results would have been different for high quality studies, the integrative review was car .ed out, based on the recognition that moderate and low quality studies constituted the best evidence available, but that interpretations must be made cautiously.

Treatment Effects Were Moderate and Heterogeneous

Ten attitude modification approaches were identified: providing information about disabilities and persons with disabilities; direct contact with persons with disabilities; providing situations for vicarious experiences related to having a disability; the use of systematic desensitization to extinguish negative attitudes; the use of positive reinforcement to modify attitudes; information plus direct contact; information plus vicarious experiences; and a category of "other" which encompassed other combinations of the first eight techniques. The tenth category was for studies that contrasted different types of persuasive messages. For two of the ten techniques--positive reinforcement and persuasive message, contrast--so few effect sizes were available that they were largely excluded from consideration.

The mean D for the 644 comparisons in our main data set was .37, a moderate level of effect. The treatment techniques were ranked by size of mean D in the following order: Persuasive Messages, mean D = .67 (N = 23); Information Plus Contact, mean D = .51 (N = 100); Direct Contact, mean D = .43 (N = 93); Vicarious Experiences, mean D = .40 (N = 58); Other, mean D = .39 (N = 71); Systematic Desensitization, mean D = .32 (N = 21), Information mean D = .29 (N = 203); and, Information Plus Vicarious Experiences, mean D = .20 (N = 62). Considerable variability in outcomes was indicated by the overall standard deviation for Ds of .61, along with standard deviations for the eight attitude modification techniques which ranged from .36 to .76, with six of them .50 or above.

Clearly, although it was possible to rank order the techniques by their mean Ds, there was much heterogeneity, with the distribution of Ds for the various techniques overlapping one another. Moreover, for each technique, there were negative Ds (the treatment group gain was less than that for the comparison group). And, there were instances in which the change for the treatment group was negative (even though the D may have been positive because the comparison group had a more negative change).

There Were Also Variations in Treatment Features

Variations occurred in such treatment features as the disabilities toward which attitude modification efforts were directed; the types of information and information delivery modes used, whether information gain was assessed and, if so, the degree of gain, in investigations of information as an attitude modification technique; the types of experiences provided in vicarious experience studies; the types of persuasive messages and how they were presented; and, the types of contact and the situations in which it occurred, the characteristics and relative status of the disabled persons, the presence of institutional support, and the competencies of the disabled persons, in contact studies. However, coding to determine the extent to which contact studies reflected factors considered theoretically important for the modification of attitudes was not fruitful because many reports contained inadequate information for scoring and, when information was available, effect sizes were often clustered in only one or two categories. In short, the various treatment techniques were heterogeneous in terms of features, as well as outcomes.

Studies Also Differed in Other Ways

Other characteristics on which the studies differed included type of comparison (i.e., whether treatment versus control, treatment versus placebo, or single-group, pre-posttest), time of posttest, type of dependent measure, length of treatment, the context for the treatment (e.g., whether an elementary-secondary school or college-university environment), setting (the specific place where the research was conducted--e.g, classroom or laboratory), and sample size. Occasionally, there were different relationships between Ds and concommitant variables across treatments; but treatments were sometimes nested within study characteristics and Ds were lacking for some treatment-characteristic combinations, so drawing conclusions about relative effects was difficult.

There were also differences in samples, such as the methods of selection and the grade-age levels and gender of the Ss. Not only did the studies vary on these sample characteristics, but there were some differential effects across treatments. However, with the nesting of treatments and small Ns or empty cells, relations to treatment outcomes were not clear. Very few reports contained any information on interactions between Ss' prior contact with persons with disabilities or Ss' personality traits and treatment outcomes.

Alternatives to the Analysis of Individual Ds Were Explored and Supplementary Analyses Carried Out

Alternative and supplementary data explorations included analyses based on a median effect size for each study rather than individual effect sizes, and analyses with outliers (Ds greater than two standard deviations from the mean) excluded. Neither of these analyses produced results markedly different from the analyses of individual effect sizes.

In addition, Ds from Treatment A vs. Treatment B comparisons (i.e., direct comparisons of different treatment techniques) and Mainstreaming studies were analyzed separately. However, there were so few effect sizes in either category that little information was gained.

Also, studies for which the statistical significance of results was reported, but information to compute effect sizes was not available, were coded and frequency data analyzed. The Missing Information and effect size studies constituted different populations. For example, the Missing Information studies were likely to be of somewhat lower quality and to have been conducted in different contexts (e.g., a higher percentage of inservice-education and work contexts) than were those for which Ds could be computed. Different attitude modification techniques also were represented in the Missing Information studies.

Differences in the posttest variances of treatment and comparison groups were also explored. Overall, treatment groups were slightly more variable at posttest than were the control groups (mean variance ratio = 1.13). It was postulated that it would be desirable for treatment groups to have higher posttest means (more positive attitudes) and lower posttest variabilities than their comparison groups. However, the relationships between posttest Ds and variance ratios for different treatments were not consistent. Treatment effects on variability appear to be a candidate for further attention in primary research studies and in integrative reviews.

Conclusions Were Drawn About Outcomes, Research Quality, and Attitude Change Research

This comprehensive integrative review confirmed the conclusions in prior reviews that the effects of attitude modification techniques are often not large and frequently are contradictory.

It was also concluded that the general quality of research in the modification of attitudes toward persons with disabilities has not been high. Among the areas of needed improvement in methodology are definition of the construct of "attitude toward persons with disabilities", the selection of assessment techniques that yield reliable and valid scores, attention to treatment validity, including verification of the independent variable, the reduction of threats to internal validity, more careful reporting of procedures and design elements, and, most important of all, the development of programmatic, replication-oriented research.

It was noted that designing and conducting studies with perfect internal validity is extremely difficult in an applied field such as attitude modification, and one major flaw will invalidate an otherwise valid study. Moreover, complex interactions between variables may account for the moderate and contradictory results from primary research studies. How to modify attitudes toward persons with disabilities should be a continuing item on the research agenda. However, how to channel the behavior of individuals who have negative attitudes so as to avoid the restrictive and dehumanizing effects on persons with disabilities should continue to be both a policy and research focus. We ought not simply assume that research will someday tell us how to obliterate negative attitudes.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

CHAPTER 1

WHY ANOTHER REVIEW OF THE LITERATURE

In the last two decades, much progress has been made toward assuring equity in education, housing, and employment for handicapped persons. The Architectual Barriers Act of 1968, Section 504 of the Rehabilitation Act of 1973, Public Law 94-142, and recent court decisions supporting the rights of disabled persons have signaled a marked shift in public policy in regard to handicapped populations. Among other changes that have resulted, increased mainstreaming in public schools and reductions of physical barriers have been major steps toward bringing handicapped persons into full participation in the educational, social, and economic life of the society. Yet, despite these auspicious beginnings, a major barrier to equal opportunities for handicapped persons remains--the attitudes of nondisabled persons toward them.

Bogdan and Biklen (1977) used the term "handicapism" to emphasize the similarity in the situations of handicapped persons, ethnic minorities, and women. All may suffer from stereotyping and prejudice. As with racism and sexism, handicapism refers to sets of assumptions and practices which reflect attitudes and lead to "differential and unequal treatment", based on "apparent or assumed physical, mental, or behavioral differences" (p. 14). Writings by Bowe (1978, 1980), Cohen (1977), and Kleinfield (1979) have illustrated well how unfounded, negative attitudes of other persons often limit the opportunities of disabled individuals, thus handicapping them.

Examples include educators who underestimate the potential of disabled* students, employers who lack confidence in the abilities and motivation of disabled potential employees, and nondisabled persons who hesitate to socialize with those who are disabled. It is of equal, if not greater, consequence that the result can be loss of self-esteem and the self-limiting of options by disabled persons, as they adopt the attitudes of others toward themselves.

Changing attitudes toward disabled persons to eradicate the effects of handicapism is not simply a matter for direct legislative or judicial action. As Itzhak Perlman, the world renowned violinist whose access to some concert halls is impeded because of his physical disability, expressed it:

> There is a real problem in society. Certainly I would like to see public transportation and architecture that everyone can use without difficulty. But what I most want funds can't buy, and that's a change in attitudes toward the disabled so that laws don't have to be enforced. (Salt Lake Tribune, 10/3/84, p. 2A).

Laws that bring handicapped persons into the mainstream of life, as PL 94-142 does in schools, may in the long run have favorable effects on attitudes. But how to modify stereotypic attitudes toward those who are disabled is a significant educational question (Shaver & Curtis, 1981a, b) for several reasons. First, a large number of people are affected directly--some 35 to 50 million persons with potentially handicapping mental and physical disabilities (Bowe, 1978, p. 17; Kleinfield, 1979, p. 32). Moreover, the

---

*The terms "disabled" and "handicapped" are not used consistently in society. Sometimes they are used interchangeably, sometimes to denote different concepts. The authors prefer to use "disabled" to refer to persons who have physical or mental impairments, with "handicapped" used to refer to those persons for whom disabilities limit their ability to function (Shaver & Curtis, 1981a, pp. 1-2). The point relevant to this report is that the attitudes of other persons towards one's disability may make it a handicap, as well as increase the extent to which it is a handicap. This distinction in usage is hard to maintain, and we have not been able to do so throughout the report. In keeping with general practice, we have used "handicapped" as the more "generic" term.

consequences of handicapism are of grave concern in a society, such as ours, in which human worth and dignity and the development of individual potential are basic values. In addition, although the deleterious effects of handicapism on the handicapped are well known, we often overlook the self-degradation of those who act on prejudice to deny others their rights (Robert Coles, Introduction to Kleinfield, 1979). Finally, the costs to society are high, in terms of both productivity and public expenditures for social services, when handicapped persons are undereducated, underemployed, and underpaid.

The Problem

In light of the significant potential effects of attitudes toward persons with disabilities, it is not surprising that a considerable amount of research has been done to investigate ways to modify those attitudes. Yet, as Jones (1984, p. vii) noted recently in an introduction to a review volume, "this varied and rich [research] literature . . . has not yet been synthesized for special education consumers and researchers," and "state-of-the-art overviews" and "critical reviews" of the research on attitudes toward handicapped persons are needed.

It is not that there have been no reviews of the research on modifying attitudes toward persons with disabilities. Prior to beginning the study reported in following chapters, we identified three articles (Anthony, 1972; Donaldson, 1980; Sandler & Robinson, 1981) and a chapter (Towner, 1984) that were focused on that literature. In addition, ten review articles* were

*Alexander and Strain (1978), Chubon (1982), English (1971), Frith and Mitchell (1981), Harth (1973), Horne (1979), Levitt and Cohen (1976), Mitchell (1976), Rabkin (1972), Segal (1978).

located that have brief sections on the topic. However, all of these reviews were found to suffer from the common weaknesses in integrative reviews which Jackson (1978, 1980) cited following his analysis of a random sample of review articles.

Although reviewing and synthesizing research findings has been a common activity in social science research, Jackson (1978, 1980) found an absence of well-defined procedures for conducting integrative reviews. Included in his list of "important weaknesses in the currently prevailing methods of integrative reviews" (1978, p. 37) were: (1) the lack of thorough, systematic searches of the literature, (2) a tendency not to represent study findings as powerfully as possible, (3) the failure to consider systematically the possible relationships between study characteristics and outcomes, and (4) the lack of adequate reporting of review methods. Based on his analysis, Jackson (1978, Ch. 6) proposed seven elements to be considered in judging the quality of reviews:

1. Topic Selection--Was the topic clearly defined and delimited?

2. Review of Previous Work--Were previous efforts to review similar bodies of literature cited and critiqued so that (a) it is clear how the present work will differ from or extend previous work, (b) an appropriate point of departure for the present work can be determined, and (c) the present work will avoid the mistakes of past reviews.

3. Selection of Studies to be Reviewed--Were the criteria for selecting studies to be reviewed clearly explicated? Was a representative or comprehensive sample of previous research on the topic reviewed, so that results of the review are generalizable to the "population" of research studies?

4. Data Collection--Were data collection procedures specifically described and defended on rational and empirical grounds? Were data collected for each study on the magnitude and direction of outcomes, the characteristics of dependent variables (study outcomes), the intervention, and other study or subject characteristics, such as age of students, implementation of intervention, methodological quality?

5. Data Analysis--Were the relationships between dependent and independent variables, including treatment and concomitant variables, examined? Were appropriate analysis techniques utilized?

6. Interpretation--Were conclusions carefully based on the data, including evidence of confounding variables? Were implications of interest to relevant audiences (e.g., practitioners, researchers, policy makers) drawn?

7. Reporting--Were results reported in such a way that the reader can tell exac.ly what procedures and operational definitions were used? Could the investigation be replicated based on the information reported? Were conclusions and recommendations clearly stated?

None of the four full reviews on modifying attitudes toward the disabled that we identified stacked up well against Jackson's criteria, nor did the brief sect.ons in broader reviews. These reviews, and others located after the study began, will be discussed in considerable detail in Chapter 2. The purpose here is to set the context for the decision to conduct another review. For example, in none of the four reviews were previous reviews critiqued and a statement made as to how the review being reported differed from, extended, or benefited from prior reviews. In fact, in only one (Towner, 1984) was a prior review acknowledged. Moreover, none of the four reviews specified how its sample of primary research reports was selected for review. That some selection was likely, especially for the reviews reported since 1972, seemed obvious. The slight increases and one decrease in numbers of reports cited (Anthony, 1972--N=31; Donaldson, 1980--N=22; Sandler & Ro.inson, 1981--N=40; and Towner, 1984--N=47) did not reflect the expected year-by-year accretion in publications. Moreover, considerable selection was also suggested by a comparison of those numbers against the nearly 200 research reports which we had already identified as potentially relevant for review.

Data collection, analysis, and reporting procedures were also identified as major shortcomings in the four reviews. In fact, the type of data collection was not evident in any of the reviews; the results of primary research projects were described verbally in terms of differences (e.g., "Soloway found that more favorable attitudes toward disabled children were demonstrated by teachers without integration experience . . . . Negative changes in attitude toward and optimism concerning the integration of exceptional learners were found by Fenton (1975) and Shotel et al. (1974)" (Towner, 1984, p. 236). As the example implies, the magnitude of differences was not reported, nor were summary statistics (e.g., means and standard deviations), statistical analyses, or the probabilities of results presented. The basic reporting style was narrative, with little effort to summarize findings systematically. The exception was Towner (1984), who provided summary tables, with studies identified as "successful" if statistically significant results were obtained. With the exception of Towner (1984), the methodologies of primary studies were not critiqued. However, Towner discussed the covariation of outcomes with methodological soundness in only general terms; the research outcomes were not expressed in quantitative terms that would allow a systematic analysis and description of methodology-result relationships. The statistical significance of results was referred to in Towner's review, but without recognition that statistical significance is relative to sample size and so does not provide a measure of research results that is comparable across studies (Shaver, 1980).

Given the shortcomings of these reviews of research on modifying attitudes toward the handicapped, it is not surprising that even the author of the most recent, thorough, and systematic review (Towner, 1984) concluded:

> The applications [of similar techniques] yielded discouraging and contradictory findings. Both positive and negative attitudinal changes, in addition to numerous reports of [statistically] nonsignificant changes, resulted from interactions [of nondisabled persons] with disabled persons as well as from the provision of educational and general information. (p. 249)

The lack of a comprehensive, quantitatively-based, integrative review of the research on modifying attitudes toward the disabled was the problem addressed by the research project presented on the following pages. The purpose of the project was to develop a clearer portrayal of the status of research in this area, useful to educational practitioners, policy makers, and researchers.

CHAPTER 2

AN ANALYSIS OF PRIOR REVIEWS

Primary research studies are to be designed and conducted within the context of a critical review of prior research, according to a commonly accepted canon of scholarship. It is a rare research report that does not include a review of prior research. As scholarly contributions to knowledge, reviews of the literature should also be based on prior efforts. According to Jackson (1978):

> Previously completed reviews of the topic or similar topics ought to be consulted to assess what is already known on the topic, to refine questions or hypotheses for the forthcoming review, to anticipate problems that may be encountered when doing the review, to gain familiarity with alternative ways of doing the review and to acquire ideas for interpreting the results of the forthcoming review. Seldom is a review topic so unique that the reviewer cannot benefit from examining previous reviews. (p. 53)

Yet, as White, Bush, and Casto (1985-86) have pointed out, "reviewing and reporting the work of previous integrative reviewers on similar topics is seldom done in reports of 'literature reviews'" (pp. 417-18).

Of course, like any evidence, the results of previous reviews of the literature should not be accepted and used uncritically. The critical examination of procedures and conclusions is essential if the prior work is to serve as a valid foundation for new efforts. Nevertheless, Jackson (1980, p. 443) found that although 27 of the 36 reviews in his sample cited prior reviews, only two provided critiques of the prior work. White, Bush, and Casto (1985-86) found that ten of 52 reviewers cited at least three prior reviews, but none provided a critical analysis of the prior work. In this chapter, we present a critique of the prior reviews of research on modifying attitudes toward disabled persons.

The Reviews

Our search of the research literature indicated that during the past two decades, over 200 investigations of methods for enhancing attitudes toward disabled persons had been reported. Despite the amount of research, there have been few published efforts to summarize and synthesize the findings. A computer-assisted search of ERIC, CEC Abstracts, Dissertation Abstracts, Index Medicus, Psychological Abstracts, and Social Science Research using broad descriptors, along with a manual search of bibliographies and reference lists in over 600 research reports and other publications pertaining to attitudes and the disabled, yielded the titles of only seven reviews (Anthony, 1972; Donaldson, 1980; Haddle, 1974; Horne, 1985; Towner, 1984) devoted to the research literature on modifying attitudes toward disabled persons or toward persons with a particular type of disability (mental retardation: Sandler & Robinson, 1981; physical disabilities: Westwood, Vargo, & Vargo, 1981). An additional eight reviews were identified that contained brief sections on the general topic (Alexander & Strain, 1978; Chubon, 1982; Horne, 1979) or on modifying attitudes toward persons with a specific type of disability (mental retardation: Harth, 1971; mental illness: Johannsen, 1969, Rabkin, 1972, Segal, 1978; physical disabilities: Pulton, 1976). The purpose of this chapter, as noted above, is to examine critically the full reviews and eight of the partial reviews with sufficient relevance* in terms of their methodological soundness, their contributions to knowledge, and the applicability of their findings.

---

*Pulton's (1976) review is not a traditional review of literature in the sense of an effort to summarize the state of research. Rather, the review is part of an effort to build a case for a particular approach to attitude change. The review is frequently cited for its conclusions, however, so we have included it in our critique. Similarly, Haddle's (1974) article is (continued on next page)

Judging the Quality of Reviews

Conducting interpretative reviews of the research literature is a common social science activity. Well-defined procedures for assimilating the findings of a number of primary studies have been formalized only recently, however (Jackson, 1980; Light & Pillemer, 1982). Prior reviews have tended to be "unsystematic . . . narrative, and subjective" (Light & Pillemer, 1982, p. 2). That traditional narrative review approach has in recent years been subjected to a great deal of criticism (e.g., Cook & Leviton, 1980; Cooper & Rosenthal, 1980; Glass, 1976, 1977; Hunter, Schmidt, & Jackson, 1982; Jackson, 1978, 1980; Rosenthal; 1984), with Jackson (1978, 1980) presenting what is undoubtedly the most systematic study of the status of the review literature.

Jackson (1978, 1980) analyzed a random sample of 36 integrative review articles. He concluded, as we noted in Chapter 1, that important weaknesses were pervasive. They included lack of attention to previous reviews, incomplete literature searches, inadequate summaries of study findings, and the absence of systematic examinations of relationships between study

---

included as a full review even though it contained two parts, the second of which was directed toward building a rationale for using systematic desensitization to modify attitudes. Despite the title of a paper by Bernotavicz (1979), the focus is on the research on visual presentations rather than on modifying attitudes toward disabled persons. Consequently, it was not included in this chapter, even though it would be an excellent source for someone interested in an intensive examination of information-media attitude change studies. Other review articles (e.g., English, 1971; Frith & Mitchell, 1981; Levitt & Cohen, 1976) were excluded because they did not address the research on modifying attitudes toward disabled persons. The brief review by Yuker et al. (1970, pp. 87-93) was not included, although often cited, because of its focus on research with the Attitudes Toward Disabled Persons (ATDP) scale. Also, Horne (in press) was not received in time to be included, but it is in most respects similar to Horne (1985).

characteristics and outcomes. Following up on these conclusions, Jackson (1978, Ch. 6) proposed seven crucial tasks to be considered when planning a review or judging the quality of an existing review. These tasks are: (1) The selection and definition of the topic, (2) the use made of previous reviews, (3) the selection of studies to be included in the review, (4) the collection of data from the primary research reports, (5) the analysis of data, (6) the interpretation of the results, and (7) reporting the review. It is not a coincidence that the components to be considered in planning or critiquing a review are similar to those to be taken into account in designing or evaluating a piece of primary research. The reviewer and the primary researcher share the goal of making accurate generalizations based on data which they collect (Jackson, 1978, p. 7).

Using the seven areas identified by Jackson (1980) as a general framework, we developed six sets of questions to provide the context for judging the quality of the seven reviews and the brief sections in the eight general review articles dealing with modifying attitudes toward disabled persons.

1. Formulating the Problem

   a. Was the problem clearly defined and delimited, and its importance justified?
   b. Were central terms clearly defined?
   c. Were questions identified that the reviewer attempted to answer or hypotheses stated that the reviewer sought to test?
   d. Were the questions and hypotheses adequately warranted by reference to theory, previous reviews, research, or soundly based insight?

2. Building on Prior Reviews

   a. Were previous efforts to review similar bodies of research cited?
   b. Were prior reviews critiqued as a basis for (1) the justification of another review as different from or an extension of prior reviews, (2) an appropriate point of departure for the review, and (3) avoiding the inadequacies and errors of prior reviews?

3. Selecting Studies to be Reviewed

   a. Was the method of locating studies (e.g., the indexes, reference lists, bibliographies used) described?
   b. Were criteria for selecting and excluding studies to be included in the review clearly explained?
   c. Was a representative or comprehensive sample of prior research on the problem reviewed?
   d. Did the sample of studies reviewed have a bearing on the problem?
   e. Was the sample biased either by being too small or by the failure to include relevant studies?

4. Collecting Data from the Primary Studies

   a. Were data collected for each study on common dependent and independent variables?
   b. Were data collection categories defended on rational and empirical grounds?
   c. Were data collection procedures specifically described?
   d. Were findings from studies recorded in terms of effect sizes?

5. Analyzing the Data

   a. Did the examination of relationships between dependent and independent variables take into account concomitant variables that might have influenced the results, including sample characteristics, assessment instruments, statistical techniques, and design factors?
   b. Did the reviewer try to account for any findings within the sample of studies analyzed?
   c. Were serious methodological weaknesses in studies identified?

6. Reporting and Interpreting the Findings

   a. Were the findings, including the results of analyses, reported clearly, for example, using summary tables to help readers comprehend the pattern of results from the primary research reports?
   b. Were the conclusions of the review sufficiently supported by the data and analyses?
   c. Did the review contain implications for policy or practice?
   d. Did the reviewer draw conclusions about attitude change theories?
   e. Did the review contain recommendations for future research or reviews?

Review of the Reviews

On the following pages, the six sets of questions are applied, in the order stated above, to the seven reviews and eight review sections. Specific examples are included to illustrate the applications and, as is appropriate in scholarly work, studies are cited so that our readers can check on our interpretations and analysis.

Such specific referencing of weaknesses in studies is usually not done, perhaps to avoid subjecting the authors who are cited to potential embarrassment. But as careful attention to reviews of literature as a scholarly activity becomes more common, critiques with specific citations, such as Slavin's (1984) criticism of prior meta-analyses, ought to become more frequent. Only in that way will readers be helped to develop a sense of which reviews can be relied on and to what extent, and of the considerations that are important in evaluating and preparing reviews.

It should be noted as well that, like the perfect piece of primary research, the perfect (uncriticizable) review of research may be well nigh impossible. Our own integrative review of research, to be reported in the following chapters, will not elicit from its critics a strongly affirmative answer to every question in the list. Furthermore, some of the reviews of research we critique on the following pages were written before the recent focus of attention on the conducting of reviews became a matter of general scholarly concern. Finally, our intent is not to demean the previous authors, but to try to learn from their efforts.

Formulating the Problem

The appropriate starting point for a review of literature, as with primary research, is a clear statement of the perplexity underlying the scholarly effort. Hence, the first set of questions to serve as a context for judging reviews has to do with problem formulation.

Rationales for conducting reviews of the research literature on modifying attitudes toward disabled persons were provided in all but one

(Haddle, 1974) review. For the most part, the reviewers acknowledged the prevalence of negative attitudes toward the disabled, both in the community at large (Anthony, 1972; Harth, 1973; Johannsen, 1969; Pulton, 1976; Rabkin, 1972; Sandler & Robinson, 1981; Segal, 1978; Westwood et al., 1981) and in specific groups, such as teachers and school personnel, mental health professionals, service providers, and employers (Alexander & Strain, 1978; Chubon, 1982; Donaldson, 1980; Horne, 1979, 1985). Either directly stated or implicit within most of these reviews was the need to mount attitude change programs in order to provide better services to disabled persons (e.g., Alexander & Strain, 1978; Chubon, 1983; Donaldson, 1980; Horne, 1979; Westwood et al., 1981) or to facilitate integration by reducing public prejudice (e.g., Anthony, 1972; Harth, 1973; Johannsen, 1969; Sandler & Robinson, 1981; Segal, 1978).

The rationales for two reviews, however, appeared to be based more on scholarly interests than on the practical problem of mitigating the effects of handicapism. Towner's (1984) orientation was theoretical. Her purpose was to examine the "variability in the results" of a number of empirical studies using factors hypothesized to be requisites to effective attitude change interventions (p. 223). Rabkin's (1972) interest, on the other hand, was historical. Her brief review of attitude change studies was one aspect of a general review of the literature that included descriptions of changing patterns in attitudes toward mental illness and the treatment of mentally ill patients, and of instruments commonly used to assess attitudes toward the mentally ill.

<u>Defining terms.</u> As in designing primary research, the problem underlying a review of research ought to be stated in terms sufficiently

precise to provide unambiguous guidance in the collection, analysis, and interpretation of data. Such specificity requires both conceptual and operational definitions of relevant variables.

The construct, "attitude", is central to the topics of the reviews of literature critiqued in this chapter. (One review [Rabkin, 1972] was included even though "opinions" was used in the title, because "attitudes" was used interchangeably with "opinions" throughout the article.) Despite the signal importance of the construct of attitude, only two reviewers provided a conceptual definition of it. Harth (1973) described attitudes as "predispositions toward behavior" (p. 150), a definition attributed to an earlier work by Osgood, Succi, and Tannebaum (1957). Johannsen (1969) used the following definition of attitude, from Nunnally (1961): a "personal disposition avoiding truth as an issue". According to Johannsen, Nunnally concluded that information and attitude were distinctly different constructs and while information refers to facts that can be either proven or disproven, attitudes may or may not directly reflect agreed upon facts (p. 219).

Although Towner (1984) acknowledged the multi-dimensional nature of attitudes and reported which of the studies she reviewed contained definitions of attitude, she did not present her definition of the construct in her report.

An operational definition of attitude was not specifically given in any of the 15 reviews. Implicit, however, was the presumption that attitudes were represented by scores on whatever attitude measures were used in the primary research.

Conceptual definitions of the constructs used to identify disability groups were also missing in most of the reviews and had to be inferred from the context of the review. The disability constructs were not operationally defined, either. Terms used frequently in titles to describe the population of concern, such as "disabled", "handicapped", "mentally ill", and "mentally retarded", lack precision; interpretations of the populations to which they refer may vary widely. Johannsen's (1969) discussion of the range of definitions accorded to "mental patient" and "mental illness" illustrates that point, as does Rabkin's (1972) reference to the changing definition of mental illness as psychiatry has moved from a medical model toward one of "psychological conceptions of problems of being" (p. 155).

Questions and hypotheses. As with the planning of a primary study, questions and hypotheses, usually formulated from theory and prior research, should provide the focus for an integrative review (Jackson, 1980). Questions or hypotheses were rarely a feature of the reviews and sections of reviews under examination. In fact, hypotheses were not directly stated in any review, and specific questions to be investigated were identified only by Chubon (1982) and Donaldson (1980). It seemed likely that the questions in these reviews had been developed from perusals of the research literature. Questions in both reviews were stated in a very general manner, without reference to specific variables or attitude change theories.

Comments. Conventional standards for problem statements as a basis for research studies were generally not followed by the authors of reviews of the research literature on changing attitudes toward the disabled. Reviews in this area have frequently been weakened by the failure to provide adequate definitions for relevant variables and by not centering the reviews on significant questions and hypotheses developed from attitude theory and research.

Building on Prior Reviews

The importance of reviewing prior reviews of literature was emphasized in the paragraphs introducing this chapter. We noted that Jackson (1980) and White, Bush, and Casto (1985-86) had not found the critical analysis of prior work to be common in their samples of review articles. The present investigation of reviews yielded similar findings. Of the 15 reviews and brief reviews examined, only seven cited prior reviews; however, in only four was it made clear that reviews of the literature were being cited (Horne, 1979; Sandler & Robinson, 1981; Towner, 1984; Westwood et al , 1981).

Lists of the prior reviews available at the time each review was written are contained in Table 1. The most-cited reviews were by Anthony (1972) and Donaldson (1980). References to Anthony's (1972) review were found in Horne (1979), Haddle (1974), and Horne (1985). The last two works did not, however, identify Anthony's article as a review of the literature. Donaldson's review was referred to by Sandler and Robinson (1981), Towner (1984), and Westwood et al. (1981).

It appears that the cited reviews were accepted without question. Neither the Donaldson or Anthony review was subjected to critical examination, although methodological weaknesses in each will be noted later in this chapter. Donaldson's review, in particular, was very favorably described in Towner (1984) and Westwood et al. (1981).

Justification of a new review. With the exception of Towner (1984), the authors who cited prior reviews provided no rationale for conducting another review; nor did they state how their review differed from earlier works. Towner did acknowledge Donaldson's (1980) review, and she explained that her review differed from Donaldson's in both "format" and "focus" (p. 223).

Table 1

References to Earlier Reviews

Review	Reviews Available	References	Comments
Full Reviews			
Anthony (1972)	Johannsen (1969)	No	
Haddle (1974)	Johannsen (1969) Anthony (1972) Rubkin (1972) Harth (1973)*	Anthony (1972)	Not recognized as a review
Donaldson (1980)	Johannsen (1969) Anthony (1972) Rubkin (1972) Harth (1973) Haddle (1974) Pulton (1976) Alexander & Strain (1978) Segal (1978) Horne (1979)*	Segal (1978)	Not recognized as a review
Sandler & Robinson (1981)	Johannsen (1969) Anthony (1972) Rubkin (1972) Harth (1973) Haddle (1974) Pulton (1976) Alexander & Strain (1978) Segal (1978) Horne (1979) Donaldson (1980)	Donaldson (1980)	Cited as a review
Westwood et al. (1981)	Johannsen (1969) Anthony (1972) Rubkin (1972) Harth (1973) Haddle (1974) Pulton (1976) Alexander & Strain (1981) Segal (1978) Horne (1979) Donaldson (1980)*	Donaldson (1980) Pulton (1976)	Both cited as reviews
Towner (1984)	Johannsen (1969) Anthony (1972) Rubkin (1972) Harth (1973) Haddle (1974) Pulton (1976) Alexander & Strain (1978) Segal (1978) Horne (1979) Donaldson (1980) Sandler & Robinson (1981) Westwood et al. (1981) Chubon (1983)*	Donaldson (1980)	Cited as a review
Horne (1985)	Johannsen (1969) Anthony (1972) Rubkin (1972) Harth (1973) Haddle (1974) Pulton (1976) Alexander & Strain (1981) Segal (1978) Horne (1979) Donaldson (1980) Sandler & Robinson (1981) Westwood et al. (1981) Chubon (1983) Towner (1984)*	Anthony (1972) Alexander & Strain (1981)	Neither recognized as reviews

Review	Reviews Available	References	Comments
Brief Reviews			
Johannsen (1969)			
Rubkin (1972)	Johannsen (1969)	No	
Harth (1973)	Johannsen (1969) Anthony (1972) Rubkin (1972)*	No	
Alexander & Strain (1978)	Johannsen (1969) Anthony (1972) Rubkin (1972) Harth (1973) Haddle (1974) Pulton (1976)	No	
Pulton (1976)	Johannsen (1969) Anthony (1972) Rubkin (1972) Harth (1973) Haddle (1974)	No	
Segal (1978)	Johannsen (1969) Anthony (1972) Rubkin (1972) Harth (1973) Haddle (1974) Pulton (1976)	No	
Horne (1979)	Johannsen (1969) Anthony (1972) Rubkin (1972) Harth (1973) Haddle (1974) Pulton (1976) Alexander & Strain (1978)* Segal (1978)*	Anthony (1972)	Cited as a review
Chubon (1982)	Johannsen (1969) Anthony (1972) Rubkin (1972) Harth (1973) Haddle (1974) Pulton (1976) Alexander & Strain (1978) Segal (1978) Horne (1979) Donaldson (1980) Sandler & Robinson (1981) Westwood et al. (1981)	No	

*May not have been available at the time the review was written.

BEST COPY AVAILABLE

Furthermore, she suggested that, together, the two reviews offered a "comprehensive analysis of the literature on modifying attitudes toward disabled persons" (p. 223).

A comparison of the two reviews revealed that Towner's review did indeed differ. Although the basic reporting style of both reviews was narrative and both contained summary tables, they varied in significant ways. Towner's review was more thorough and methodological. For example, Towner's discussion of study methodologies was considerably more extensive than Donaldson's brief one-sentence allusion to some studies whose weak designs seriously threatened the generalizability of their findings (p. 505). Moreover, Towner's application of attitude change theories in her analyses of studies was much more systematic than Donaldson's attempt to explain the theoretical bases for certain attitude change strategies. And, Towner examined more than twice the number of studies reported in Donaldson's review. The justification she presented for doing another review was evidenced in the review she wrote.

Comments. Using prior reviews as a basis for subsequent reviews has not been a general feature of the literature on changing attitudes toward the disabled. Less than half (47%) of the reviews examined in this study referred to earlier reviews, and in none of the reviews were previous reviews critically examined. In only one of the most recent and systematic reviews (Towner, 1984) was there an attempt to build on prior works; however, only one of the 13 available reviews and brief reviews was referenced.

Selecting Studies to be Reviewed

Sample selection is an important, even if often ignored, factor in primary research; similarly the methods used to locate primary research reports are an important factor in a review of research. Procedures should be used that will locate the maximum number of primary studies (Jackson, 1980). A search employing computerized information retrieval systems, but excluding printed bibliographies and reference lists in primary research reports is, for example, likely to identify a set of studies that is not representative of the entire body of research. The findings of a review based on such a body of research could be severely limited, even misleading. Since the adequacy of the literature search affects the generalizability of the conclusions in a review, Jackson (1980) argued that it is the responsibility of a reviewer to report the search strategy.

As with primary research, sampling bias is a serious threat to the external validity of reviews of research. Glass (1976) has argued that all studies that can be located on a topic--i.e., the accessible population of primary research reports--should be included in an integrative review. If, however, the accessible population of primary reports is not reviewed, the sampling procedure should also be reported along with the search procedures (Jackson, 1980; White et al., 1986), so that the reader of the review can judge whether the sample represents the population of studies to which generalizations are made.

The importance of addressing the method of sample selection can be illustrated by comparing the number of primary research reports identified for the review of literature reported in later chapters with those identified in prior reviews. Our literature search yielded 273 studies that met specific criteria (as described in the Chapter 3) for inclusion in our review of research on the modification of attitudes toward the disabled. An additional 394 studies were discarded as not suitable* for the present review. By the end of the period from 1950 (the earliest study located) to July, 1986, the body of attitude change studies included at least 706 titles (including reports cited in prior reviews, but not included on any of the lists of reviewed or discarded reports for this review), of which 239 were theses or dissertations.

The sum total of individual studies concerned with attitude change cited in the seven reviews and eight brief reviews was 192**, slightly less than one-quarter of the studies available. The median number of primary studies referenced in the full reviews was 32 ($\overline{X}$ = 36); in the brief reviews, the median was 10 ($\overline{X}$ = 12)***. A correlation coefficient (Pearson product-moment *r*) of .64 indicated a moderate relationship between number of primary studies cited in each full review and the number of studies available when the reviews were in preparation. A similar relationship (*r* = .18) was not found for the eight brief reviews.

---

*Of these, 363 were deemed irrelevant because they were, for example, correlational studies, used instruments judged not to fit our definition of attitude, or attitudes toward mainstreaming rather than toward disabled persons were assessed. An additional 31 studies were discarded due to lack of information. (See Appendix D.)

**Sixty-two per cent (N=119) of these studies investigated the effects of a variety of interventions (e.g., information about the disabled, personal contact with disabled persons, simulations of disabilities) on attitudes toward the disabled. Of the remaining citations, 24% (N=46) examined relationships between attitudes toward the disabled and variables such as amount of reported contact with disabled persons, knowledge of disabilities, and membership in mainstreamed classes, and 13% (N=24) assessed the efficacy of courses in psychology, special education, nursing and so forth for enhancing university students' attitudes toward the disabled. Two percent (N=3) of the studies compared the attitudes of different professional, student, and community groups.

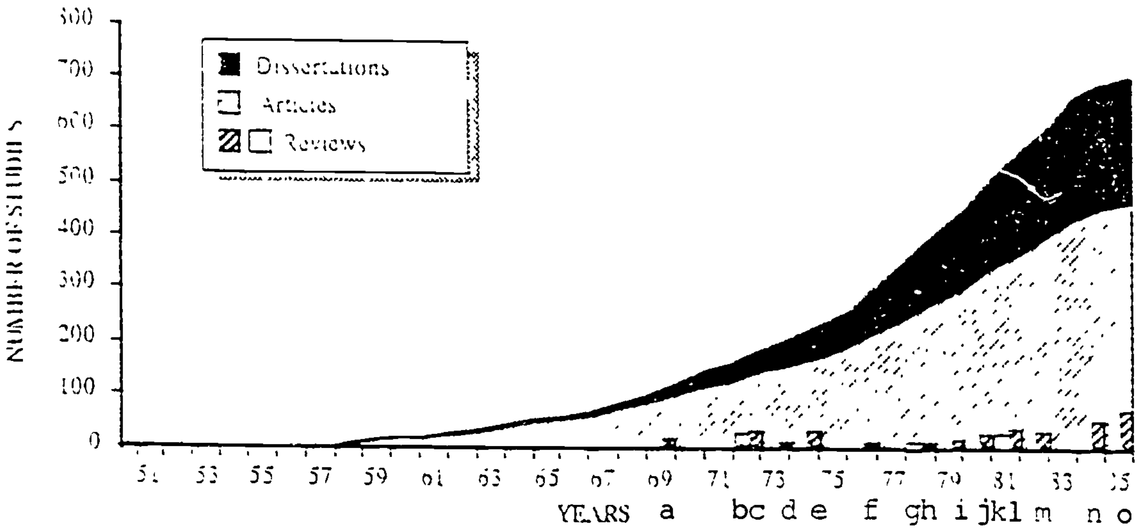
***Median and mean numbers of studies referenced are rounded to whole numbers, as references are cited in toto.

The relationship between reports available and reports cited is depicted in Figure 1. The cumulative numbers of dissertations and other reports of primary studies available from 1951 through 1985 are indicated by the shaded areas, and the top line on the graph indicates the total cumulative number of reports available. The bars on the graph show the number of reports cited in the individual reviews which are identified by letter in the list below the graph.

In interpreting the information in Figure 1, as well as in contemplating our discussion of the absence of a common core of cited studies in the next section, it is important to remember that although the majority of the reviewers were concerned with disabilities generally, two of the full reviews were directed at attitudes toward persons with specific disabilities (mental retardation: Sandler & Robinson, 1981; physical disabilities: Westwood et al., 1981), while five of the brief reviews were also limited in scope (mental retardation: Harth, 1973; mental illness: Johannsen, 1969; Rabkin, 1972; Segal, 1978; physical disabilities: Pulton, 1976). In addition, there was some focusing on groups whose attitudes were of concern (e.g., school psychologists: Horne, 1979; professionals: Chubon, 1976). Consequently, it would not be correct to assume that all of the available reports of research would be relevant to the defined topic of every review.

A common core? Seventy of the 192 studies cited in the reviews were included in two or more reviews. However, the seven most frequently referenced studies (Cleland & Champers, 1959; Clore & Jeffrey, 1972; Granofsky, 1956; Hicks & Spaner, 1962; Lewis & Cleveland, 1966; Warren, Turner, & Brody, 1964; Wilson & Alcorn, 1969) were cited in only five of the 15 reviews. Another 14 studies (Altrocchi & Eisdorfer, 1961; Anthony, 1969;

40



<u>Reviews and Number of Reports Cited</u>

a. Johannsen (1969), 12
b. Rabkin (1972), 25
c. Anthony (1972), 30
d. Harth (1973), 10
e. Haddle (1974), 31
f. Pulton (1976), 7
g. Segal (1978), 5
h. Alexander & Strain (1978), 6
i. Horne (1979), 15
j. Donaldson (1980), 24
k. Westwood et al. (1981), 29
l. Sandler & Robinson (1981), 36
m. Chubon (1982), 27
n. Towner (1984), 47
o. Horne (1985), 70

Figure 1: Research Reports Available to Reviewers and Number of Reports Cited

Brooks & Bransford, 1972; Donaldson & Martinson, 1977; Evans, 1976; Glass & Meckler, 1972; Holzberg & Gewirtz, 1963; Lazar, Gensley, & Orpet, 1971; Rapier, Addson, Carey, & Croke, 1972; Rusalem, 1967; Sadlick & Penta, 1975; Shotel, Iano, & McGettigan, 1972; Strauch, 1970; Yerxa, 1971) were referenced four times. The remaining 49 studies were cited either three times (17 studies) or twice (32 studies).

The few times that even the most frequently cited studies were included across reviews suggests that even the reviewers concerned with disabilities in general did not draw on a common core of research for their conclusions. That observation is confirmed by the data in Table 2, a matrix of common sources which was prepared to illustrate the extent to which reviewers cited studies in common. The greatest number of common sources was shared by Anthony (1972) and Haddle (1974). Of the 30 studies referenced in Haddle, 18 had been cited previously in Anthony's review. Other pairs of reviews that cited a number of studies in common were: Horne (1985) and Towner (1984), with 17 shared citations; Donaldson (1980) and Towner (1984), with 14 shared citations; and, Towner (1984) and Westwood et al. (1981), with 14 citations in common. For the most part, the overlap in reference lists was not great. Two of the 15 reviews (Alexander & Strain, 1978; Segal, 1978) contained few citations in common with other reviews.

The data from the matrix support the conclusion that there was not a core body of research common to the reviews; few studies appeared to have had acquired the status, or visibility, where their inclusion was requisite to an adequate review of literature pertaining to modifying attitudes toward the disabled. The data also suggest that later reviewers did not rely on reference lists in prior reviews to obtain studies to examine.

Table 2

Matrix of Common Sources

		Full Reviews							Brief Reviews							
		Anthony (1972)	Haddle (1974)	Sandler & Robinson (1978)	Donaldson (1980)	Westwood et al. (1981)	Towner (1984)	Horne (1985)	Johannsen (1969)	Rabkin (1972)	Alexander & Strain (1978)	Pulton (1976)	Segal (1978)	Harth (1979)	Horne (1979)	Chubon (1982)
Full Reviews	Anthony (30)[a]		18	2	3	2	2	3	3	5		2	2	2	2	5
	Haddle (31)	18		2	6	2	2	2	2	3		2	1		1	6
	Sandler & Robinson (36)	2	2		2		2	12	2	1		1		4		3
	Donaldson (24)	3	6	2		12	14	6				3			2	6
	Westwood et al. (29)	2	2		12		14	6				3				4
	Towner (47)	2	2	2	14	14		17			1	3			2	8
	Horne (70)	3	2	12	6	6	17		1		2	2		2	4	6
Brief Reviews	Johannsen (12)	3	2	2				1		2				1		2
	Rabkin (25)	5	4	1					2				2	1		4
	Alexander & Strain (6)						1	2							2	
	Pulton (7)	2	2	1	3	3	3	2								
	Segal (5)	2	1							2						2
	Harth (10)	2	1	4				2	1	1						
	Horne (15)	2	2		2		2	4			2					
	Chubon (27)	5	6	3	6	4	8	6	2	4		1	2			

[a]Number of primary research reports cited.

Dissertations. Dissertations and theses, which comprised over one-third of the studies located for our review, were almost totally ignored in the prior reviews. The number (N=27) cited was slightly more than one-tenth of the number available. The median and mean numbers of dissertations and theses referenced in the full and brief reviews were, rounded to whole numbers, 6 ($\bar{X}$ = 5) and 3 ($\bar{X}$ = 2), respectively. Moreover, the majority (84%) of dissertation citations were references to abstracts in Dissertation Abstracts and Dissertation Abstracts International, rather than to the actual dissertations. One of the two references to Wyrick's (1968) magistral thesis was, in fact, a reference to the work in a secondary source.

Search techniques. Given the lack of overlap among reference lists, it is probable that methods for locating studies varied among the reviews. The search method was reported in only one review (Chubon, 1982). Chubon described his search as including a manual search of the journals published for the "helping professions" (p. 25) and computer-assisted searches of ERIC, Dissertation Abstracts, and Psychological Abstracts for the period 1960-1979. Of the 102 articles he located, 62 were discarded because they did not describe empirical studies. Twenty-eight of the remaining articles were cited in the brief section on modifying attitudes toward the disabled in his general review of literature. For the period Chubon reviewed, the population of studies concerning attitude modification in this area numbered at least 105 dissertations, 5 theses, and 275 articles, unpublished papers, and project reports.

It was not possible to identify the search strategies for the remaining 14 reviews. Neither could the reviewers' criteria for including or excluding studies be ascertained. As a consequence, it was difficult to judge the representativeness of the sample of studies cited in most reviews, and the possibility of sampling bias as a threat to the generalizability of the conclusions in each of these reviews has to be seriously considered. In two reviews (Alexander & Strain, 1978; Horne, 1979), however, sampling bias was beyond doubt as references were limited to reports of successful interventions only. Bias was also highly likely when only one study out of the many available was cited in a discussion of an intervention (Pulton, 1976). Donaldson (1980) also provided an example of a limited sample of primary studies, using only two studies of simulation, one positive and negative, to draw a conclusion.

Study relevance. A second concern with sampling is the appropriateness of the research studies for the specific questions the review is intended to address. Reviews which attempted to identify effective strategies for modifying the attitudes of particular groups, such as health professionals (e.g., Chubon, 1982), educators (e.g., Alexander & Strain, 1978), or peers (e.g., Horne, 1985), tended to cite studies that were relevant to these groups. On the other hand, reviewers who sought strategies for changing societal attitudes (e.g., Anthony, 1972; Donaldson, 1980) had to extrapolate to the general population from the findings of studies conducted with specific populations, such as university psychology students, nursing students in internship programs, and special education students in practica. Only occasionally did reviewers remind readers of the importance of viewing the conclusions from a limited set of findings with caution (e.g., Anthony, 1972).

Another difficulty was presented when reviewers drew inferences for specific contexts from studies not directly relevant. Sandler and Robinson

(1981), for example, reviewed the literature in an attempt to locate "factors which may be related to improved public attitude toward development of group homes for mentally retarded people in the community" (p. 98). Only two of the 33 studies they examined had a direct bearing on this problem. The other 31 reports included, among other things, interventions conducted in schools and universities, the effects of mainstreaming on the attitudes of nondisabled peers, and correlations between the amount of reported contact with disabled persons or the amount of knowledge about disabilities and attitudes toward the disabled.

Comments. The lack of information concerning literature search strategies and sampling procedures cast doubt on the representativeness of the studies cited in most reviews and severely limited the generalizability of the conclusions of the reviews. Failure to include this information was a particular problem because of the relatively small number of studies cited in each review. Doctoral dissertations were a neglected source of primary studies. When dissertations were referred to in reviews, frequently only the abstracts in Dissertation Abstracts were cited. A search of the reference lists in the reviews failed to reveal any studies that were commonly cited once they were available.

Collecting Data from the Studies

Along with identifying primary studies for an integrative review, procedures for collecting data from the studies must be established. The procedures for collecting data from primary studies for a review should meet standards similar to those for data collection in primary research. That is, they should be "systematic", "well-planned" (Borg & Gall, 1983, p. 840), and "organized in a manner which facilitates analysis" (Gay, 1976, p. 218).

Furthermore, since the methods for collecting the data influence the outcomes of analysis and the credibility of the interpretations in the review (Jackson, 1980), they should be described carefully.

None of the 15 reviews contained a description of data collection procedures. How data were collected could not be inferred in any of the reviews, since, for the most part, the results of primary studies were reported in narrative form. For example, there was no way of knowing if some sort of coding sheet was used or if notes were taken while reading studies. Failure to include information on data collection, or to present results in a manner that makes the method evident, appears to be prevalent in reviews of research (Jackson, 1980; White et al., 1986).

Although data collection procedures were not described, some inferences can be drawn about adequacy of collection. For example, the reporting of primary studies in most reviews was organized under broad categories according to intervention techniques, with "contact" and "information" the most frequently used headings. There was a tendency to describe intervention strategies only in a very general manner. Important study characteristics--such as the treatment setting, who conducted the treatment, and treatment length--were ignored, suggesting that data were not collected on these variables. For example, treatment length, which may be a significant factor in the effectiveness of an intervention, was mentioned for some studies in only five reviews (Donaldson, 1980; Haddle, 1974; Horne, 1985; Rabkin, 1972; Towner, 1984) and for most studies in only one (Anthony, 1972) review.

Describing interventions. The reviews would have been strengthened by adequate description of the intervention techniques. For instance, in discussing "enforced contact" as a strategy for modifying attitudes toward the disabled, Sandler and Robinson (1981) stated, ". . . Aloia, Beaver, and Pettus (1978), Leyser and Gottlieb (1980), and Marlowe (1979) have also reported the use of carefully planned interventions which improved the social status of integrated EMR children" (p. 99). In each of these studies, however, there were important differences not only in sample characteristics, but in treatments. These differences were not identified, nor was the reader informed whether the differing interventions were equally effective.

Horne's (1979) review provides a similar example. Contending that the research has indicated that promoting peer acceptance will lead to positive attitude change, Horne stated that "positive peer interaction . . . may be facilitated . . .", and cited Kirby and Toler (1970) and Strain and Timm (1974) to support that point. However, distinctly different strategies were employed in each study. It was not indicated in the review that the former article reported a study in which a 5-year-old boy's interaction with peers increased when he passed out candy, while the latter article described a program which utilized praise and contact to increase appropriate social behaviors in a "disordered pre-school child". Again, the question is raised whether such information was collected.

Another serious problem occurs when the primary studies are referred to in such a way that the reader is not even able to infer the general categories of intervention techniques. An example comes from Chubon (1982). Following a statement about the mixed results with treatments designed to "enhance attitudes of teachers and students majoring in various areas of education toward disabled students", Chubon stated, "while some attitude change programs seemed to produce the desired results . . . others have produced no changes" (p. 26). He cited Kuhn (1971), Parish, Eads, Reece, and

Piscitello (1977), Wilson and Alcorn (1969), and Zukerman (1975) as relevant studies. Without referring directly to each study, the reader is unaware that the intervention techniques included exposure to blind children (Kuhn, 1971), an introductory special education course (Parish et al., 1977), simulations of disabilities (Wilson & Alcorn, 1969), and gaming (Zukerman, 1975). The implication is that data were not collected on important treatment attributes such as these.

A similar example is taken from Horne (1985). Commenting on the efficacy of programs to enhance teachers' attitudes toward disabled students, Horne wrote, "Sometimes changes have occurred in negative directions . . . ." (p. 158). References for this statement were Bradfield, Brown, Kaplin, Richert, and Stannard (1973), Warren, Turner, and Brady (1964), and Sellin and Mulchahay (1965). Again, without referring directly to these studies, the reader does not know that the treatments consisted of a single trip by high school seniors to a state institution for the mentally retarded (Sellin & Mulchahay, 1965), tours of institutions for the sight impaired, hearing impaired, and the mentally retarded conducted within the context of an integrated "psychology-education-sociology" program for sophomore college students (Warren, Turner, & Brody, 1964), and an inservice training program designed to instruct teachers of integrated classrooms in individualized instructional techniques (Bradfield, Brown, Kaplin, Richert, & Stannard, 1973). Again, it is open to question whether data were collected on these central treatment characteristics.

<u>Program description-research confusion.</u> Some authors cited narrative descriptions of programs along with reports of primary research, which raises questions about their data collection procedures. Johannsen (1969), for

example, in a discussion of employers' attitudes toward hiring ex-mental patients, referred, respectively, to programs described by Murray (1958) and Brennan and Margolin (1954) with the phrases "promising results were forthcoming" and "success was also reported". These references were included with references to primary research studies and the title of Johannsen's article described it as a "review of empirical research" (p. 218). Consequently, a reader might assume that both the Murray and Brennan and Margolin articles reported the findings of empirical studies. However, neither was a report of research; both were narrative accounts of programs advocated by the authors for encouraging the acceptance of ex-mental patients in the workplace.

Conclusions-findings confusion. In a somewhat similar vein, questions can be raised about data collection when reviewers treat recommendations from the summary and conclusions sections of reports as though they were research findings. Rabkin (1972), for instance, examined the research pertaining to modifying mental hospital attendants' attitudes toward patients. She stated that "Middleton's (1953) work suggests that training [for attendants] ought to include a thorough indoctrination about etiology, treatment results, examples of success, and reasons for failure, with periodic repetitions of this training" (p. 163). In his study, Middleton compared the attitudes of attendants and non-attendants in a mental hospital. In the conclusions section of his article, he recommended the type of instruction referred to by Rabkin. However, such instruction was not provided to the subjects in his study.

A similar example comes from Johannsen's (1969) brief review of the research on changing attitudes toward mental patients. He suggested that the effectiveness of personal contact as an attitude change strategy may depend upon others' perceptions of mental patients' behavior as being normal. He concluded with the statement, "Halfway houses . . . serve a public education function by enabling relatives to see the patient move gradually into the community" (p. 224), and cited Pettit (1956) as the source. Pettit's study investigated the attitudes of relatives toward deinstitutionalizing "long-hospitalized" mental patients. In the closing section of his article, Pettit suggested that halfway houses might ease the movement of mental patients into the community. This was an opinion, and the adequacy of halfway houses for modifying the attitudes of relatives was not examined.

Other problems. Clearly, collecting data in such a way that researcher's conclusions and recommendations are kept distinct from findings would enhance the validity of reviews. Other problems that may stem from inadequate data collection are presenting studies incorrectly, failing to consider all of the results from a piece of research, or citing irrelevant studies. Instances of such practices were identified in 11 of the 15 reviews. The following examples were selected from Sandler and Robinson (1981), Alexander and Strain (1978), Towner (1984), and Horne (1979).

Reporting studies incorrectly

In their discussion of information as a factor in changing attitudes toward the disabled, Sandler and Robinson (1981) indicated that Begab (1969) ". . . attempted to improve the attitude of social work students by providing them with coursework on mental retardation and found that knowledge alone had little effect upon attitude. When coupled with direct contact with mentally retarded persons through a fieldwork experience, however, there was positive increase in attitude" (p. 99). A perusal of the abstract of Begab's dissertation (cited by Sandler and Robinson) revealed that he did not attempt to change students' attitudes; he conducted a survey of students enrolled in several schools of social work to investigate the "impact of differences in curricula and experiences on social work students' attitudes and knowledge about mental retardation" (Begab, 1969, p. 4111-A).

Alexander and Strain (1978) included Lane's (1976) dissertation abstract in their review of educators' attitudes toward disabled children and mainstreaming. They stated that his study "showed that a background in special education can help alleviate stereotypes or prejudice toward exceptional children". However, the purpose of Lane's study was to "investigate the effects of labels conveying ethnic group membership and retardation on evaluative statements made by prospective teachers" (Lane, 1976, p. 1491-A). The subjects were dual majors in elementary and special education. No comparison was made with a control group of subjects who did not have a special education major. The following hypothesis was not supported: "Knowledge of a child's ethnic group label or label indicating the presence or absence of mental retardation would result in more negative ratings for those labeled retarded than not so labeled" (Lane, 1976, p. 1491-A). Lane concluded that "Neither the label nor the ethnic identification were [sic] found to significantly affect ratings and no interaction effects were demonstrated" (p. 1491-A). The relationship between a background in special education and attitude toward exceptional children was not a concern in the study.

Partial results

When only partial results of a primary study were reported, suggesting partial collection of data, misleading conclusions were sometimes drawn about the effectiveness of an intervention, or factors that had a bearing on the outcome of the study were not described. For instance, Towner (1984) reported that simulation exercises were successful in modifying the attitudes of elementary children in a study by Dahl, Horsman, and Arkell (1978). Towner's conclusion, however, was based on a statistically significant F-ratio for the difference between the adjusted posttest means of the experimental and control students on only one of four items in a social distance scale. F-ratios for the differences between the adjusted group means on the remaining three items and on two additional attitude measures were not statistically significant. Dahl et al. (1978) stated that the results "suggest that simulations may have limited value for changing Grade 5 students' attitudes toward their handicapped peers" (p. 574).

Selective reporting

An example in which the selective reporting of findings, perhaps reflecting the selective collection of data, resulted in the loss of valuable information about an important mediating variable comes from Sandler and Robinson (1981). They reviewed several studies pertaining to the effects of mainstreaming mentally retarded children on the attitudes of their non-disabled peers and concluded that "investigations involving EMR children integrated into regular classrooms have consistently shown more negative attitude change among nonhandicapped children related to increased exposure to EMR children" (p. 98). Among the studies cited was Goodman, Gottlieb, and Harrison (1972). A perusal of this study revealed that gender was a significant factor in the acceptance of mainstreamed mentally retarded children; although boys rejected integrated mentally retarded children more than segregated ones, the acceptance or rejection of mentally retarded children by girls was independent of educational setting. This important relationship was disregarded in the Sandler and Robinson review.

Horne's (1979) review provided an example of an irrelevant citation that might be due to incomplete data collection. Referring to studies by Kearney and Rocchio (1956) and leBue (1959), Horne wrote, "Studies done using teacher populations have included an exploration of the relationship between information about the handicapped and attitudes toward the handicapped . . . as well as the effect of teaching experience or contact" (p. 63). In both studies, the Minnesota Teacher Attitude Inventory was used to assess attitudes, and in neither article was reference made to handicapped children nor was attitude toward the disabled measured.

Outcomes. Another important aspect of a research review is the treatment of the dependent measures used in the primary studies. For the most part, the reviewers' reports of data pertaining to dependent variables suggested that data collection was inadequate. Dependent variables were not adequately described, nor was it possible to discern how the reviewers gathered data concerning study outcomes.

In only three (Haddle, 1974; Rabkin, 1972; Towner, 1984) of the 14 reviews were the means of assessing dependent variables identified for individual studies. Of the remaining 11 reviews, three (Alexander & Strain, 1978; Anthony, 1972; Segal, 1978) reported dependent measures for several but not all of the primary studies cited, and eight (Chubon, 1982; Donaldson, 1980; Harth, 1973; Horne, 1979; Johannsen, 1969; Pulton, 1978; Sandler & Robinson, 1981; Westwood et al., 1981) made no mention of how study outcomes were measured.

In addition, the findings of studies were generally reported only narratively, suggesting that quantitative data, either statistical significance or effect sizes, were not collected. There were two exceptions. Statistical significance levels were reported by Haddle (1974), and Towner (1984) provided summary tables with studies identified as successful if "statistically significant results" were obtained. Donaldson (1980) also provided summary tables in which she grouped studies according to "positive change", "no change", and "negative change" (p. 506). Whether these headings referred to statistical significance or some other criterion could not be determined from the article.

Typical of the narrative statements reporting the findings of studies were the following:

> Quay, Bartlett, Wrightsman, and Catron (1961) used three methods of presenting material to a group of attendants. They found that the formal lecture method was more effective in changing reported attitudes than either the discussion method or use of a booklet. (Rabkin, 1972, p. 165)

> Voeltz (1980) has demonstrated that providing structural contact experiences to elementary school children is related to more positive attitude toward integrated severely handicapped students. (Sandler & Robinson, 1981, p. 99)

> Evans (1976) has also reported that interaction with disabled persons which was structured to alleviate interpersonal discomfort was effective in enhancing the attitudes of psychology students, and Hersh, Carlson, and Lossino (1975) found that social work students' attitudes towards retarded persons seemed to be enhanced by interaction with families having retarded children. (Chubon, 1982, p. 27)

> Negative changes in attitude toward and optimism concerning the integration of exceptional learners were found by Fenton (1975) and Shotel et al. (1974). (Towner, 1984, p. 236)

> In another summer workshop, administrators and classroom teachers were involved in: (1) a practicum placement working with handicapped children for 3 hours per week; (2) two additional hours of observation; (3) enrollment for 9 hours of graduate credit; and (4) weekly sensitivity sessions. Post-testing using a semantic differential showed significant and positive attitudinal changes (Brooks & Bransford, 1971). (Horne, 1985, p. 158)

The 15 reviews were replete with similar statements in which effect sizes were not given or statistical tests or significance levels were not reported.

One of the dangers of not collecting information on levels of statistical significance so that it can be reported in statements alluding to positive findings in primary studies is that the reader may be tempted to assume that statistical significance was reached. In five of the 14 reviews, such a conclusion could be erroneous. For example, Horne (1979) referred to a specific study to support her advocacy of programs designed to modify the behaviors of disabled students which inhibited their acceptance by nondisabled peers and programs to "facilitate acceptance on the part of regular classroom members" (p. 64). She commented, "Although few such programs have been presented in the literature, Simpson, Parrish, and Cook (1976) have demonstrated the efficacy of using such procedures with elementary school pupils" (Horne, 1979, p. 64). The article by Simpson et al. described two studies, the first conducted with primary children and the second with grade 5 students. The designs for both studies included control groups, and the Attitude Toward Disabled Persons Scale was used to assess attitudes. The *t*-test was used to analyze the differences between means for the primary children. The difference between the posttest scores of the experimental and control groups was not statistically significant. ANOVA was used to analyze the differences among the mean change scores of the three fifth grade experimental groups and the one control group (or two control groups--the report is unclear). The *F*-value was not statistically significant. Horne's statement might lead readers who did not have access to the primary report to assume that the results were statistically significant.

A second example comes from Anthony (1972). He referred to a study reported by Cowen, Underberg, and Verrillo (1958) with the comment, "These researchers found that individuals who had had contact with the blind tended to have more negative attitudes than individuals reporting no contact" (p. 118). Cowen et al. reported developmental data for the Attitude Toward Blindness Scale. Their sample consisted of university students in educational psychology courses who were divided into groups according to reported contact or no reported contact with blind persons. The mean score for the contact group was 54.53 (SD = 12.21), while the mean score for the no contact group was 53.58 (SD = 12.21). Not only is the difference visibly small, especially relative to the standard deviation, but the t value for the difference between the means of the two groups was 0.39, clearly not statistically significant (p. 300). In contrast with Anthony's comment, Cowen et al. concluded that, ". . . in the present study, though clearly not significant [underlining added], there is a slightly higher mean score (more negative attitudes) reported by those who have had previous contact with the blind" (Cowen et al., 1958, p. 300).

Comments. The failure to describe data collection procedures was common to the 15 reviews. The problems in the way data were presented that seemed likely to reflect inadequate data collection included the following: (1) studies were grouped together under summary statements that were not accurate for all of the studies cited and that ignored differences in independent and dependent variables; (2) studies were cited with no information provided about treatment variables; (3) comments in the conclusion or summary sections of primary articles were referred to as if they were findings from the studies; (4) program descriptions were referred to as though they were research studies; (5) studies were presented inaccurately, including reports of partial findings and the citation of irrelevant studies.

In addition, narrative reporting of the findings from primary studies was common to the reviews. Only two of the 14 reviews even reported whether primary results were statistically significant. In some reviews, findings were reported in a manner that implied statistical significance when, in fact, significance had not been reached.

Analyzing the Data Collected from Primary Studies

As with primary research, after data collection the next stage in conducting a review of research is the analysis of the data collected from primary studies. The analysis is the basis for the reviewer's inferences from the findings. Analysis is complicated by a number of factors (e.g., sample characteristics, design, intervention implementation, the assessments used) that may vary with outcomes and should be considered when attempting to draw conclusions (Jackson, 1980).

Important decisions that could significantly affect the quality of the review face the reviewer during the analysis of data stage. Among the matters that must be decided are how to attempt to get at relationships between outcomes and the concomitant variables that might have confounded findings, how to treat the results from studies with multiple, even conflicting, outcomes, how to weight studies that vary in degree of methodological soundness, and what analyses to conduct to explain contradictory findings among primary studies?

Concomitant variables. None of the 15 reviews being critiqued in this chapter examined the effects of concomitant variables such as age of subjects and gender. As will be noted later, Towner (1984) did at least examine methodological deficiencies, even though she did not relate them directly to differences in outcomes. Most of the 15 reviews dealt with simple treatment-dependent variable relationships, despite the number of sample attributes (e.g., age, sex, intelligence, education) and intervention characteristics (e.g., setting, length of treatment) that might have been related to changes in attitudes toward disabled persons. That approach might have been in part a reflection of the design and the analysis strategies in many of the primary studies which were reviewed. Particularly in the earlier ones, the t-test was often first used to test the difference between the pretest means of the experimental and control groups; if no statistically significant pretest difference was found, then the t-test was used to test pre-post mean differences for each group. If the experimental group mean gain was statistically significant and the control group mean gain was not, it was concluded that the difference between the two was significant. When the posttest means for treatment groups were compared, using the _t_-test or one-way analysis of variance, other factors that might have had a bearing on the outcomes were usually not included in the analysis.

Even when the primary studies involved complex designs with complex analyses, however, the tendency was to report the findings of these studies in simple treatment-outcome terms. The reference to Hafer and Narcus's (1979) study in Westwood et al. (1981) was typical of the treatment of complex findings in most of the reviews.

The purpose of the Hafer and Narcus study was to investigate the effects on 76 college students' attitudes toward disabled persons of viewing a film (_Like Other People_) designed to present the needs and feelings of persons with cerebral palsy as similar to those of nonhandicapped people, as compared to viewing a Laurel and Hardy comedy. Whether pretesting would have an effect was also investigated. The _Attitude Toward Disabled Persons_ scale (ATDP) was administered to two of the four randomly assigned groups as a pretest. All groups took the ATDP immediately following the viewing of one of the two films. Six weeks later, the ATDP was administered again to the total sample.

Multiple classification analysis of variance revealed that there was a statistically significant mean difference between the groups which saw the two films, with the Laurel and Hardy group having the higher (more positive) mean. There also were statistically significant interactions between the "specific movie seen and whether a group was pretested or not" and between "specific movie and time of administration of the posttest" (p. 99). Nonpretested students who saw the two movies had similar posttest means, but pretested students had a higher mean with the Laurel and Hardy movie than with Like Other People. The mean difference between viewers of the two films detected at the first posttest was, however, not evident at the follow-up testing. Despite this complexity, the Hafer and Narcus study was reported as follows in the Westwood et al. (1981) review: "[O]ther studies involving college students (Hafer & Narcus, 1979; Wyrick, 1968; Yerxa, 1971), nursing students (Rosswurm, 1980), high school students (Forader, 1970), and grade school students (Perkins-Karniski, 1978) also produced equivocal results with three showing positive change (Perkins-Karniski, 1978; Rosswurm, 1980; Yerxa, 1971) and three showing no change (Forader, 1970; Hafer & Narcus, 1979; Wyrick, 1968)" (p. 221).

Study quality. As was indicated in the footnote on page 20, of the 192 primary studies that were cited in the seven full and eight brief reviews, 119 were investigations of the efficacy of interventions for modifying attitudes toward disabled persons. An additional 24 studies assessed the effects of university programs and courses on students' attitudes toward the disabled. These 143 studies comprised 74% of the individual studies referenced in the 15 reviews. As part of the integrative review to be reported in the following chapters, 114 (about 80 %) of the 143 studies were coded for treatment validity and internal validity, using the coding instrument presented in Appendix B. The two types of validity were coded for the other 20 percent this review of reviews.

None of the 143 primary studies was judged to be "excellent" in treatment validity (Category C.9.b.) or "high" in internal validity (Category E.3.i.). The treatment validity of 55% (N = 79) of the studies was coded as "fair", and that of the remaining 45% (N = 64) was coded "poor". Major threats to treatment validity were multiple treatment interference, shortcomings in treatment implementation, test by treatment interaction, and experimenter effects. The internal validity of 36% (N = 52) of the studies was judged to be "medium"; that of the remaining 64% (N = 91) was coded as "low". Selection, history, and instrumentation were found to be serious threats to internal validity in many studies.

Contrary to the above findings, only a few reviewers expressed concern for the general quality of the primary studies they cited. Towner's (1984) conclusion that the findings of many of these studies "can only be characterized as contaminated" (p. 251) was in accord with earlier comments by Anthony (1972), Chubon (1982), Donaldson (1980), and Haddle (1974). However, only Towner (1984) identified the studies she judged to be methodologically weak and attempted to examine systematically the methodological deficiencies in the studies. Anthony limited his criticism to contact studies in which the independent variable was the self-reports of respondents as to the amount of their contact with disabled persons. He argued that such studies were "methodologically deficient" (1972, p. 119), because each subject decided for himself or herself just what was meant by contact. Donaldson questioned the generalizability of several studies she considered to be poorly designed, but there was no indication that these studies were assigned a low rating when the effectiveness of particular interventions was weighed. Similarly, although Chubon (1982) and Haddle (1974) commented that many studies had weak designs, they did not identify specific studies. It was, therefore, difficult to determine whether adequacy of design was weighted in their discussions of findings.

Theory base. Whether outcomes vary between studies with and without a theoretical focus is an important analysis question. Of 224* reports of primary research coded for the meta-analysis reported in following chapters, only 20% (N = 44) explicitly identified the attitude change theories upon which the interventions were based. Only five of the reviews (Chubon, 1982; Donaldson, 1980; Haddle, 1974; Harth, 1973; Towner, 1984) mentioned this shortcoming. And, with the exception of Towner, comments concerning the failure of most researchers to ground their studies in theory were general in nature; studies that lacked a theoretical base were not specifically identified. Whether this lack of a theoretical base for the primary studies was considered in drawing the conclusions in these reviews is unclear.

Attitude assessment. Whether outcomes covary with quality of outcome assessment is another pertinent analysis question. Critical comments concerning the assessment of attitudes were, however, infrequent in the reviews, and relationships between instrumentation and outcomes were not discussed. Although six reviewers (Alexander & Strain, 1978; Anthony, 1972;

---

*This number excludes studies of mainstreaming in classrooms and studies for which the information to compute effect sizes was not available.

Haddle, 1974; Rabkin, 1972; Segal, 1978; Towner, 1984) identified some or all tests used in the primary studies they examined, T. ner alone pointed out inadequacies in the reporting of assessment in many primary studies. In particular, Towner criticized the lack of reliability and validity data in the instrumentation sections of most primary research reports. Furthermore, Towner noted that instruments developed for individual studies were often poorly described and test development data were seldom given. She did not, however, use that information in analyzing the variance among study outcomes. Chubon (1982) also criticized the means for assessing attitudes in the studies he reviewed, but his comments were general, with references to neither specific studies or tests. Nor did Chubon specifically consider weaknesses in instrumentation when he judged the effectiveness of the interventions he reviewed.

A perusal of the primary studies included in the 15 reviews disclosed numerous examples of poorly designed and inadequately reported assessment. Much like the primary reports coded for the integrative review reported in the following chapters*, validity and reliability data we infrequently reported in the primarv studies included in the 15 reviews. This was true even for those instruments such as the Opinions A..ut Mental Illness scale (OMI) and the Attitude Toward Disabled Persons scale (ATDP) for which extensive documentation exists. The failure to provide satisfactory descriptions and developmental data for tests prepared for sp ic studies was common, too. Examples of the latter problem were found in, among others,

---

*With 704 effect sizes computed for 214 studies, reliability coefficients for the attitude assessments were reported for 411, or 58%; validity was mentioned for 330, 47%, of the 704 effect sizes.

Harasymiw and Horne (1975), Quay, Bartlett, Wrightsman, and Catron (1961), Shotel, Iano, and McGettigan (1972), and Stephens and Braun (1980). Nevertheless, the reviewers who referenced several (Chubon, 1982; Donaldson, 1980; Harth, 1973; Rabkin, 1972) or all (Sandler & Robinson, 1981) of these studies apparently overlooked these deficiencies in their analyses of study findings.

Conflicting results. Table 3 records the results from the primary studies that were included in the reviews, arranged according to review authors and attitude change strategies. As would be expected, most reviewers included studies in which the interventions were deemed to be successful in modifying attitudes toward the disabled and studies which produced nonsignificant or negative findings. Deciding how to handle discrepant findings is an essential part of the analysis of data for a review, and has been the subject of a number of articles. Light and Pillemer (1982) proposed that contradictory findings provide the reviewer with "an opportunity to examine and explain variations in outcomes" (p. 6). In an earlier article, Light and Smith (1971) argued that to ignore discrepant findings is to "assume that genuinely contradictory results can never be a valid description of reality" (p. 438).

Systematic approaches to the examination of conflicting findings were described in Jackson (1980), Ladas (1980), and Light and Smith (1971). They suggested that conflicting findings may result from a number of factors, including sampling error, grouping under the same attitude change strategy studies with different intervention characteristics, methodological inadequacies, and instrumentation. The implication is that the reviewer is responsible for attempting to explain divergent findings from a set of primary studies.

Table 3

Outcomes Arranged by Author & Type

	# Citations/ # Studies	CONTACT # Citations	CONTACT +	CONTACT 0 and -	CONTACT Success	KNOWLEDGE # Citations	KNOWLEDGE +	KNOWLEDGE 0 and -	KNOWLEDGE Success	CONTACT PLUS KNOWLEDGE # Citations	CONTACT PLUS KNOWLEDGE +	CONTACT PLUS KNOWLEDGE 0 and -	CONTACT PLUS KNOWLEDGE Success	VICARIOUS EXPERIENCE # Citations	VICARIOUS EXPERIENCE +	VICARIOUS EXPERIENCE 0 and -	VICARIOUS EXPERIENCE Success	OTHERS/NOT IDENTIFIED +	OTHERS/NOT IDENTIFIED 0 and -
Anthony (1972)	30/27	13	4	9	No	8	0	8	No	13	13	0	Yes	0	0	0		0	0
Haddle (1974)	31/29	6	3	3	No	8	1	7	No	15	12	3	Yes[a]	0	0	0		0	0
Sandler & Robinson (1978)	36/35	28	14	14	Yes[a]	6	0	6	/	2	1	1	/	0	0	0		0	0
Donaldson (1980)	24/23	13	9	4	Yes[a]	12	7	5	Yes[a]	0	0	0		2	1	1	Yes	0	0
Westwood et al. (1981)	29/27	14	10	4	Inc.	7	3	4	Inc	0	0	0		8	6	2	Inc.		
Towner (1984)	47/47	14	10	4	Inc.	6	3	3	Inc	13	7	6	Inc	9	4	5	Inc	5	[illegible]
Horne (1985)	69/68	25	15	10	Inc.	11	7	4	Inc.	10	8	2	Yes[a]	7	1	6	/	1[illegible][b]	4
Johannsen (1969)	12/10	9	8	1	Inc.	3	2	1	Inc	0	0	0		0	0	0		0	0
Rabkin (1972)	25/24	3	2	1	X	7	4	3	Inc.	15	14	1	Yes	0	0	0		0	0
Alexander & Strain (1978)	6/5	0	0	0		3	3	0	/	3	3	0	/	0	0	0		0	0
Pulton (1976)	7/6	3	0	3	X	4	1	3	/	0	0	0		1	1	0	Yes	0	0
Segal (1978)	5/4	3	3	0	/	0	0	0		2	2	0	Yes	0	0	0		0	0
Harth (1979)	10/10	5	2	3	No	3	0	3	No	1	1	0	Yes	0	0	0		1	1
Horne (1979)	15/14	0	0	0		0	0	0		9	9	0		0	0	0		[illegible]	[illegible][c]
Chubon (1982)	27/27	0	0	0		3	2	1	/	11	10	1		0	0	0		5	[illegible]

Note "Success" as judged by reviewers. An x means the author(s) did not make an explicit statement about success.

[a]Qualified conclusion.
[b]Includes cooperative learning and peer teaching studies.
[c]For these studies, the author did not indicate whether the findings were positive or negative.

In few of the 15 reviews was there evidence of systematic analysis of factors that might account for contradictory findings among primary studies or consideration of such findings as an opportunity to advance knowledge about attitude modification. In three reviews (Alexander & Strain, 1978; Horne, 1979; Segal, 1978) only successful studies were cited, so discrepancies could not be examined. Of the remaining nine reviews, however, only three provided explanations for differing findings.

Towner's (1984) examination of non-significant and negative findings was the most adequate attempt to explain discrepant findings. She discussed the covariation of outcomes with methodological soundness and theoretical underpinnings and concluded that unsuccessful results could not be attributed "to any single factor" (p. 252). She suggested that poor reports of studies, weak methodologies, and failure to ground interventions in theory inhibited the drawing of conclusions about the effectiveness of particular attitude change strategies.

Although they reviewed fewer studies than Towner (1984) and their analyses of contradictory findings were less thorough, Anthony (1972) and Donaldson (1980) were less cautious in their explanations of these findings. Anthony's discussion pertained only to differences among studies investigating contact as a change strategy, since studies cited for the knowledge change technique and the contact plus knowledge technique were considered either all successful or all unsuccessful. He noted that in correlational studies of the relationship between reported contact and attitude toward the disabled, there was a tendency to find positive associations; whereas in experimental-type studies in which contact was the independent variable, the findings were frequently nonsignificant or negative. Anthony suggested that this might have been due to the effects of self-reports of contact. Such estimations, he argued, are highly subjective, with the definition of "contact" usually left up to the individual subject. He questioned the validity of contact as an independent variable in these studies, in contrast to studies in which the amount or type of contact was controlled by the researcher.

Donaldson (1980) did discuss discrepant findings. In regard to contact as a change treatment, she observed that studies tended to yield successful results when contact was "structured" to expose subjects to disabled persons who behaved in a non-stereotypic manner. However, no explanation was given for the positive results from some studies in which "unstructured" contact was used.

Donaldson (1980) also addressed the inconsistencies in findings among studies in which information was used as the change agent. Using as her frame of reference a Lewinian model, in which discomfort reduction is posited as the mechanism for explaining modified attitudes, she noted that all of the successful studies attempted to "put subjects at greater ease through verbal messages or subtly sanctioned staring" (p. 510). Reduction of discomfort, she maintained, was not a feature of the unsuccessful studies. The unsuccessful studies cited were Cole (1971), Forader (1970), Granofsky (1956), Wallston, Blanton, Robinson, and Pollchinck (1972), and Wyrick (1968). It is difficult to determine whether discomfort reduction was or was not a feature of several of these studies. The descriptions of Cole and Granofsky's studies in Dissertation Abstracts, which Donaldson cited, are too brief for that purpose. The reference to Wyrick's thesis was based on a single sentence in Evans (1976) which stated, "Wyrick (1968) assessed the effects of a course in rehabilitation psychology on the attitudes of nondisabled undergraduates and found no significant change" (p. 573). Wallston et al. (1972) was not available to us.

In addition, Donaldson (1980) attempted to account for the differences in findings between the two simulation studies she reviewed. She noted that the successful intervention permitted subjects in wheelchairs to observe the reactions of nondisabled persons to persons (themselves) who were apparently disabled. During the unsuccessful simulation, the nondisabled persons were obviously simulating disabilities (e.g., wearing a blindfold). Donaldson contended that it was the difference in the subjects' opportunity to observe the reactions of nondisabled persons to them as handicapped rather than as role players that resulted in the discrepant findings, although the reason for the effect was not discussed.

Characterizing results. Whether there are discrepancies in findings or not, the reviewer must characterize and summarize results as a basis for drawing conclusions about the effectiveness of particular types of interventions. Towner (1984) clearly used statistical significance. The methods of analysis used for that purpose were not described in any of the other reviews. It seems reasonable to infer from the contexts of the reviews, however, that Anthony (1972), Haddle (1974), Segal (1978), and, perhaps, Rabkin (1972) employed the box-score or voting method, in which statistically significant and nonsignificant results are summed, to cumulate findings.

Comments. A number of deficiencies in the analysis of data seriously threaten the quality of most of the reviews. Information pertaining to relationships among sample and intervention characteristics was frequently lost because of the general tendency to consider the findings of complex studies in simple treatment-outcome terms. Furthermore, with several exceptions, there was a lack of attention to internal validity, and poorly and well-designed studies were accorded equal status in the discussion of interventions. Poorly designed and inadequately reported assessment of attitudes also escaped attention in most reviews. In fact, in several reviews, no mention was made of the means used in primary studies to assess attitudes. Few reviewers attempted to explain contradictory findings among primary studies, and in no review was this done in a systematic manner. It appeared that the box-score or voting approach was used to arrive at an overall judgment of the efficacy of strategies in several of the reviews.

Integrating and Reporting the Findings

If, as Jackson (1980) suggested, the "methodology of primary research" can be "used to conceptualize the methodology of integrated reviews" (p. 442), then the final section of a review is analogous to the "Discussion and Conclusions" section in a report of a primary study. It is expected that here the results of the study will be discussed in terms of the original hypotheses or questions, and practical and theoretical implications drawn from these results. Here, also, the limitations of the study are to be identified and recommendations for further research made.

The headings for final sections in 11 of the 15 reviews clearly identified them as summaries or conclusions. In fact, the term, "conclusion", was included in the headings of final sections in Anthony (1972), Horne (1979), Johannsen (1969), and Segal (1978), while "summary" was used in Haddle (1974), Harth (1973), Johannsen (1969), Rabkin (1972), and Towner (1984). The word, "implications", was used in the headings of concluding sections in Anthony (1972), Donaldson (1980), and Westwood et al. (1981). One brief review (Alexander & Strain, 1978) contained no final conclusions section; one (Pulton, 1976) did not summarize the research in the final section, but proposed an attitude change strategy; and, Sandler & Robinson (1981) made recommendations that were not based on their own review of primary studies. Conclusions as to the effectiveness of change strategies were drawn in most of the 15 reviews, even though often briefly stated. They are summarized in Table 4.

The practical implications of the conclusions drawn by the reviewers are open to question. Most reviewers who identified effective strategies for modifying attitudes toward the disabled did so only in general terms. Typical of these reviewers were Anthony (1972), Harth (1973), Horne (1979), and Segal (1978). Anthony, for example, concluded that "attitudes of nondisabled . . . can be influenced positively by providing . . . an experience which includes contact with disabled persons and information about the disability" (p. 123). In a similar vein, Harth proposed that "rather direct, well organized procedures are required" for "bringing about significant positive changes in attitudes" (1973, p. 161).

The exceptions were Donaldson (1980) and Sandler and Robinson (1981) who offered specific suggestions that would be useful, though not necessarily sufficient, to readers seeking information on which to base attitude change programs. The remaining two reviewers (Haddle, 1974; Towner, 1984) indicated that the evidence was too inconclusive to support the recommendation of any intervention as most effective for enhancing attitudes toward disabled persons.

In the reviews in which conclusions about the effectiveness of change strategies were not included in a concluding section, inferences were stated

Summary of Conclusions

Author	Contact	Information	Contact + Information	Vicarious/Simulation	Other Comments
Anthony (1971)	Studies with wide variety of disabled persons, no consistent changes (p. 119) Contact in and of itself does not change attitudes significantly (p. 120) May even reinforce neg attitudes (p. 123)	Regardless of how info presented, negligible effect (p. 120, 121)	Consistently favorable impact (p 121, 123) Limited research with college volunteers or trainees in helping professions Dearth of data on other age groups, nonvolunteers, and nonhelping professions (p 123)		Dearth of expt'l studies (p 120) Need to include behavioral measures (p 124) Little known about the time needed (varied in length from 6 hrs to 2 yrs, p 123)
Haddle (1974)	No substantial results with contact alone (p. 93)	Most studies produced no significant results (p. 92)	Info and contact tend to produce more significant results (p 95) But studies poorly designed (p 95) Most significant studies required extensive contact—often 40 hrs/wk (p 95)		Cites Anthony (1972) that most Ss were volunteers and college age (p 96) Most studies lacked good expt'l designs (p 96)
Donaldson (1980)	Contact per se not effective (p. 505) Structured contact, pos. change (p. 505); unstructured social or prof'l contact, results equivocal (p 505) Factors in pos. change: (1)=status (age; social, educ'l, vocational status; helping relation) (p 505), (2)=disabled don't act in stereotyped manner (p. 507) Short, structured non-stereo. experiences, short term impact (p. 511)	No causal relationship between limited info and attitude change (p 508) If info. confirms negative stereotypes, negative effect (p. 511) Studies of courses not helpful because content unspecified and confounded with contact, media exposure, instructor characteristics (p 508)		Simulation only 2 studies Can be effective if can observe reactions of nondisabled persons (p. 508)	Paucity of research. "literature contains relatively few studies" (p 505) Failure to test theories (p 529) Behavioral outcomes + long term effects need investigating (p 512)
Sandler & Robinson (1981)	Effects of contact equivocal (p. 98)	Effects assessed by few researchers (p. 99) Controlled studies needed (p. 100)	Cited 1 study that [illegible] and contact together beneficial (p 101)		
Westwood et al (1981)	"Results tend to be inconclusive at best" (p 221)	"Educational programs" produced equivocal results, results "inconclusive at best" Program content unknown (p. 221)		Simulation "results are inconsistent" (p 222)	Contact "earlier" studies didn't produce change (p 221) Need to study various media (p. 221)
Towner (1984)	Various approaches with different populations equally effective similar techniques with different disability groups yielded discouraging and contradictory findings Positive and negative [illegible] addition to non-significant results, from contact and information. Modes of presentation, including simulation, didn't produce significant differences (pp 249-51)				Instrumentation seemed to have no effect (p 251) Generally paper & pencil techniques (p. 224) Few attempts to address the complexity of attitudes (p 224) Most reported no theoretical base (p 224) Findings contaminated by methodological faults (p. 251)
Horne (1985)	Results inconclusive (pp 156, 163-4) Limited # of studies, differences in treatments, methodology, & disabilities (p. 156-7)	Inconsistent results (pp 163-4)	More successful, with prof Is, but still not consistent (p 151)	Role playing children's books—results mixed (p 173)	Interactions rarely assessed (p. 152, 166) Do immediate posttest results hold up? (p 155)
Fulton (1976)	Contact a factor but not with all social settings	Results with info equivocal (p 86)		Role play has potential (p 86-7)	Very few experiments that have positively changed attitudes toward physically stigmatized (p 35)
Johannsen (1969)	Equivocal results (p 224)	Not much is known about relative effectiveness of techniques (p 224)			
Rabkin (1972)		Results conflict (p. 167)	Contact with patients and formal instruction effective (p 166)		Questionnaires, few efforts to measure changes in behavior (p 163)
Harth (1973)	Social contact not enough (p. 161)	More direct the procedure, the better the results (p. 160)	Effectiveness of knowledge through direct contact supported (p 160)		No consistent line of research no theoretical base (pp. 161-2)
Alexander & Strain (1978)*					
Segal (1978)	Can reinforce neg attitudes if bizarre behavior (p. 215)		"Educated contact" necessary (pp 215 216)		
Horne (1979)			[illegible]ed info and contact (p [illegible]3)		
Chubon (1982)	Some indication that prof'l experience negatively related (p. 28)				Lack of definition of terms (p 27) Methodology poor—lack of theory, standardized definitions, refined measurement devices Need to build on findings and experiences of other researchers (p 27)

*No conclusions based on the research in regard to methods for modifying attitudes toward disabled persons could be found in the article.

following discussions of specific interventions (Rabkin, 1972; Westwood et al., 1981), included in introductions to such discussions (Johannsen, 1969; Pulton, 1976), or sometimes embedded within the discussions of specific interventions (Horne, 1985). Suggestions for attitude change programs based on these findings were also stated as broad generalizations.

The practical significance of the recommendations in the reviews is uncertain, not only because of the general manner in which the recommendations were stated but also because the validity of the reviewers' data collection methods can be seriously questioned. In most of the reviews, such problems as the lack of representativeness of samples of primary studies, the grouping of primary studies into loosely defined categories, and the tendency to ignore complex designs threatened the usefulness of the recommendations.

Theoretical Implications

Attitude change theory might have been advanced had the theoretical bases for or implications of the interventions been examined, but few reviewers did so. In fact, research results were discussed in terms of attitude theories in the conclusions sections of only two reviews (Pulton, 1976; Towner, 1984).

Pulton (1976) utilized role conflict theory to explain attitude change following simulations of marginal disabilities in one study (Clore & Jeffrey, 1972). He suggested that nondisabled subjects who role play being a disabled person experience both inter- and intra-role conflict. Inter-role conflict occurs because nondisabled subjects who simulate disabling conditions must behave as if disabled while continuing to be nondisabled persons. Intra-role conflict occurs when significant others demand "immediate explanations" for

80

the apparent disability or when nondisabled persons expect subjects to be competent in managing their disabilities. Pulton argued that subjects who role play marginal disabilities experience more psychological strain than subjects who role play severe disabilities since marginally disabled persons have more behavioral choices available to them, and hence face more conflicts. These conflicts, he maintained, result in greater acceptance of disabled persons. Consequently, Pulton advocated "emotional role playing" as a means for enhancing attitudes toward the disabled. Pulton did not discuss the theoretical underpinnings of the studies he reviewed which employed contact and knowledge as change strategies.

Towner's (1984) discussion of attitude theory was thorough and systematic. She extrapolated elements from attitude theory to use in reviewing each of 47 studies, and she concluded that most successful approaches were applications of the attitude change theory proposed by Hovland, Janis, and Kelley (1953). This theory has been labeled in the literature (Kiesler, Collins, & Miller, 1969; Insko, 1967) as "stimulus response" or "behavioristic" theory. Because Towner found so few methodologically sound studies among the 47 she examined, she argued that her findings that success appeared to be related to the number of theoretical elements employed in a study should be considered tentative.

One other reviewer (Donaldson, 1980) cited specific attitude theories to explain the efficacy of particular interventions as change strategies. Although Donaldson made no mention of theory in her conclusions section, she included a section in her review entitled, "Theoretical Models". Here, Donaldson suggested that together Lewin's theory of attitude change and the theory described in Communication and Persuasion (Hovland et al., 1953) best explained attitude change in six of the studies she examined. She failed to explain why subjects in five studies that apparently did not fit this model also experienced enhanced attitudes toward the disabled. And, as noted earlier (p. [46]), it was difficult to determine that elements of these theories were absent from the unsuccessful studies she cited.

Qualifying Conclusions

Of the 15 reviewers, only Anthony (1972), Chubon (1982), and Towner (1984) qualified their conclusions by reference to the primary studies they reviewed. Both Chubon (1982) and Towner (1984) cited weaknesses in the design and instrumentation of primary studies as limiting factors in the generalizability of their findings. Due to the general lack of methodological soundness in the studies reviewed, Chubon referred to his findings as "soft or preliminary" (1982, p. 29) and Towner characterized her findings as "contaminated" (1984, p. 251). Anthony (1972) suggested that his conclusions pertaining to contact and information were limited because of the restrictive nature of the samples used in the primary research studies. He noted that "college students who volunteered" and "trainees in the helping professions" (p. 123) comprised most samples. His conclusion that contact plus information was an effective attitude change strategy could not be generalized beyond these groups.

Research Recommendations

Recommendations for future research were included in the results of seven of the 15 reviews. These recommendations varied from simple pleas for "successful experiments" (Pulton, 1976, p. 87) and for researchers to "coalesce the disparate findings [of primary studies] and to build upon the work of one another" (Chubon, 1982, p. 29), to multiple suggestions for designing studies and selecting and reporting the instruments used to assess attitudes (Towner, 1984).

The most frequently mentioned recommendation was to base research on theory and to design research to test competing theories of attitude change (Harth, 1973; Donaldson, 1980; Towner, 1984). Furthermore, Donaldson and Towner suggested that future research should examine the relationship between attitude and behavior, as well as address the question of the long term effects of attitude change interventions. Donaldson also proposed that future studies should be directed toward investigating social forces that encourage the devaluation of disabled persons.

Research designed to assess the effectiveness of a variety of media for presenting persuasive communications was advocated by Donaldson (1980) and Johannsen (1969). Donaldson also recommended research to explore the differential effects of live versus media presentations of nondisabled persons interacting with disabled persons in a nonstereotypic and positive manner in studies of desensitization and modeling as attitude change approaches. Additionally, she called for research to identify and analyze the factors that contribute to positive attitudes in university courses, integrated settings, and disability simulations.

Alexander and Strain's (1978) recommendations for future research reflected their interest in modifying teachers' attitudes toward disabled students and mainstreaming. They favored studies designed to identify factors in preservice and inservice teacher education programs that would have positive affects on attitudes toward both disabled children and their integration into regular classes.

Horne (1985) reiterated Donaldson (1980) and Towner's (1984) suggestion that in the future, researchers should investigate the long term effects of attitude change strategies, and she agreed with their recommendation that research should be directed toward exploring the relationship between scores on attitude measures and behavioral changes. Additionally, Horne argued that the relationship between subject characteristics (such as predisposition toward disabled persons, age, and sex) and the effectiveness of particular interventions should be examined, along with strategies for modifying attitudes in specific situations (e.g., the attitudes of a particular group toward a certain disability).

The focus of the recommendations for future research was on primary studies. No reviewer proposed additional or alternative types of reviews.

Comments

Concluding sections in most reviews did not hold up well when examined against criteria for preparing conclusions in reports of primary studies. Less than half of the reviewers referred specifically to their own findings in drawing conclusions. Most conclusions concerning the effectiveness of particular intervention strategies were stated in broad generalizations that were of doubtful utility. Few reviewers acknowledged the limitations placed on their findings by poorly designed primary studies, and recommendations for future research were found in only half of the reviews. Similarly, the reviews made little contribution to an increased understanding of attitude change theory.

Discussion

Seven full and eight brief reviews of primary research on the modification of attitudes toward disabled persons were located by a comprehensive search of the literature. These reviews were examined for methodological soundness and for their contribution to practical knowledge and attitude change theory using questions developed from the work of Jackson (1978, 1980) and others, with the primary research process as a model.

Although building on prior works is a standard approach for advancing knowledge in a field, most reviewers ignored previous, but relevant, reviews. As a consequence, the reviewers did not draw on the findings of earlier reviews; neither did they use inadequacies in prior reviews as a means for improving the quality of their work. Most of the reviews were presented as though they were unique in the literature.

The possibility of sampling bias was present in each review. Methods of locating primary studies were seldom reported; moreover, the limited reference lists of studies and the small number of primary studies that were cited in more than one review cast serious doubt on the representativeness of the samples. Consequently, the generalizability of the findings of the reviews was dubious.

Many of the primary studies reviewed were low in treatment and internal validity; and, although this was mentioned in several reviews, it could not be determined how or if such studies were weighted when decisions concerning the effectiveness of particular interventions were reached. It seems apparent, given that lack of discussion, that treatment and internal validity were not explicitly considered in most reviews. Including poorly designed and executed studies in the reviews without examining the association between design quality and outcomes compromised the integrity of interpretations and conclusions.

A number of significant methodological weaknesses were found in most of the reviews. Primary studies were placed into loosely defined intervention categories, with the result that important differences in sample and intervention characteristics were frequently disregarded. Narrative reports of programs and reviews of literature were cited as though they were primary studies. In several reviews, primary studies were misinterpreted and irrelevant studies were cited. Furthermore, there was a general tendency to report the findings of complex primary studies in simple treatment-outcome terms and, in some cases, to report only partial results. Moreover, even the statistical significance of findings was not presented in most reviews, and none reported an effect size metric independent of sample size. And, studies which failed either to identify the dependent variable or to provide reliability or validity data for project developed instruments appeared to be accepted uncritically. Contradictory findings were not adequately analyzed or explained in most reviews. In view of these concerns, the findings of the reviews should be treated with caution.

Important deficiencies were also noted in the conclusions of most reviews. The practical value of the conclusions is questionable, as interventions judged to be effective were most often described in broad generalizations. Few reviewers attempted to examine the theoretical underpinnings of change strategies and little contribution to attitude theory was made by the reviews. Additionally, few reviewers acknowledged limitations to the generalizability of their findings, even though sample bias was a threat in all reviews. Specific recommendations for future research were rare.

Implications for the Present Study

This review of reviews has important implications for the quantitative, integrative review reported on the following pages. Many of the problems we identified appear to be common to past narrative integrative reviews of the literature (Jackson, 1980). Particularly germane to the present study are the potential sampling bias and other methodological problems common to most of the 15 reviews examined above.

Sampling bias was a threat to the generalizability of the findings of each review. The small number of primary studies included in the reviews, the slight overlap of samples of primary studies, and the failure of reviewers to describe how samples were selected suggest the possibility that the samples of studies for the reviews were not representative of the population of primary works reporting investigations of interventions for modifying attitudes toward the disabled. This finding indicated that a careful sampling approach or a comprehensive literature search should be considered as an integral part of additional reviews and that future reviewers should report literature search procedures thoroughly.

The ineffectiveness of the reviewers' practice of grouping interventions under broad descriptors in a manner that failed to take into account specific sample and intervention characteristics indicates the need for more precise coding and reporting of design and sample characteristics. Donaldson's (1980) recommendation that future research be directed toward identifying specific factors in university courses, integrated settings, and disability simulations that contribute to enhanced attitudes toward the disabled is consistent with this recommendation.

Suggestions for sub-categories within generic groupings of interventions were found in several of the reviews. Towner (1984), for example, organized and analyzed primary studies according to general sample characteristics, and her description of attitude change techniques had implications for coding interventions in a more accurate manner. The conclusion in several reviews that contact was an effective change strategy only if certain conditions were met (Donaldson, 1980; Harth, 1973; Segal, 1978; Westwood et al., 1981), and that the efficacy of information as a means for modifying attitudes was related to the way in which it was presented (Sandler & Robinson, 1981), also implied subcategories for obtaining detailed descriptions of interventions. And, Anthony's (1972) conclusion that length of treatment in an attitude change study may have a bearing upon the results drew attention to the need to include this factor in describing interventions.

The failure of most reviewers to explore relationships between quality of research design and results in the primary studies they examined implied that data for our review should be collected and organized so that associations between design quality and outcomes could be examined. Similarly, the tendency in several reviews to ignore poorly designed and inadequately reported attitude assessments in primary studies suggested that comparing results of studies differing in quality of instrumentation would be an appropriate strategy in the present study. Examining quality of design and instrumentation might provide insights into the reasons for contradictory findings. And, the lack of attention in most reviews to the theoretical underpinnings of specific change strategies indicated that study-theory relationships should be addressed in the present study. Most importantly, the absence of a comprehensive, systematic, integrative review of the research suggested the need for such an effort to determine whether the generally indefinite conclusions about the effectiveness of types of interventions for modifying attitudes accurately reflect the state of available research knowledge.

CHAPTER 3

HOW THE REVIEW WAS CONDUCTED

The review of literature has taken on greater status as a full-fledged scholarly research activity in the years since Glass (1976) first advocated the conducting of meta-analyses. Earlier, Feldman (1971) had expressed concern about the "half-hearted commitment" of behavioral scientists to reviewing and integrating the research literature. He suggested that the lack of more intense attention to conducting reviews "might account in part for the relatively unimpressive degree of cumulative knowledge in many fields" (p. 86). In his landmark presentation, Glass (1976) pointed out that scholarly values and attitudes which emphasized original research militated against the commitment of effort that is necessary to extract knowledge from the "staggering number of individual studies" now available.

Careful analyses of the literature, not superficial synopses, are needed. And, in fact, Glass argued, "a good review is the intellectual equivalent of original research . . . ; we need more scholarly effort concentrated on the problem of finding the knowledge that lies untapped in completed research studies" (p. 4). However, as Jackson (1980) and Light and Pillemer (1984, pp. 3-4) have noted, the traditional narrative revie of research has tended to be subjective, lacking in scientifically sound procedures, and inefficient, especially when findings are to be extracted and summarized from a large number of studies, for example, 30 or more. If reviews of literature are to meet scholarly criteria such as are applied to original research, it is crucial that a method of reviewing be used that overcomes the shortcomings of the "traditional literary format"--which often results in reviews that tell little more than the directions of findings and, sometimes, their probability level (Rosenthal, 1984, p. 10).

Alternative Approaches to Integrating Primary Research

The most commonly used approach to integrating research has been the narrative review, such as those critiqued in Chapter 2. Typically, such a review is based on a small group of easily obtained research reports, gathered from fairly prominent journals and *Dissertation Abstracts*, using criteria that are not clearly specified. The reviewer usually offers a brief verbal synopsis of each report; sometimes the methodology is critiqued and the credibility of any conclusions challenged. Often the reviewer concludes that the existing research is inconclusive and calls for additional research, using better techniques and more precise methodology.

In a variation of the narrative review approach, the reviewer begins with a small group of readily available articles, but eliminates all that have striking design or analysis flaws. The findings from the remaining "acceptable" studies are summarized as the knowledge on the topic under review. Unfortunately, judgments as to what constitutes good research frequently differ from reviewer to reviewer. Also, the criteria for selecting "methodologically superior" articles are often overly restrictive, with the result that a small and frequently unrepresentative sample of articles is considered. Moreover, as Smith and Glass (1980) have pointed out, even "methodologically good" studies often result in contradictory findings, creating considerable difficulty in deciding what conclusions should be reached.

A more systematic approach to integrating the outcomes of primary research is what Light and Smith (1971) and others (e.g., Hedges & Olkin,

1980) have referred to as the "vote-count method". With this method, each relationship between a treatment and a dependent variable is tallied as positively statistically significant, negatively statistically significant, or not statistically significant. The size of the sample utilized in the particular primary research study is rarely taken into account. Since with larger sample sizes there is a greater probability of concluding that results are statistically significant, the voting method discriminates systematically against studies with small samples. In addition, use of the voting method implies incorrectly that statistical significance indicates the degree of importance of relationships. Consequently, erroneous or misleading conclusions may be drawn. As Glass (1977, p. 358) pointed out, if nine small-sample studies of a method of modifying attitudes yielded not-quite-significant results in one direction, while a tenth large sample study yielded statistically significant results in the same direction, the vote would be one for and nine against—a conclusion quite at odds with what seems sensible.

In considering how to improve on the voting method, Light and Smith (1971) concluded that ". . . progress will only come when we are able to pool, in a systematic manner, the original data from the studies" (p. 243). Unfortunately, users of this procedure must disregard any studies for which research data are not obtainable, and original data from studies are rarely easy to obtain. For example, Glass (1977) reported that Wolins (1962) wrote to 37 authors asking for the data from studies they had published in the preceding two years: Five did not reply, 21 reported that their data were irretrievable, two refused to share the results, and four sent their data too late to be useful. Our efforts to obtain only the information necessary to compute effect sizes (reported later in this chapter) verified the likely difficulties in obtaining data for secondary analyses.

The approach to integrating the results of prior research to be used in this project was first proposed by Glass (1976) as "meta-analysis". Properly implemented, the meta-analysis approach meets all of the criteria for high quality integrative reviews proposed by Jackson (1980) and referred to in Chapter 2. Briefly, in conducting a meta-analysis, the reviewer: (1) locates either all studies or a representative sample of all studies on the defined topic; (2) converts the findings of each study, regardless of study quality, to a common metric--that is, computes an effect size for each relevant finding; (3) codes the various characteristics of each study that might have affected the results (such as type of treatment, methodological quality, sample attributes, and type of dependent measure); (4) uses statistics to summarize study outcomes (effect sizes) and to examine the covariations of outcomes and study characteristics; and, (5) draws conclusions based on the results of those analyses.

Clearly, meta-analysis is not a technique, but an approach to reviewing the literature--one might say, a point of view about reviews--which emphasizes the gathering and analysis of comprehensive, systematic, quantitative data from primary research reports. In his critique of previous efforts to integrate the findings of social science research, Jackson (1980) concluded that "the meta-analytic approach is an important contribution to social science methodology. It is not a panacea, but it will often prove to be quite valuable when applied and interpreted with care" (p. 455). And Gage (1982), in discussing the past and future of educational research, referred to meta-analysis as one of the more important methodological advances in recent years.

Some educational and psychological researchers, have raised questions about the use of the meta-analysis approach (e.g., Eysenck, 1978; Gallo, 1978, Mansfield & Bussey, 1977; Shaver, 1979a, b; Simpson, 1980; Slavin, 1984). Some have questioned the results of specific meta-analyses; others have raised concerns about the meta-analysis approach per se. Most of these criticisms and cautions have been responded to in the literature (e.g., Glass, 1978, 1980; Glass & Smith, 1978; Carlberg et al., 1984), and will not be discussed here.

A major dilemma is how to capitalize on the advantages of the meta-analytic approach with large numbers of primary research reports so as to avoid the pitfalls of traditional narrative reviews, but without becoming "over-quantified" and losing touch with the subtle variations in individual studies (Light & Pillemer, 1984, Slavin, 1986, Wolf, 1986). As Jackson (1978) has pointed out, "there are usually trade-offs between the quantity of data and the quality of the data or analyses" (p. 16). How to consider individual studies or discuss critical issues in reference to specific studies, as has often been done in traditional narrative reviews, poses serious dilemmas for quantitative reviewers faced with large numbers of findings to integrate. Slavin (1986) has proposed, for example, that the reporting of individual studies should be as specific and detailed as possible, as one type of effort to preserve a positive aspect of narrative reviews.

The most important point that the concerns and questions about the meta-analytic approach have demonstrated is that meta-analysis, like specific research procedures such as random assignment, is not a fail-safe approach. If applied carelessly, many problems will occur. However, as Rosenthal (1984) noted:

> The alternative to the systematic, explicit, quantitative procedures [of meta-analysis] is even less perfect, even more likely to be applied inappropriately, and even more likely to lead us to error. There is nothing in the set of meta-analytic procedures that makes us less able to engage in creative thought. All the thoughtful and intuitive procedures of the traditional review of the literature can also be employed in a meta-analytic review. However, meta-analytic reviews go beyond the traditional reviews in the degree to which they are more systematic, more explicit, more exhaustive, and more quantitative. Because of these features, meta-analytic reviews are more likely to lead to summary statements of greater thoroughness, greater precision, and greater intersubjectivity or objectivity. (p. 17)

It is important to note that the term "meta-analysis", first proposed by Glass (1976) to refer to "the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings" (p. 3), is not a unitary concept, nor is it without ambiguity in use. Other researchers (e.g., Hedges & Olkin, 1985; Hunter, Schmidt, & Jackson, 1982; Rosenthal, 1984) were involved in the parallel development of quantitative review methods that are forms of meta-analysis as defined immediately above, but that do not necessarily fit exactly the meta-analytic steps defined earlier (p. 66). Bangert-Drowns (1986) has presented an excellent discussion of five forms of meta-analysis, including Glass's approach, other approaches developed parallel in time, and elaborations of Glass's approach.

One variation of Glass's approach utilizes the study rather than the study finding as the unit of analysis. Exploring variables that covary with outcomes is still a central concern, but poor quality studies or studies in which the treatments do not meet clearly defined criteria may be excluded from the sample. Bangert-Downs (1986) refers to this approach as "study effect meta-analysis" (p. 393), as contrasted with "Glassian meta-analysis" (p. 391).

In another approach, the "combined probability method" (Bangert-Drowns, 1986, p. 394), the analyst also uses the study as the unit of analysis. The focus, however, is on producing an average effect size and a combined probability statement for all of the studies reviewed; less attention is paid to exploring outcome variation.

A stronger inferential statistics orientation is represented by the "approximate data pooling with tests of homogeneity" approach. Here, the intent is "to approximate the pooling of the subjects from all of the studies into one large comparison" (Bangert-Drowns, 1986, p. 394). As part of that process, tests for the homogeneity of effect sizes and adjustments of effect sizes are employed.

Finally, a variation of the last approach, also aimed at determining a general estimate of treatment effect, was labeled by Bangert-Drowns (1986) as "approximate data pooling with sampling error correction" (p. 395). With this study-effect-size method, the variance in effect sizes due to sampling error variability is estimated and subtracted from the total variation to determine if the remaining variance is large enough to justify the investigation of moderator variables.

A sixth variation of meta-analysis, not mentioned by Bangert-Drowns, is Slavin's (1986) "best evidence synthesis", a combination of elements of Glassian meta-analysis and traditional narrative reviews. With this approach--which is, like Glass's, aimed at making statements about treatment effects, not at estimating a pooled sample result--the analysis of study effect sizes from the available reports that are highest in internal and external validity is to be coupled with tables that describe study characteristics.

As Bangert-Drowns (1986) has noted, the choice of a quantitative approach for conducting an integrative review should be based on the purpose for the review. Is the intent to try to determine what the available research has to say about the effectiveness of a treatment or treatments, or is it to approximate a pooled sample in order to arrive at a "generalizable estimate of a treatment effect"? Our purpose was clearly the former. For that reason, we have adopted an approach--which we prefer to call quantitative-integrative, rather than meta-analytic--which draws primarily on Glass's (1976, 1977) conception of meta-analysis and recognizes Slavin's (1986) advice to provide the reader with information on primary studies to the extent possible within the constraints of reviewing large bodies of research.

Procedures for this Review

As others (e.g., Jackson, 1978, p. 7; Cooper, 1982) have noted, the tasks involved in doing an integrative review of literature are basically the same as for primary research. To lay the basis for the study, a topic must be selected, a problem specified (as has been done in Chapter 1), and a critical review of prior related work completed (Chapter 2). Then, the study must be conducted: subjects (in a review, research reports) identified and selected, instrumentation selected or developed, data gathered and analyzed, and results and interpretations reported.

Interestingly, the one missing element in planning a review, as compared to primary research, is the design of the study to control for various threats to internal validity. Reviews of literature are by definition, post hoc efforts to draw conclusions from data already gathered under research conditions beyond the control of the investigator (reviewer). Some (e.g., 100 mm

A5

1.0 1.1 1.25 1.4 1.6 1.8 2.0 2.2 2.5 2.8 3.2 3.6 4.0 4.5 5.0 5.6

ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz1234567890

ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz
1234567890

1.0 mm

1.5 mm

2.0 mm

Shaver, 1979a, b) have argued that post hoc analyses (integrative reviews) of sets of primary studies that often were carried out without planning based on the careful analysis of prior primary research are not likely to be any more productive than the post hoc analyses of primary data about which researchers have been skeptical because of the difficulty, if not impossibility, of controlling for extraneous factors. Whether that skepticism is valid for quantitative integrative reviews is still open to question.

In any event, the description of procedures for this comprehensive integrative review of the research on modifying attitudes toward disabled persons follows the traditional format, sans Design section. With the problem statement and review of literature presented in earlier chapters, the rest of this chapter addresses the identification and selection of studies, instrumentation, data collection, and analysis.

Accessible Population

The purpose of this study was to conduct a comprehensive integrative review of the literature. That is, the target population was all English-language reports of research identifiable through an extensive search strategy conducted in this country and Canada. The intent was to obtain all such reports, to the extent possible. There was, therefore, no sampling procedure; and, as will be reported later, only a few o the identified reports could not be obtained, although some that were relevant had to be discarded because adequate information was not reported.

Because of the inclusiveness of study identification, although certainly not perfect, we consider the set of primary research reports we reviewed to be an accessible population, not a sample. As Jackson (1980, p. 453) has pointed out, the decision as to whether a set of studies is to be considered a population or a sample is a serious one, with implications for whether inferential statistics should be used to analyze the data gathered from the reports (discussed in the Analysis section of this chapter). The matter is complicated by the tendency of reviewers to want to draw conclusions about populations of phenomena (e.g., attitude change treatments) not about a population of research reports. That distinction is a fine line to draw and walk, but we shall try to do so and be circumspect in this report.

It should also be recognized, as noted above, that including all obtainable research reports in the review is not a universally accepted strategy (Bangert-Drowns, 1986). Some critics of Glass's meta-analytic approach (e.g., Eysenck, 1978) have argued that only methodologically superior studies should be included in integrative reviews, on the ground that poorly designed studies cannot yield valid and, therefore, useful information. The counter-argument is that such a restriction may frequently eliminate studies from which important information can be gained. As Glass (1977) has noted, researchers do not typically set out to perform studies that are deficient; ard, once a less than perfect study has been done, its findings should not be disregarded totally. He argued:

> Many weak studies can add up to a strong conclusion. Suppose that in a group of one hundred studies, studies 1 to 10 are weak in representative sampling but strong in other respects, studies 11 to 20 are weak in measurement but otherwise strong, studies 21 to 30 are weak in internal validity only, studies 31 to 40 are weak only in data analysis, etc. But imagine also that all 100 studies are somewhat similar in that they show a superiority of the experimental over the control group. The critic who maintains that the total collection of studies does not support strongly that conclusion of treatment efficacy is forced to invoke an explanation of multiple causality (i.e., the observed difference can be caused either by this particular measurement flaw or that particular design flaw or this particular analysis flaw or . . .). The number of multiple causes which must be invoked to counter the explanation of treatment efficacy can be embarrassingly large for even a few dozen studies. Indeed, the multiple defects explanation will soon grow into a conspiracy theory or else collapse under its own weight. Respect for parsimony and good sense demands an acceptance of a notion that imperfect studies can converge on a true conclusion. (p. 356).

Of course, it is possible that methodologically weak studies will yield biased or misleading results. For example, in a meta-analysis of research on the treatment of hyperactivity (White, Myette, Baer, & Taylor, 1982), an empirical approach was taken to determining whether including "weak" studies biased the results. Each study was classified according to well-defined criteria for methodological quality (e.g., type of control group, reliability or fakability of outcome measures, "blinding" of judges, duration of intervention), and associations between design strength and outcomes were analyzed. "Weaker" studies did yield different results than "stronger" studies. That is, when all of the studies in the set were considered, drugs appeared to be more effective in reducing the symptoms of hyperactivity than when the analysis was limited to those studies in which control groups were used, which met minimum standards of internal validity, and in which objective measures were used to select hyperactive children for the study and to measure outcomes. Consequently, more credence was placed in the results of the "stronger" studies. However, had the results been similar regardless of study quality, including all of the studies would have allowed the more complete examination of data to answer other important questions (e.g., the influence of age of child or duration of treatment).

In summary, for the review reported here, a comprehensive approach was taken to identifying and securing research reports. As will be reported, the covariation of results with methodological adequacy was examined and conclusions drawn accordingly. Use of this approach does not condone future primary research studies with weak designs, poor measurement techniques, or inappropriate analyses. In fact, the results reported in later chapters provide strong evidence for the importance of careful study design.

Definition of scope. Having stated that the target population is all English-language reports of research to modify attitudes toward disabled persons that are identifiable through the usual search mechanisms available in this country, it was still necessary to provide more specific guidelines by which those searching for reports could discern relevance so that decisions as to what reports to include would be made consistently. A statement consistent with Chapter 1 was written for the guidance of project staff (see STATEMENT OF GENERAL PURPOSE AND POPULATIONS in Appendix A). When a person was not certain whether to include or exclude a report, the opinion of another staff member was obtained. In cases where the project director was not one of the first two conferees and doubt remained, the report was brought to him for discussion and resolution of the dilemma.

In initially identifying reports, the titles of reports, and any descriptions of the studies when the titles were obtained from the reference lists of other reports or reviews, were examined for words that indicated that the intent of the research was to investigate methods for modifying attitudes toward disabled persons. In screening actual reports for inclusion, only those reports in which it was clear that the intent was to investigate (1) the modification of attitudes (2) toward persons with disabilities or handicaps (whether specific disabilities or in general) were included.

Journal articles, master's theses, doctoral dissertations, and other unpublished papers identifiable through conventional computer and hand search techniques and through the bibliographies of other reports were eligible for inclusion. Because the emphasis was on reports in English that were accessible in this country, most of the identified studies had been conducted in North America, particularly in the USA; but research from other countries was not excluded.*

Of specific interest, then, were empirical investigations of the effects of interventions, or treatments, on the attitudes of nondisabled persons toward persons with disabilities. Correlational research was excluded. For example, contact with disabled persons is one common type of intervention. However, if a researcher went into a school district, obtained information from students about the amount of their prior contact with disabled persons, and then compared mean attitude scores for reported contact-hour groups or correlated amount of reported contact with scores on an attitude toward disabled persons measure, that study did not qualify. In addition to studies with experimental and quasi-experimental designs, single-group studies that involved a planned intervention and the collection of pretest and posttest data were included, as well as static group designs used to investigate program effects. Mainstreaming studies were included if the effects on attitudes toward disabled persons had been investigated.

Research reports with any age or occupational samples were of interest, as long as the research was directed toward changing attitudes toward persons with disabilities or handicaps.

"Disabled or handicapped persons" was defined in terms of conventional special education categories, as reflected in Public Law 94-142, to include: mentally retarded, hard of hearing, deaf, speech impaired, visually handicapped, seriously emotionally disturbed (or, mentally ill),

---

*Of 705 effect sizes from our population of studies, 653, or 93 percent, were based on U.S.A. samples of subjects. Of the remaining 52 effect sizes, 15 (2%) came from Canadian samples, 8 (1%) for Australian or New Zealand samples, and 29 (4%) from 5 other countries (Ghana, Israel, Jamaica, Thailand, and Turkey).

orthopedically impaired, deaf-blind, multi-handicapped, and learning disabled, as well as general categories such as "the disabled", "the handicapped", or "physically disabled". Studies of subjects from populations such as "disadvantaged students", "disruptive students", or "slow learners" were not included.

Attitudes toward disabled or handicapped persons was the dependent variable of interest in identifying and selecting primary reports. It was recognized that, consistent with common definitions (e.g., Triandas, Adamopoulas, & Brinberg, 1984), researchers might consider attitudes (which we defined, to provide context, as "interrelated beliefs about and feelings toward an object which predispose the person to act in certain ways") as having cognitive, affective, and/or behavioral components. It was also recognized that "attitudes" might be assessed in a variety of ways, including paper-and-pencil tests with items that are cognitive-affective mixtures, assessments of changes in voluntary interactions with disabled persons, or reactions on projective-type tests. Measures which assessed only knowledge about the disabled did <u>not</u> qualify for selection, unless clearly considered by the research report author(s) to be attitude assessments; nor did measures which assessed attitudes toward mainstreaming qualify. General measures of attitudes toward children or other people were <u>not</u> included, unless specifically aimed at disabled persons or a particular type of disability, through instructions to the Ss or because of the context of the study—e.g., an attempt to change parents' attitudes toward their disabled children.

Measures such as sociometric scales, friendship choices, or observations of interactions were considered relevant only if clearly considered by the researcher(s) to be assessments of attitudes. Even if considered in the report to be attitude assessments, observational or other data were not included if the behaviors or responses of nondisabled Ss toward disabled persons, or the direction of behavioral or response change, could not be identified.

The intent in conducting a meta-analysis type of study is to obtain effect sizes to be analyzed. Typically, studies from which effect sizes cannot be obtained are discarded. In a major departure from that practice, we also selected studies if it could be determined that a result was statistically significant and the study otherwise met the guidelines. This methodological decision was consistent with a later recommendation by Slavin (1986). Our purpose in including such studies in our population was to investigate whether the statistical significance and direction of findings differed between studies for which effect size information was available and those for which it was not.

The search. An effort was made to identify and obtain every report of research in the target population, as defined above. In the quest for research reports, the search of the literature began with a computer search that included ERIC, CEC Abstracts, Dissertation Abstracts, International Index Medicus, Psychological Abstracts, and Social Science Research, using the descriptor "attitude change" with the broad descriptor, "disabilities", as well as with descriptors specific to types of disabilities such as "mental retardation" and "deaf". The computer search was updated twice during the duration of the project. Hand searches of Psychological Abstracts, Education Index, and Dissertation Abstracts International were also done. Also, the references in Attitudes and Disability: An Annotated Bibliography, 1975-1981 (Regional Rehabilitation Research Institute on Attitudinal, Legal, and

Leisure Barriers, George Washington University) were checked. In addition, the reference lists in all of the reviews cited in Chapter 2 were searched, as was the reference list in each primary research report we obtained, whether or not it was decided to include the report in our review.

Data were not kept on the yield of our various sources, but it was clear that the computer search yielded the smallest number of reports and the reference lists of primary research reports the largest number. This search outcome is consistent with the experience of the staff of Utah State University's Early Intervention Research Institute, who have conducted major meta-analyses of early intervention research. From 10 to 30% of the available primary research reports are likely to be identified via computer searches, with the majority coming from the reference lists of other reviews and primary research reports. The shortcomings of our computerized searches, as helpful as they were, also reflected our prior experience with searches that did not yield relevant reports that we knew were in the data base because, for example, one of us had authored them. And, it is consistent with the lament by Bracey in his Research column in the January, 1986 Kappan that a colleague who did an ERIC search of research on reading comprehension "could find only 46% of the articles he knew to exist" (italics in the original, p. 395). Clearly, a search strategy in which computerized indexes are the only source, or even one in which only computerized and print indexes are used, is not likely to meet the ideal of finding all available studies in the defined area of interest. This caveat is particularly noteworthy in a time when some reviewers assume that a computer search is adequate to the task.

Copies of some 667 primary research reports that were judged potentially relevant based on title and abstract or reference in a review or primary research report were obtained through a variety of sources. The journal and the ERIC microfiche collections in the Utah State University, University of British Columbia, Simon Fraser, and Western Washington University libraries were utilized. In addition, the Interlibrary Loan Department of the Utah State University library sent out 218 requests for reports for us, of which 187 (86%) were received. Included were 77 dissertations, many of which had been identified in Dissertation Abstracts International. (No dissertation abstracts were included in the review because of the limited amount of information they contain.) In addition, hard copies of 154 dissertations not available through Interlibrary Loan or from the authors were purchased from University Microfilms, Inc. (That represented a considerable expense at $25 plus $2.25 shipping per dissertation; UMI personnel were very cooperative.)

Each of the 667 primary research reports obtained was screened and 273 were judged to be relevant to the review topic. Bullock and Svyantek (1985) have underscored the importance of providing a listing of the studies which constituted the data set for a quantitative review. It is equally important to report those studies explicitly excluded from the analysis so that readers can comprehend the scope of the data set and judge its adequacy as a population or sample of studies. The 363 reports that were discarded as irrelevant for our analysis are listed with brief identifying information in Appendix D, as are the 31 discarded for lack of information.

Some comments on specific discards may help to make clear how the criteria for inclusion were applied. Two of the discarded reports dealt with attitude change, but not within the context of our review--which was a concern with how to make attitudes toward disabled persons more positive in order to enhance equality of educational, social, and economic opportunity.

In his master's thesis, English (1966) attempted to make more negative the attitudes toward blind persons of 57 undergraduate students (as compared to the attitudes of 59 control Ss) in two introductory sociology classes. The Ss listened to 3 ten-minute tape recordings represented to be: (1) comments by a sighted student, disgusted at being exploited by her blind roommate; (2) a homosexual person who justified homosexuality based on his blindness; and, (3) a blind person who expressed feelings of self-pity and resentment of sighted people. Attitude change in a negative direction was predicted for the experimental group. Analysis of pre-posttest data from the Attitude Toward Blindness Scale indicated a lack of statistical significance, although our computation of a d (based on pre-posttest mean gains and using as the standard deviation the pretest standard deviation for the experimental group pooled with the pretest and posttest standard deviations for the control group) yielded an effect size of -.19.

England's intent to produce more negative attitudes not only ran counter to our concerns about modifying attitudes, but raised a serious and obvious ethical question, which was not addressed in the thesis. We continue to be perplexed at the justification for the research, given no stated practical or theoretical reason to attempt to make college students' attitudes toward a group of persons with a disability more negative. England's methodology and silence on the ethical issue are, for example, in stark contrast to the decision by Oberle (1975) not to include a "negative contact condition", even though it might have added to the understanding of attitudes toward disabled job applicants, because to do so would be "imcompatible [sic] with the ethics of experimentation with human subjects and the overall philosophy of rehabilitation" (p. 108).

Another discarded article resembled the English (1966) thesis, in that attitudes toward a Down syndrome child were expected to become more negative as a result of group discussion of feelings about the target child (Siperstein, Bak, & Gottlieb, 1977). However, this study was justified on the grounds of better understanding what might happen as more handicapped children are integrated in regular classrooms. Also, rather than an effort to change attitudes in a negative direction through use of materials presenting negative stereotypes of persons with a disability, the Siperstein et al. study was an investigation of the effects of discussion of a situation in which a Down syndrome child appeared as academically incompetent. In any event, the intent of the study was not to investigate whether an intervention would improve attitudes toward disabled persons, and it was discarded.

A further illustration of the types of decisions made in discarding studies involves an often cited study by Gottlieb (1972) which was not included in our review. In that research, children chose whether to play a ring-toss game with a mentally retarded child under two conditions--expectation of success or lack of success (a ring toss from 3 or 12 feet) and level of reinforcement (5 cents or 50 cents for winning). Volunteering was confounded with the expectation and reinforcement variables; and, the study, justifiably, got at variables related to attitude change with voluntary contact rather than at the effects on attitudes of manipulating contact and of manipulated contact plus reinforcement (which was not manipulated vis-a-vis contact, but was a function of ring-toss performance).

As is typical in using the meta-analysis approach, a central interest was in obtaining effect sizes to use in quantitative analyses of the literature. The procedures for computing effect sizes will be described in the next section. It is, however, pertinent to a discussion of the number of articles obtained in the search to note that we wrote to authors for information when the description in their report was inadequate for effect size computations.

Addresses were sought in the 1986 National Faculty Directory, the American Psychological Association's 1984 Membership Register, and the American Educational Research Association's 1985-86 Biographical Membership Directory. If an author's address was not available in any of those sources, a letter was sent to the address on the article or, in a few cases, to the major professor for a dissertation. Multiple letters were sent for a report when there were multiple authors or when a follow-up seemed in order, e.g., to request additional information.

One hundred and forty-six letters were sent to authors to request information for 117 reports. The results are presented in Table 5. In the case of 53 studies (45%), nothing was heard. (For 74, or 51%, of the 146 letters sent, there were no responses.) For 13 reports (16 letters), the letters were returned by the Post Office as undeliverable or someone wrote to say some such thing as that the author was dead or had moved leaving no forwarding address; for three reports, we were informed that the person to whom we wrote was not the author. For 23 reports (20%), authors wrote to tell us the information we had requested was not available--a courtesy we very much appreciated, and which was acknowledged (as were all letters received) with a short thank-you letter. For 14 reports, information was sent that was different from that requested and did not permit effect size computations. And, for 14 reports, we received information that allowed the desired effect size computations.

Our 12% fruitful response rate for studies (10% of letters) is close to what has been reported by others--for example, Wolins (1962), as reported by

Table 5

Responses to Information Requests

Type of Response	Type of Report: Dissertations	Articles	Project Reports	Papers	Total
None	25/30	24/38	3/4	1/2	53(45%)/74(51%)
Author unavailable					
Letter returned by P.O.	2/2	5/8			7(6%)/10(7%)
Author dead, moved		3/3			3(3%)/3(2%)
Wrong person	2/1	/1		1/1	3(3%)/3(2%)
Info. not available	1/1	22/23	/1		23(20%)/25(17%)
Info. provided					
Incorrect	2/2	9/11	3/3		14(12%)/16(11%)
Correct	2/2	9/9	1/1	2/3	14(12%)/15(10%)
Total	34/38	72/93	7/9	4/6	117/146

Note: In each cell, number (or percent) of reports for which information was requested is to the left of the slash; the number (or percent) of letters sent to obtain information is to the right of the slash.

Glass (1977), received useful data from 5 of 37, or 13.5%, of the authors queried; Hyde (1982), as reported by Orwin and Cordray (1985), got information from 2 of 18 authors, or 11%. Yet, it was disappointing to find again how rarely data could be obtained from authors. It was particularly disappointing to receive no response at all to 51 percent of our information request letters. Despite the return of nine percent of the letters by the Post Office or someone at the receiving institution, it is likely that some of the 74 letters for which we had no response went astray. There apparently were, however, a fairly sizable number of authors who simply chose not to respond, not a very collegial reaction.

Although the information requested was obtained for only 14 of 117 reports, all of the remaining 103 reports were not discarded. In some cases, information was available to compute one or more effect sizes other than that for which additional information was requested. Also, as noted above, contrary to what is the case in most meta-analytic reviews, it was decided that valuable information would be gained by collecting data from reports for which effect sizes could not be computed but for which it could be determined whether results were statistically significant at the .05 level. As a result, only 31 reports were discarded for lack of information. They are listed in Appendix D.

The remaining 273 reports were the accessible population for the integrative review. They are listed in Appendix E, and a brief description of each study is presented in Appendix F.

Instrumentation and Data Collection

The basic meta-analytic approach involves quantifying the outcomes of primary research studies using a common metric and coding various study characteristics so that it can be determined whether outcomes covary with the treatment variable and with any other study characteristics. The classification system used to code primary studies is, therefore, fundamental to data collection and data analysis. It must be comprehensive enough to "capture" the factors which are contributing to variance among studies, but not be so complex as to make coding overly burdensome. There are at least three other major considerations in developing a coding instrument: (1) That the data be collected in a usable format; (2) that the coding instrument adequately reflect the substantive area under review; and, (3) that appropriate nontreatment study characteristics be coded.

The first consideration may seem mundane, but it is practically very important, especially if the meta-analysis is of sufficient scope that a mainframe computer will be used for analysis. In regard to format, a coding instrument developed at Utah State University's Early Intervention Research Institute for a meta-analysis of early intervention research with at-risk children (White & Casto, 1985) was of great value, as was consultation with the principle investigators. The second major consideration, that the coding instrument reflect adequately the substantive area under review (in this case, efforts to modify attitudes toward disabled persons), raised the concern that we would be able to code correctly the attitude-change interventions reported in the literature. Our prior review of research, reported in Chapter 2, helped to ensure that, as did the prior reading of a number of the primary research reports and tryouts of the instrument on research reports as it was developed. The third major consideration, that the coding instrument contain appropriate categories for identifying study characteristics that might be expected to covary with and perhaps confound treatment effects, was addressed based on the literature on research design (e.g., Campbell & Stanley, 1963; Cook & Campbell, 1979) and meta-analysis.

The result of our instrument development was a 20-page coding instrument with some 162 categories (see Appendix B), including a 2-page "prior contact" supplement with 14 categories and a 2-page "contact" supplement with 18 categories. Development of the main instrument took place over a 3-month period; revisions continued until the scoring of new reports could be accomplished reliably, with no distortion of studies to fit the categories, and no important information being left out.

Modifications. Even with the care in development and tryout, it proved to be impossible to anticipate all important coding alternatives and some changes were made after scoring was in progress. For example, we became sensitive to the fact, not commented on in other reports of meta-analyses we had read, that a positive effect size might actually reflect a negative treatment effect. That is, a treatment group could have a decline in scores, indicating more negative attitudes, from pretest to posttest, but because the control group's scores dropped even more, the effect size would be positive. That offended our sense of what the researchers were attempting to do and of what we wanted a positive effect size to indicate in our review. Consequently, we added a category to identify studies with positive effect sizes, but with no or negative mean attitude changes for a treatment group.

Another example of "in process" revisions had to do with "prior contact". After coding a number of studies, it seemed evident that we were not collecting adequate data on that important potential moderating variable. Consequently, the Prior Contact Coding Sheet was developed. By the same token, a Contact Coding Sheet was developed to obtain more information than we originally anticipated to be necessary on the extent to which contact treatments reflected variables considered theoretically important (see, e.g., Yuker et al., 1970; Yuker, 1986; Makas, 1986).

In each instance of a coding change, all studies already coded were checked against the revised category, and new scores entered as necessary. This was done for the Prior Contact and Contact Coding Sheets after all of the reports had been coded with the main coding instrument.

Coding conventions. The coding instrument, including the Prior Contact and Contact Coding Sheets, contains ten sets of categories. As is typical in meta-analyses, an extensive set of conventions for coding studies was also developed. These are presented in Appendix C.

The coding conventions begin with general instructions for the coders. Included in these directions are such items as how to use the various supplementary sheets (for identifying effect sizes, indicating that further information on a study should be requested, and making comments on the study or on the use of conventions that might later help in interpreting the results), what to do with blank spaces, how to complete the checklist at the beginning of the coding instrument, how to handle multiple reports of a single study, and how to round decimals in computing effect sizes.

Instructions for coding individual categories were also provided. The conventions pages are numbered for the convenience of the coder to indicate the set of categories being discussed, as well as the page number on the coding instrument where the category being discussed is located. For example, "A-2/CI 1-2" indicates that the page is the second page of the coding conventions for Section A of the coding instrument (A-2) and that the categories being discussed are on pages 1 and 2 of the coding instrument (CI 1-2).

In addition to the conventions for coding categories, there is a section on computing effect sizes and a section of addenda added to the conventions for clarification as the coding progressed and items of concern were raised that could not be handled easily with revisions in the coding conventions. All told there are approximately 90 pages of conventions, plus 10 pages of effect size computation information and 5 pages of addenda. A discussion of the sets of coding categories follows.

A. General Information.

As the title of this section of the coding instrument suggests, it contains categories to gather miscellaneous but important information about the studies. Categories A.4. and A.5. call for the recording of information about the year in which the report was made available and the type of report. These have been common pieces of information gathered in meta-analyses because of interest in any trends in research over time and in whether findings are related to how the author chooses, or is allowed, to make public his or her results.

The number of effect sizes for each primary report (Category A.6.a.) was recorded to provide ready information during data analysis. By the same token, each effect size was assigned an identification number for convenience of analysis.

As Jackson (1980) and Green and Hall (1984) have noted, those doing meta-analyses have tended to focus on the main effects of studies. These are what we have defined as primary effect sizes (Level 1 in Category A.6.c.). Although main effects were to be the primary focus of analysis, it was decided that in gathering data for this integrative review, we would also code information to allow the possibility of analyzing the data for interactions. Secondary effect sizes are those, then, that involve interaction effects or comparisons of treatment condition within levels of a classification independent variable (in this case, the classification variables of gender, testing, age, personality, and prior contact with disabled persons were selected because of their assumed importance in the literature). Category A.6.d. "Type of comparison" was included to separate out the basic types of primary comparisons (codes 1-4) and secondary comparisons (codes 5-15).

The next three categories, dealing with populations and whether the study was a replication, were included because of a general interest in the extent to which authors of research reports define their target and accessible populations and to which replication is a part of educational and psychological research (Shaver & Norton, 1980a, b).

B. Description of Sample(s).

A number of the characteristics of the subjects (Ss) in the samples used for primary research were deemed to be of potential importance in attempting to integrate and interpret the research literature. These included: sample size; how the sample was selected (Category B.2., based on the categories that Shaver and Norton [1980a, b] used in their reviews of the research literature); the percentage of males in the sample (because some researchers have found females to have more positive attitudes toward disabled persons and to be more amenable to attitude change interventions); the context within which the intervention or treatment occurred (e.g., was it an educational or a work context); in the case of an educational context, the educational level (age) of the students and, for university students, major (some of the majors [Category B.6.] that may seem surprising--such as philosophy--came from a perusal of the research studies in the pool to be coded); the occupation of non-student subjects; and, the Ss' prior experience with disabled persons. An additional sample attribute of interest was the country from which the subjects came (Category B.9.).

C. Treatment/Intervention.

Describing the attitude modification interventions that had been investigated was a primary concern. One important attribute of those interventions was whether they had been grounded in theory. As noted in Chapter 2, prior reviews (especially Towner, 1985) have lamented the lack of theoretical bases for attitude modification efforts. In order to determine the role of attitude change theory in the studies in our integrative review, the coding instrument included categories for identifying whether an explicit basis for the treatment could be identified and, if so, whether it was based on theory, on prior research, or on practical experience (C.1.).

Categories were also included in which to indicate the theory that explicitly or implicitly underlie the intervention (C.2.a.). Five attitude change theoretical positions were identified, based on the summations and discussions of theory by Insko (1967), Kresler, Collins, and Miller (1969), Thompson (1975), Triandis (1971), Wagner and Sherwood (1969), and Zimbardo, Ebbesen, and Maslach (1977). (Also see Watts, 1984.) The five categories of theory are: (1) Stimulus-response and behavioristic; (2) conditioning; (3) consistency-equilibrium; (4) social judgment; and, (5) functional. A category was also included for scoring a theory base that was a combination of these positions. Conventions for coding theoretical bases are included in Appendix C.

The use made of theory was also scored. The uses to be coded were: none (mentioned, but not used in presenting the design or discussing the results), a brief or well-developed discussion of the theory base, use of theory only for post hoc interpretation, or whether the coder had to infer the underlying theory because it was not mentioned in the report (C.2.b.).

Setting

A significant element of treatment--that is, the setting in which it took place, whether, for example, in a regular classroom, a research laboratory, or an institution for the mentally retarded--was also coded (C.3.).

Treatment Characteristics

Several categories were devoted to describing the intervention techniques. Based on the prior review of literature, three predominant approaches had been identified. These were: the conveying of information about disabilities and disabled persons; providing direct contact with disabled persons; and, providing vicarious disability experiences, such as through simulations. In addition, combinations of those three major techniques were used. Another technique, related to the presentation of information, was the use of persuasive communications. And, a few researchers used positive reinforcement or systematic desensitization in an effort to modify attitudes. Each of these techniques was included in Category B.4., along with codes for control and placebo groups.

If the conveying of information was coded as the treatment or a major part of the treatment, the type of information and delivery mode were coded in Categories C.1.a.(1) and C.1.a.(2). If direct contact was the treatment technique or a major part of it, the type of contact was coded in Category

B.4.b. Types of vicarious experience were coded in Category B.4.c., and types of positive reinforcement and types of persuasive messages coded in the Categories B.4.d. and B.4.e.

It should be noted that any description of the treatment in the report was used in coding, not the label attached to the treatment by the report author(s). This was particularly important in studies of contact as an attitude modification technique. As Makas (1986) has noted, some researchers (e.g., Donaldson, 1974; Sadlick & Penta, 1975) have labeled media presentations of disabled persons as "contact". We did not code those as contact treatments, but as use of persuasive messages and vicarious experience, for example, as was deemed appropriate.

Another significant aspect of an attitude change treatment is the intended attitude target. In this case, the question was attitudes toward persons with what disabilities were to be modified by the intervention? That was coded in Category B.5.

A potentially important methodological dimension of a treatment (coded in Category B.6.) is whether it was conducted by the experimenter himself or herself, by project assistants, or by regular nonproject personnel in their work setting, such as regular classroom teachers (Category B.6.). As the outcomes of treatments or interventions could covary with duration (see, e.g., Makas, 1986). Whether information on length of treatment was available was coded and, if available, the duration (the total number of hours of treatment was of particular interest) was entered (Category B.7.).

Treatment Verification

One of the major problems affecting the interpretation of primary research studies is the frequent lack of verification that the treatment was actually implemented as intended (Cook & Campbell, 1979; Hunter et al., 1982, pp. 95-96; Ladas, 1980; Shaver, 1983). Consequently, a set of categories (B.8. through B.8.e.) were devoted to treatment verification. Coding included whether the report contained any indication that treatment implementation was verified and, if so, what type of verification took place, how it was reported, and the degree of implementation that was claimed by the author(s). In addition, the coder estimated the extent to which implementation actually took place as intended; and, because of the importance accorded replication in discussions of research (Shaver, 1979a; Shaver & Norton, 1980a, b), a category was included for indicating whether the description of treatment in the report would provide an adequate basis for replication should another researcher wish to do so.

Treatment Validity

It is possible that a treatment which was executed exactly as intended might not validly represent the anticipated independent variable. That is, the experiences of the subjects may not be what was intended by the researcher. This concept of treatment validity is similar to Cook and Campbell's (1979) concept of the construct validity of presumed causes and effects. Using this concept, one asks: If the intervention operations are presumed to represent a construct with an assumed cause and effect relationship to attitudes, are there other confounding variables which would invalidate cause and effect inferences? Cook and Campbell (p. 60) refer to the Hawthorne effect as one such confounding variable. The question is whether an increase in performance was due to the treatment as conceptualized by the researcher or due to the Ss' awareness of the attention accorded them? The coding instrument contains categories for coding the potential impact of confounding variables that might covary with the independent variable, thus threatening the validity of the treatment. These potentially confounding variables (which are, along with the Hawthorne effect, the extent to which the treatment was actually implemented, the John Henry effect, treatment diffusion, Ss' dissatisfaction or resentment, novelty or disruption effects, experimenter effects and expectations, the confounding of treatment and experimenter, test by treatment interaction, and multiple treatment interference) constitute the ten subcategories of Category B.9. Based upon the coding of those subcategories, the coder made an overall judgment as to whether the general treatment validity was "excellent", "fair", or "poor" (B.9.b.).

Our subcategories for treatment validity are frequently included in lists of external validity (e.g., Bracht & Glass, 1968; Campbell & Stanley, 1963). This makes conceptual sense, as external validity--that is, the extent to which one can generalize results--depends to a large extent upon whether the treatment was validly implemented. If there are serious threats to treatment validity, the person who tries to generalize from the research results is left with the puzzling conundrum as to what it is that might be generalized to other treatment situations.

Can't Tell Option

It should be noted that contrary to what is common with meta-analysis coding systems, in the treatment validity subcategories, and elsewhere in the coding instrument, the coder was provided with "can't tell" as an option. In many meta-analyses, if study characteristics or their effects are not described in a report, coders are asked to make a reasoned judgment as to the nature of the study and any threats to validity. Sometimes coders are instructed to assume that no threat was present if there is no evidence of the threat. Bullock and Svyantek (1985) have, however, indicated the importance of coding for missing information to indicate when adequate data were not reported to code a category. That is the strategy we adopted. The coder was, however, forced to make a decision about general treatment validity, with the number of "can't tell" codings a factor in that decision (see coding convention for Category C.9.k. in Appendix C).

Mainstreaming Studies

As noted in Chapter 2, mainstreaming studies that included explicit concern with the modification of attitudes toward disabled persons were included in the accessible population of reports. As a particular type of treatment, mainstreaming was accorded a major category (B.10.) with 15 subcategories to assess the type of instruction in mainstreamed classes, any special personnel support or special training for disabled students, nondisabled students, or parents that was provided, the length of the mainstreaming treatment, and the Ss for whom attitude outcomes were assessed.

D. Dependent Measures.

A critical aspect of primary research that ought to be addressed in an integrative review of literature is the instruments used to assess the dependent variables which are of interest as outcomes. The validity of scores is a primary concern. The centrality of appropriate assessment to the drawing of conclusions about treatment effects is indicated by Cook and Campbell's (1979, pp. 60-61) decision to include whether the "proposed dependent variables . . . tap into the factors they are meant to measure" as a part of the cause-and-effect construct validity of experimental designs.

Among the prior reviews of research on modifying attitudes toward disabled persons, Towner (1985) in particular raised questions about the validity of attitude assessments, noting that many primary researchers did not even define the construct of attitude which underlay their research.

Reactivity of assessment is a concern in the validity of attitude scores (Matkin, Hafer, Wright, & Lutzker, 1981; Wilson & Putnam, 1982). Consequently, coding included not only whether test validity was discussed and the type and source of validity, if reported, but an estimate of the reactivity of the measure as well. The coder then made an overall judgment of the adequacy of instrument validity ("low", "moderate", "high"), with the convention that highly reactive measures were automatically scored no higher than "moderate".

Underlying validity, of course, are the matters of the reliability of scores, the types of instrumentation used, and how data collection and scoring were carried out (for example, were data collectors and scorers blind to treatment group membership when that was important?).

These various attributes of instrumentation and data collection were coded in Section D. of the coding instrument. Categories included, in addition to the attributes mentioned above, whether the posttest was administered immediately at the conclusion of treatment or delayed for more than a day, whether there was follow-up posttesting, and, if available, the time from intervention conclusion to posttesting.

E. Internal Validity.

How to deal with the quality of the primary research reports is a major issue raised in the literature on quantitative integrative reviews. Bangert-Drowns (1986) has provided an excellent summary of the debate over study quality. While Glass (1976, 1977) advocated that all studies be included in a quantitative review, regardless of quality, so that the covariation of study outcomes with study quality could be investigated, others (e.g., Eysenck, 1978; Mansfield & Busse, 1977) have raised questions about that strategy. In particular, the investigation of study quality-outcome covariation is not likely to be fruitful if there is a constant bias or flaw in the research in a field (Bryant & Workman, 1985, pp. 635-636; Shaver, 1979a) or if there is a lack of variability, in particular a lack of well-designed studies against which to compare the outcomes of studies with methodological flaws (Bangert-Drowns, 1986).

As noted in Chapter 2, the clear intent of this study was to follow the lead of Glass and investigate the relationship between study quality and outcomes. That was the thrust of the sets of categories, discussed above, that deal with treatment validity and the dependent measures. It was also the purpose of the set of categories dealing with internal validity. In these categories, each study was coded according to the particular design used, the method of assigning subjects to groups, and the extent to which there were threats to internal validity, based on the standard Campbell and Stanley (1963) listing, with one important addition. The total treatment validity score (Category C.9.k.) was included as a potential threat, on the grounds that it is nonsensical to discuss whether an outcome can be attributed to the treatment (i.e., to consider internal validity questions) if the treatment was not implemented adequately.

Although some reviewers using quantitative techniques add up scores on individual subcategories to obtain a total quality of internal validity or methodology score (e.g., Bullock & Svyantek, 1985), that approach was not adopted for this review. Coders were given guidelines (see the conventions in Appendix C) for arriving at a judgment of general internal validity ("low", "medium", "high") based on the understanding that internal validity is not a matter of simply adding up the category scores for individual threats, because any one threat may be fatal. That is, a study may be well-designed in most regards, receiving high scores on every category of internal validity but one, and thereby lack internal validity. For example, if the experimental groups were exposed to different histories that had clear potential for differential effects on outcomes, then internal validity is low regardless of how well other threats have been controlled. Although the coder had available a "can't tell" option for coding specific threats to internal validity, a forced choice was made as to whether the general internal validity was low, medium, or high.

F. Results.

The major concern in coding the results of the primary research studies reviewed was to record the effect size(s) for each study. In addition, whether or not the result was statistically significant at the .05 level was coded (Category F.1.). Coding also included whether the authors qualified their conclusions about treatment effectiveness in terms of possible threats such as the type of sample and design flaws (Shaver & Norton, 1980a, b) and whether the author deemed the treatment to have been effective or not (Categories F.2.a. and b.).

Effect Size Availability

In Category F.3.a., the coder indicated whether an effect size was available. As was mentioned earlier in this chapter, a positive effect size may actually reflect no change or a negative change by the experimental group that was accompanied by a greater negative change by the control group. Consequently, if an effect size was available, it was coded to indicate whether the change for the treatment group was positive, on the one hand, or nil or negative, on the other. In addition, if no effect size was available but the level of statistical significance (above or below .05) could be determined, whether each difference on a relevant outcome measure favored the treatment or the comparison group was coded.

Effect Sizes

Next, the coder entered any effect sizes that could be computed and coded information about the sources of the statistics used in computing or estimating each effect size. The procedures to be followed in selecting statistics and computing or estimating effect sizes are detailed in the COMPUTATION OF EFFECT SIZES and CONVENTIONS ADDENDA sections of the coding conventions (see Appendix C).

Our major indicator of effect size was Glass's Delta* (Glass, McGaw, & Smith, 1981), which we labeled *D*. To compute a *D*, the difference between the experimental mean and the control group mean is divided by a standard deviation, if available, which is free of treatment effects. *D* is in contradistinction to Cohen's (1977) *d* in which a standard deviation based on the within group variance for all groups pooled is used for the

*Some reviewers treat Delta as if it were *the* effect size, rather than one type of effect size (see, e.g., Walberg, 1986, p. 216). Walberg also seems to consider Deltas and correlation coefficients to be equivalent (see his Table 7.2, pp. 218-19). However, they are not the same metric. For example, a point biserial coefficient will be considerably less than the Delta for the same result (if *D* = .5, r_{pb} = .24; if *D* = 1, r_{pb} = .45; if *D* = 2, r_{pb} = .71; see Appendix G). And, correlation coefficients from correlational studies would have very different meanings than either *D*s or point-biserial coefficients from experimental designs.

standardization of mean differences. Glass, McGaw, and Smith recommend use of the control group standard deviation. As our purpose was to obtain the most stable estimate of variance in the untreated population, we extended Glass's Delta by pooling the variances available for untreated groups--including treatment group pretest and control group pre and posttest variances--to obtain the standard deviation by which the difference between means was standardized.

Standard deviations. Raw score standard deviations, or estimates of those standard deviations, were used in computing Ds. When the only standard deviation available in a report was from an analysis of covariance or was a standard deviation for gain scores, an unadjusted standard deviation was estimated (Glass et al., 1981; McGaw & Glass, 1980).

As Kulik and Kulik (1986) have pointed out, the decision to use raw score standard deviations for computing effect sizes, rather than using standard deviations from which major sources of variation have been removed, for example, by covariance, regression analysis, or blocking--is not a trivial matter. Effect sizes computed with adjusted, or reduced, standard deviations (called "operative" effect sizes) will "vary not only as a function of size of the raw-score treatment effect but also as a function of the experimental design used to investigate the effect" (p. 7). Although operative effect sizes are useful in statistical power analyses, they are not directly comparable from study to study unless the same research design and analysis were used in each. However, effect sizes computed with raw score standard deviations (called "interpretable" effect sizes) can be interpreted along a common scale because they are conceptually equivalent to one another.

Estimating Ds. When the means or the standard deviation for computing a D was not available, but the result from a test of significance, such as an F-ratio or t-ratio, was available, D was estimated based on procedures spelled out in Glass et al. (1981).

Variations on D computations. Generally, the procedures followed in selecting statistics and computing or estimating Ds were those suggested by Glass (1977; Glass et al., 1981) and elaborated in worksheets prepared by Karl White, of Utah State University's Early Intervention Research Institute, for use in his own meta-analytic work and in conducting workshops on meta-analysis. However, how to code the information from a Solomon four-group design had not been addressed in the literature, to the best of our knowledge. The Solomon four-group design is, of course, a combination of two designs--a pretest-posttest, control group design and a posttest-only, control group design. It was tempting, on those grounds, to obtain two Ds from each such design. However, that would have compounded the problem of nonindependent multiple Ds from individual studies. Consequently, we computed the two Ds and pooled them (weighting by n, if the design was not balanced) to obtain one effect size (see Appendix C, CONVENTIONS ADDENDA, #7).

Another procedural variation had to do with results reported as percentages. Glass (1977, pp. 369-70; Glass et al., 1981, pp. 35, 136-146) has recommended the use of the probit transformation when data are reported in terms of dichotomies. However, this transformation (which involves taking the difference between the standard normal deviates for the percentages in the two groups being compared), is subject to technical problems (Glass et al., 1981, p. 138) and in practice often yields suspiciously high estimates

123

of D. An alternative procedure is to cast the proportions for the two groups in a two-by-two table, compute a Phi coefficient (or compute it from the chi-square value for such a table, if reported), and then use that coefficient to estimate D (see Appendix C, COMPUTATION OF EFFECT SIZES). The Phi coefficient procedure appeared to yield more reasonable estimates than did probit transformations, and it was used in the few instances (12 out of the 644 Ds used in our main analysis) in which percentages (dichotomies) were reported.

Source of D. When a D could be computed, the source (whether calculated or, for example, estimated from a t-ratio), the scale of mean differences (that is, whether between raw gain scores--the preferred unit--posttest differences, or covariance adjusted means), and type of the standard deviation (whether a control group standard deviation, a pooled standard deviation, or an estimated standard deviation) were coded (Categories F.3.b.2], 3], and 4]).

Correlations. Effect sizes were also coded as correlations because, for some people, an independent-dependent variable relationship expressed in terms of a point biserial coefficient (especially if squared) is more amenable to interpretation than when expressed as a standardized mean difference. When the point biserial coefficient was not reported, which was typical, D was computed first and the D converted to a point biserial coefficient (see the Correlation Coefficients section of the COMPUTATION OF EFFECT SIZES section of the conventions in Appendix C).

In analyzing the data and reporting the results, we focused almost exclusively on Ds: they are commonly used and readily interpretable, and the magnitude of the analysis made it unfeasible to analyze redundant indicators of effect size. A table for transforming Ds to point biserial coefficients is available in Appendix G for those who think more readily in terms of correlations and variance explained.

For interaction effect sizes (which were secondary effect sizes), a _D_ could not be computed. Eta^2, usually available or computable from ANOVA tables, was used as an appropriate indicator of effect size.

Variance ratios. Researchers often focus in their analyses on central tendencies (usually means), disregarding the possibility that treatments may have affected variability in scores. It did seem possible that attitude modification treatments would, in some cases, increase or decrease dispersion because of differential effects on people, depending, for example, on their pretreatment attitudes (see, e.g., Amir, 1969). Consequently, variance ratios were computed for any primary effect sizes for which variances were available.

Median effect sizes. Considerable concern has been expressed in the literature about Glass's strategy (see, e.g., Glass et al., 1981) of basing meta-analyses on effect sizes rather than on studies, when individual studies yield multiple effect sizes (see Bangert-Drowns, 1986, for an excellent synopsis of the debate over the nonindependence of multiple effect sizes from individual studies). Some review methodologists recommend obtaining a single effect size for each study or for each type of outcome in each study. Rosenthal (1984), for example, has recommended the use of the median effect size as a stable although conservative estimate of the effect size for an individual study.

Rosenthal's recommendation confirmed our decision to compute a median overall _D_ for the posttest outcomes of each study and for each set of follow-up posttest outcomes that was available. Median _D_'s were computed only for primary effect sizes. If within-study replications were reported, medians were computed separately for each replication. In addition, a median D was computed for each type of assessment (Category D.6.) used in the study, for the posttest and any follow-up posttest outcomes, and for within-group replications.

As will be indicated in our discussion of results, use of the median values in our analyses was neither easy nor particularly fruitful. Use of a median effect size is particularly feasible when reviewing studies which do not have multiple types of outcomes or assessments, intra-study replications, and/or repeated posttesting. Our set included all. Of particular concern were the studies with multiple types of assessment: they yielded multiple medians, which made analysis difficult; but the alternative, the overall median D, was of questionable meaning. Consequently, median Ds were used in the analysis primarily as a check on the results obtained with individual Ds. That is, some comparisons were made to determine if the results would have been strikingly different using median rather than individual Ds. The lesson gained was that despite the recommendation to use median Ds in analyses, much work needs to be done on how to best compute and identify them for easier use in analyses that involve a large number of studies, with many having multiple assessments.

G. Supplemental Information.

As discussed above, treatment verification and treatment validity categories were included in the coding instrument to assess the extent to which the independent variable was executed as intended and with construct validity. When the conveying of information about disabilities and persons with disabilities is the technique for modifying attitudes under investigation, a further check on the treatment should be undertaken. That is, it should be determined whether the subjects' knowledge of the information did increase. Consequently, information treatment studies were coded according to whether data on information gain were reported, how those data were reported, and the conclusions that could be drawn in terms of the amount of information gain (Categories G.1.a., b., c.). In addition, an effect size was computed for the information assessment, with a median effect size entered if there was more than one measure of knowledge. This information was entered in Category G.1.d.

In addition, it became clear during the coding that we were dealing with three major types of studies that should be distinguished: course evaluations, program evaluations (e.g., evaluations of graduate programs in rehabilitation therapy), and experimental treatments. Also, while most classroom mainstreaming studies compared mainstreamed versus nonmainstreamed students, it also was anticipated that some might include the analysis of data to determine the effects on students who had been exposed to mainstreaming for varying amounts of time. These types of studies were coded in Category G.2.

H. Coding Summary.

The number of minutes spent coding each study was recorded, as well as who did the coding.

I. Prior Contact Coding Sheet.

As noted above, it was decided that it would be important to have more information about assessments of amounts and types of Ss' prior contact with disabled persons than was obtained with Category B.8. That information was obtained using the Prior Contact Coding Sheet. Included was a code (Category I.5.) to indicate whether prior contact was implicit (e.g., psychiatric nurses who had been working in mental hospital wards prior to an inservice course) even though the researcher did not deal explicitly with prior contact as a moderating variable. Also coded were how prior contact was assessed, including the type of prior contact setting and the definition of degree of contact, if available; the uses made of any prior contact information; and, the type of disability with which prior contact had occurred. In addition, the minutes spent in coding this additional information was entered.

J. Contact Coding Sheet.

In the discussion of prior reviews of the literature on modifying attitudes toward disabled persons in Chapter 2 (see Table 4 in particular), we indicated that there is consensus that contact per se is not likely to produce more positive attitudes toward persons with disabilities and may even reinforce negative attitudes. The elements in contact that are likely to affect whether contact results in more positive attitudes were alluded to by reviewers such as Donaldson (1980), Harth (1973), Segal (1978), and Westwood et al. (1981). The summary of those factors by Yuker et al. (1970) is frequently cited in the literature, and Yuker (1986) has provided a more recent synopsis. The theoretical bases for the enumeration of factors related to the effects of contact on attitudes toward disabled persons are typically found in Allport (1954) and Amir (1969). In a recent paper not available to us until the final report was in preparation, Makas (1986) has discussed those factors again. (She presents a well-developed argument that the inconsistent results of the disabilities-contact attitude-modification research are not due to inadequacies in the theory but to methodological inadequacies in the research studies.)

A contact coding sheet was developed based on the prior discussions of the theoretically important dimensions of contact in modifying attitudes. Categories were included to gather information on the relative status of the nondisabled Ss and the disabled persons with whom they had contact, including age, educational-vocational discrepancies, and helping relationships (Categories J.5.a.-d.). An overall judgment of status was coded, based on the prior subcategories and any indicators of social status available in the report.

Aspects of the type of contact were also coded. Included were the extent to which contact was voluntary, the extent of intimacy involved, the extent to which cooperation and/or competition was involved in interactions, the extent to which the interaction itself produced reinforcement and/or the extent to which the nondisabled Ss were reinforced for interacting with disabled persons, the extent to which the contact situation was pleasant, and the extent to which there was modeling either by peers or significant others of positive interactions with disabled persons (Category 7.a.-f.). Also coded was the existence or nonexistence of institutional, authority, or peer support for positive interactions and attitudes (Category 8.).

The characteristics of the disabled persons may have an impact on the outcomes of contact. Coded as characteristics of persons with disabilities were the type of disability which they had, whether they purveyed negative stereotypes, and the extent to which they were likely to be viewed as competent by the nondisabled Ss (Categories 9.a.-e.).

The characteristics of the nondisabled Ss are also theorized to be important factors in the outcomes of contact. In Category 10.a. through 10.d., we coded whether personality attributes of nondisabled Ss or their prior attitudes toward disabled persons were assessed and, if they were and the results analyzed, what the relationships were to post-treatment attitudes toward persons with disabilities. (Prior contact with disabled persons by the nondisabled Ss, another potentially potent predictor of the effects of attitude modification treatments, had already been coded with the prior contact coding sheet.) Again, the minutes spent in coding these categories were also entered on the coding sheet.

ES Information Missing Coding Instrument

As noted earlier, in most meta-analyses studies are rejected if information is not available for computing effect sizes; we decided, however (and later found that Slavin [1986] concurred), that important information would be lost by simply discarding those studies. Yet, without effect sizes to be used in our analyses, it did not seem economical to code the "missing effect size" studies as completely as those for which effect sizes were available. Consequently, an ES Information Missing Coding Instrument was developed by striking from the full coding instrument categories which seemed likely to be of little value for analysis without effect size data. (The ES Information Missing Coding Instrument is included in Appendix B.) Studies without effect size information were coded only if information about the statistical significance of results was available.

Coding Time

The time to be spent in coding is of concern to those planning quantitative, integrative reviews. In planning this integrative review, we estimated an average coding time of 3 hours per report, based on the meta-analysis coding experience of staff at USU's Early Intervention Research Institute. That figure turned out to be a slight underestimate.

The front page of each coding instrument has spaces for entering the times at which scoring began and ended. The next to the last category is "minutes spent coding". Reliability checks were not kept on that category and it must be recognized that times entered, although seemingly accurate to the minute, were often rough estimates. A major contributor to unreliability was the interruptions that inevitably occurred as raters scored--another rater needing counsel, or telephone calls and drop-ins from students, faculty, or other personnel as concurrent projects and university business had to be conducted.

In any event, as can be seen in Table 6, the mean scoring time of 175 minutes (2.9 hours) for the main coding instrument is amazingly close to the 3-hour estimate. However, with the mean time for the prior Contact Coding Sheet (12 minutes/.2 hours) and the Contact Coding Sheet (21 minutes/.35 hours) added on, the total time is 208 minutes (3 hours and 28 minutes, or 3.47 hours), and 3 hours is a 13 percent underestimate. It is also important to note that the coding time figures presented in Table 6 are probably underestimates or actual time spent in coding the reports. They tend to reflect the time actually spent in coding and are less likely, for example, to include time getting ready to code--selecting a report, perusing it to be sure it is relevant and that information necessary for coding is available (which often involved consulting with other staff members), and discussion of general coding issues that arose as raters grappled with applying the coding categories to individual reports.

As would be expected, the average time of scoring did differ among types of report. Table 6 presents that information for the main coding instrument. Combinations understandably took the longest time to code on the average (200

Table 6

Time to Code Reports for
197 T X C, T X P, Pre-post Studies

Main Coding Instrument

Report Type	N	Mean (min./hr.)[a]	Standard Deviation (min./hr.)[a]
Journal Article	54	138/2.3	88/1.5
Dissertation	103	195/3.2	90/1.5
Thesis	7	152/2.5	48/.8
Convention Paper	5	89/1.5	35/.6
Unpublished Report	10	176/2.9	96/1.6
Combination[b]	18	200/3.3	128/2.1
Total	197	175/2.9	96/1.6

Prior Contact Coding Sheet

N	Mean	Standard Deviation
197	12/.2	10/.2

Contact Coding Sheet

N	Mean	Standard Deviation
4/[c]	21/.35	13/.21

[a]Data presented are minutes/hours.
[b]Two or more reports of the same study, usually a dissertation-journal article combination. See Chapter 4 for more detail.
[c]This coding sheet was used only on those studies in which contact had been coded as an attitude modification technique in a treatment by control, treatment by placebo, or single-group, pre-posttest design.

minutes/3.3 hours), with the average time for dissertations very close (195 minutes/3.2 hours), and convention papers taking the least time on the average (89 minutes/1.5 hours). The larger standard deviation for combinations reflects the variety of associations of reports. While a combination was typically a dissertation and a journal article, it might also have been a journal article and a convention paper, or two articles.

Rater Reliability

In conducting a quantitative review of research, as with any data collection endeavor, there must be concern with the reliability of the data collected. However, as Glass and his associates (1981, p. 75) have noted, the situation in estimating the reliability of scores is somewhat different in meta-analyses than that in usual assessment settings where score inconsistencies can reflect both lack of stability in the phenomena being observed (or assessed) and inconsistency in judgments by observers (or assessors or raters). Because the data for meta-analyses come from written reports, the first source of inconsistency is eliminated and the principal source of measurement unreliability is rater inconsistency. Consequently, while one must be cautious in primary studies to distinguish between observer consistency and score reliability, with the former an ingredient of the latter, in meta-analyses the two are the same.

The estimation of the reliability of scores is crucial to the adequate interpretation of results. Clearly, as Jackson (1980) has also pointed out, in gathering data for a review of literature, two types of rater consistency are of concern—that is, consistency among the raters coding the same studies (inter-rater agreement) and consistency between the same rater coding a particular study at different points in time (intra-rater agreement or absence of drift). Both types of reliability were addressed in this study.

Most of the data collection for this integrative review was conducted by four raters--the principal investigator, a collaborator from another university, and two doctoral-level graduate students. Because of the difficulties of communication and because he took on other project tasks, the collaborator scored the fewest studies (about 14% of the effect sizes). The principal investigator coded approximately one-fourth of the effect sizes, while one graduate student coded approximately 28 percent and the other 33 percent of the effect sizes. As will be noted later, the two graduate assistants did the coding with the Prior Contact Coding Sheet (one coding about 45% and the other about 55% of the effect sizes), and the principal investigator and one of the graduate students did the coding with the Contact Coding Sheet.

As Orwin and Cordray (1985) have pointed out, little attention has been given to how to assess overall agreement for more than two raters. As did Bullock and Svyantek (1985) and Stock et al. (1982), we decided that percentage of agreement was the most accurate statistic to use for assessing inter-rater and intra-rater reliability. We recognized, however, that this method of estimating rater reliability was not without its problems.

The major difficulty in the use of percentage of agreement, from our point of view, is that an overall estimate of high agreement might be obtained despite substantial disagreement on some individual items (Orwin & Cordray, 1985), obscuring the need for corrective measures to enhance agreement (and validity). We countered that possibility by making the checking of both inter-rater and intra-rater agreement a group process. That is, with the exception of the times when the research collaborator was not present because he was coding on another campus, all raters met for reliability sessions. Eacn rater in turn indicated how he or she had coded each category, and the project director filled in a reliability check sheet. For intra-rater reliability checks, at least one other rater (usually the project director) read from one completed coding instrument and completed the reliability check sheet while the rater for whom the check was being completed read codings from the other completed coding instrument. This procedure not only provided the opportunity to ascertain any serious disagreements on category definitions, but to discuss even mild disagreements to enhance consistency.

A rigorous criterion for reliability--90% agreement--was set, even though a criterion of 80% agreement is commonly used. The 90% criterion was particularly stringent for inter-rater reliability because any categorization on which two or more of the raters disagreed was coded as a disagreement. Adopting Glass's convention on "near misses" (Glass et al., 1981), what we defined as "one space" disagreements were coded as agreements.

Inter-rater reliability. It was originally proposed that every tenth report, selected randomly, would be scored independently by two of the four coders on a rotating basis. Once the project was underway, however, it was decided that all of the raters would participate in each inter-rater reliability check in order to capitalize on the discussions of coding disagreements as a means of enhancing rater consistency, as noted in the prior section.

An inter-rater reliability check was conducted when any one of the raters had completed approximately 10 reports. The number of studies coded by different raters during a given period of time varied, depending upon factors such as report length, how difficult it was to extract information from a report, and the project duties that took varying amounts of the raters' time. As a consequence, at any one inter-rater reliability check, one or more raters may have coded slightly more than 10 studies and one or more may have coded less than 10 studies.

Once the coding instrument was judged to be in its nearly final form (early in December of 1985), coding checks were begun and from that time until the end of January 1986, 12 separate reliability checks were completed. The results were as follows: 88%, 82%, 87%, 91%, 81%, 86%, 91%, 66%, 86%, 89.5%, 94%, and 88% agreement, for the three raters who were on campus at Utah State University. When the combined scores of those three coders were compared with those of the coder who was off-campus for the last three reliability checks, the result was 95%, 93%, and 92% agreement. At that point, it was decided to go ahead with data collection.

On February 14, 1986, an inter-rater reliability check for all four coders yielded 89% agreement, which was deemed close enough to the 90% criterion to proceed. (When coded by the less stringent procedure of counting the number of agreements and disagreements per category, instead of scoring each category on a dichotomy of either all four coders agreed or disagreed, there was 97% agreement.)

The next inter-rater agreement check was on March 6, 1986. The three on-campus coders had a 90% agreement; when the combined scores of those three coders were compared with those of the off-campus coder, there was 98.5% agreement.

On March 24, a third check was done. The three on-campus raters had 96% agreement (with 91% perfect agreements). There was 90% agreement between the off-campus rater's codings and the combined codings for the three on-campus raters.

On April 15, an inter-rater check for the three on-campus coders yielded 96% agreement; on May 5, 93% agreement; and, on May 19, 85% agreement. By April 15, the off-campus rater was no longer coding reports, so was not included in the checks. Because criterion was not met on May 19 (85% agreement), a second study was coded, yielding 90% agreement.

The last inter-rater reliability check on the main coding instrument was conducted on June 4, 1986, with 94% agreement.

Prior contact rater agreement

The inter-rater agreement rate was also checked for the Prior Contact Coding Sheet. Because the sheet was rather simple and straightforward to use, no problems were anticipated and none were encountered. A check among the three on-campus raters yielded a 100% agreement before coding began. Coding was then conducted by the two graduate assistants. Again, an inter-rater reliability check was conducted for every 10 reports. All but one yielded 100% agreement; for the one, agreement was 90%.

Contact rater agreement

The distinctions to be made in coding the Contact Coding Sheet were more subtle, and it was more difficult for the principal investigator and graduate assistant to attain adequate inter-rater agreement. Eleven reliability checks were conducted over a two-week period in January 1987 before independent scoring commenced. The initial reliability sessions also served as formative evaluations of the coding conventions, although the categories in the coding instrument had stabilized by then. During and after each reliability check session, conventions were revised prior to further coding. The percentages of agreements attained were: 72%, 90%, 72%, 90%, 90%, 86%, 83%, 78%, 89%, 94%, and 94%. An inter-rater reliability check was again conducted for approximately every 10 reports. Given the relatively small number of contact studies (N = 41), only two checks were made. Agreement was 88% for the first one and 94% for the second.

Effect size accuracy

As an additional caution, because effect sizes are such a central part of a quantitative review, every effect size was checked for accuracy. Not only were computations redone, but the selection of statistics from the report for computing each effect size was reviewed for corrections. Thirty-one errors were detected (and corrected), for an overall mean accuracy rate of 94%. Taken by rater, one was 99% accurate, two were 97% accurate, and one was 83% accurate.

The lowest accuracy rate was for the off-campus coder, which was not surprising considering the amount of communication which took place among the on-campus coders. Because it was assumed that the purpose of coding was to obtain the most valid possible representation of the studies, the on-campus reviewers commonly discussed with one another difficult coding choices when they were not coding for a reliability check (that coding was, of course, done independently). In terms of accuracy of computing effect sizes in particular, potential errors were occasionally picked up through this interactive process.

Our experience with this study indicates that it is feasible to use raters who cannot communicate in person with one another whenever they wish. The success of the off-campus coder in using the coding instrument is also tentative evidence for the replicability of our results (for the pitfalls of data collection in replicating meta-analyses, see Bullock & Svyantek, 1985, and Stock et al., 1982). Our experience also suggests, however, that frequent communication among raters is a desirable feature of data collection for enhancing inter-rater agreement and reducing even a small margin of error.

Intra-rater reliability. As originally proposed, after coding approximately 30 reports, each rater recoded one of the reports at the beginning of the sequence, with the particular one selected by the project director. Recoding was, of course, done without benefit of the first coding sheet. Again, the criterion was 90% agreement.

Due to different rates of coding reports (as noted above), one rater had three intra-reliability checks, one rater had two intra-reliability checks, one rater had one intra-reliability check, and one rater coded fewer than 30 reports so had no checks. For the first rater, the percentages of intra-coder agreement were 98%, 96%, and 93%; for the second rater, the figures were 91% and 92%; and, for the third rater, there was 96.5% agreement.

For the Prior Contact Coding Sheets and the Contact Coding Sheets, there were so few studies for which prior contact was assessed that no intra-rater reliability checks were conducted.

Summary. Overall, the reliability of scoring was deemed to be adequate, based both on the percentages of agreement and the general consensus that seemed obvious as the coders discussed studies during the reliability checks and in communicating with one another as problems arose during the coding of nonreliability check reports

Data Analysis

As Glass and his associates (1981, pp. 197-200) have pointed out, the role that statistical inference should play in meta-analyses is anything but clear. There are a number of reasons for not using inferential statistics in an integrative review such as reported here. The first and perhaps most obvious is that the data to be analyzed constitute an accessible population, not a sample (although, of course, there is always the sticky conundrum, pointed out by Jackson [1980, p. 453] and Cooper [1982, pp. 294-5], of whether to consider the primary studies being reviewed or the populations and settings in which the findings might be applied as the appropriate target in drawing conclusions). Glass et al. (1981, pp. 199-200) discussed, without great clarity, the advisability of dealing with sampling error even when the studies reviewed are considered to be an accessible population, proposed inferential techniques to be used, and recounted the experience of being chided by Tukey for not presenting standard errors for mean effect sizes. Nevertheless, the use of inferential statistics in analyzing data considered to come from an accessible population would appear to be more a perpetuation of ritual than a rationally justified procedure.

The statistical inference Zeitgeist has amazing tenacity. Along with Glass et al. (1981), other books on quantitative review techniques have sections on the use of inferential statistics (e.g., Cooper, 1984; Hedges & Olkin, 1985; Hunter et al., 1982; Rosenthal, 1984; and, Wolf, 1986), and reports of meta-analyses frequently contain inferential statistics results. The misplaced emphasis on inferential statistics is nowhere better illustrated than by the dysfunctional recommendation that statistical power be increased by pooling the data from primary studies (Hedges & Olkin, 1986).

The overreliance on and misinterpretation of inferential statistics in primary research have been commented on (see, e.g., Carver, 1978; Shaver, 1979a, 1980, 1985a, b). The use of an indicator of the significance of research results which is dependent upon sample size, as statistical significance is, is no more appropriate in analyzing the findings from primary research reports in an integrative review than it is in primary research. As Wilson and Rachman (1983) have pointed out, there is with meta-analyses, as in primary research, the "danger that [the use of] sophisticated statistical techniques [will] serve [to] obscure damaging flaws in the evidence" (p. 55).

In this study, we attended to the prior criticisms of the misuse of inferential statistics to analyze data in primary studies and followed the logic of the irrelevance of inferential statistics with population data. The basic analytic approach, therefore, was descriptive. Glass's (1976) recommendation was accepted that variables which might have moderated treatment effects be investigated, an approach which can be recast as determining to what extent treatment effects appear to be nested within other variables, such as the age-level of subjects and the type of research design. Considerable effort was put into attempting to disentangle treatment techniques from study characteristics in order to determine what conclusions about treatment effects might be legitimately drawn and, conversely, to determine what confounding factors might be accounting for the inconsistent results referred to in prior reviews.

As with tests of statistical significance, we also eschewed other inferential procedures such as tests of homogeneity of effect sizes (Hedges, 1982), the estimation of effect size from a series of independent experiments (Rosenthal & Rubin, 1982), adjustments to increase the accuracy of effect sizes (Hedges, 1981) (which in the experience of the Early Intervention Research Institute staff at Utah State University tend to produce negligible and practically unimportant adjustments [also see Bangert-Drowns, 1986]), and weighting procedures for minimizing cumulated effect size variances (Bangert-Drowns, 1986; Green & Hall, 1984).

Basic descriptive statistics were computed--means, modes, medians, standard deviations, and ranges--and frequency distributions are sometimes reported, too, in the Results section. Two and three-way frequency tables were used to illustrate possible interactions between variables, especially treatment techniques and study characteristics. The approach was to demonstrate how the characteristics of the studies in our accessible population, particularly the treatment techniques, were related to the size of effects.

Index of triviality. Even with an accessible population of studies and without the use of inferential statistics, the researcher is faced with issues of inference. For example, how does one decide how large a difference between mean Ds (e.g., the mean Ds for two treatments or for two categories of study quality) must be in order to be considered "important"? The question is, of course, analogous to the question of how to determine if the result of a piece of research, expressed as a difference between posttest means, is of practical significance. And our solution was analogous as well, although our initial question was somewhat different. It was, at what point could we consider a difference in mean effect sizes to be so trivial that we would be justified in not paying further attention to it? Such an index of triviality is not the clear converse of an index of importance. That is, a difference below the index would be judged clearly trivial, but one greater than the index would not necessarily be of practical importance.

In addressing the question of practical importance and triviality, we treated mean Ds as we would treat the means for data collected on Ss in a primary piece of research: we computed an effect size by subtracting one mean from the other and dividing by a standard deviation. In this case, the means were population values, as was the standard deviation. Consequently, the use of Cohen's (1977) symbol, d, to represent the effect size is appropriate, because he defines that effect size as a population parameter (pp. 9, 20).

Because we had only one standard deviation to deal with—the population parameter, which was .61*--it was not necessary to compute a d for each comparison. Once a minimum d was established for purposes of judging triviality, it could be determined how large a difference in means had to be in order to attain that value.

Setting such a convention is a dubious process. Cohen (1977, p. 12) has cautioned that establishing conventions for judging when an effect size is of acceptable, or unacceptable, magnitude is as arbitrary--and as potentially useful, yet subject to misuse--as the use of the common .05 criterion for statistical significance. Glass et al. (1981, p. 104) argued strongly against the establishment of conventions by which to label regions of an effect size metric with such adjectives as "small", "moderate", or "large". As they correctly noted, an effect size can only be interpreted meaningfully in context, that is in terms of the benefits to be achieved given the cost of producing the result (also see, Shaver, 1985a, b).

It should be noted that the arguments of Cohen (1977) and Glass et al. (1981) are directed against the setting of conventions that would be used across primary research studies or integrative reviews, as has been the case

*This is the standard deviation for the population of studies that constituted the main analyses reported in Chapter 6. It is nearly identical to the standard deviation of .62 for the 705 Ds in the total data set.

with the .05 level of statis cal significance. In a particular piece of research, such as our review of the research on modifying attitudes toward persons with disabilities, one must address the matter of a minimum interpretable effect size either directly or through implicit, unarticulated, standards. We chose the former, but the answer was not easy to come by, even following Cohen's one guideline for such decisions, "that they not be unreasonable" (p. 12).

Setting the standard for a minimum effect size is particularly difficult when it is not clear what benefits are indicated by different values of a dependent measure. This is a particular problem in the attitudes toward disabled persons research. Data are not available that allow the valid formulation of clear expectations as to the effects of a difference in scores on an assessment such as the Attitudes Toward Disabled Persons Scale (Yuker et al., 1970) which was widely used in the primary studies. How different must their scores be before Ss will differ in their thoughts the next time they encounter a person with a disability or, more importantly, will behave differently toward a person with a disability or toward other nondisabled persons who make stereotypic statements about disabilities or the persons who have them? Or, how different must such scores be to indicate, if they ever will, that the Ss will take different stances on public policies that affect educational, economic, and social equity for persons with disabilities? In the absence of empirically-based answers to such questions, establishing a minimum effect size for interpretability is, at best, a loosely bounded guessing game, albeit it a necessary one.

We proceeded by using Cohen's (1977) standard for a "small" effect size (pp. 25-26), $\underline{d} = .2$, as a starting point. With a population standard deviation of .62, a mean difference of .12 would yield a *d* of .2. The corresponding r_{pb} is .10, an obviously small correlation that suggests an obviously small amount of common variance (i.e., r^2_{pb} = .01 or 1%). Overlap of distributions is another corollary to *d* as an indicator of effect size magnitude (e.g., Cohen, 1977; Glass et al., 1981). Cohen (p. 22) indicates, for example, that with a *d* = .20, the overlap of distributions is 85.3 percent (14.7% nonoverlap). This interpretation depends, however, on the assumption that the distributions are normal, an assumption not tenable with some of our data.

The same line of reasoning was applied to the analysis of the variance ratios computed when variances were available for the treatment groups (see Chapter 6). For the 453 variance ratios that could be computed for the 644 effect sizes used in the main analyses, the standard deviation was .91 (with a mean of 1.13, very close to the value for equal variances). To yield a *d* of .2, a difference in mean variance ratios would have to be .18 (the same as obtained using the standard deviation of .88 for the 705 effect sizes), our index of triviality for variance ratios.

It did not seem likely that anyone would argue that we would be overlooking important differences with a difference in mean *D*s of .12 or in mean variance ratios of .18 as an index of triviality. Also, those indices were consistent with our earlier intuitive judgments, in inspecting data, about mean differences that hardly seemed worth attending to.*

---

*Cohen (1977) specifies *d* = .5 as a medium effect size and *d* = .8 as a large effect size. If applied to our data set, a difference in mean *D*s would have to be .30 to be "moderate" by that criterion and .49 to be "high"; a difference in mean variance ratios would have to be .45 to be a "moderate" effect size and .73 to be "large". If a commonly used criterion of practical significance, *d* = 1.00, were used, the difference in mean *D*s would, of course, have to be .61 and the difference in mean variance ratios, .91.

Minimum ns. Another question that had to be addressed was how many *Ds* must a result be based on before it would be considered sufficiently stable for interpretation. That is, just as replications are important to establishing the stability of a finding in primary research, so is it important in conducting a review of research not to over-interpret findings that may not be stable. Even when dealing with a population of studies, the confidence one has in a result must be based on, among other things, the number of data points on which it is based. With few guidelines to go by, it was decided that a mean *D* had to be based on at least 10 *Ds* to be considered sufficiently stable for interpretation. Although our principal set of primary research reports had a mean of 3.3 effect sizes per report, the median and mode were both 2. Consequently, it can be anticipated that a mean *D* based on 10 *Ds* will typically reflect data from approximately 5 separate studies.

Our criterion for multiple effect sizes should not be confused with a demand for replications, in the sense of either repeated observations on randomly assigned units or planned repetitions of studies. To confuse the presence in the literature of several studies on a topic with replication as a planned research strategy, as, for example, Bangert-Drowns (1986, p. 398) and Jackson (1980, p. 445) seem to have done, would be an error. We simply sought a minimum sense of stability, which is significantly different from addressing the congruence among studies implied by the replication terminology.

Binomial effect size display. It was anticipated before the study began that the binomial effect size display (BESD) (Rosenthal, 1984; Rosenthal & Rubin, 1982) might be a helpful interpretive device. Using the BESD involves displaying results in a 2-by-2 frequency table according to treatment condition and success rate. The example used to demonstrate application of the BESD has to do with the numbers of persons who are alive or dead following exposure to a treatment or a control condition. In such a case, "success" is obvious.

As the earlier discussion of the meaning of attitude scores in arriving at an index of triviality indicated, establishing "success" and "failure" is not so simple a matter in attitude modification research. Preece (1983) has noted that the use of a median split to effect the necessary dichotomy is meaningless in many cases. Ours is one of them. With no feasible means of arriving at a valid success rate, the BEMD turned out not to be useable for our interpretative purposes.

Data Availability

As noted earlier, lists of the studies that were coded, as well as those discarded, are contained in Appendices D and E. Descriptive data for the studies coded are included in Appendix F. The coding instrument and conventions for coding are in Appendices B and C. In addition, the complete data set is on magnetic tape and can be obtained at cost from the project director.

Summary

In this chapter, the procedures followed in conducting our integrative review have been spelled out, with considerable detail in regard to the obtaining of primary research reports and the coding instrument used for data collection. The outlines of our approach to data analysis have been sketched. The details will be filled in as some information about our population of studies is presented in the next chapter, followed by the results of analyses of our data.

CHAPTER 4

SOME ATTRIBUTES OF THE POPULATION OF STUDIES

The major concern in conducting this integrative review of the research on modifying attitudes toward persons with disabilities was, of course, to answer the question, what does an analysis of the population of research studies indicate about the effectiveness of different techniques for modifying attitudes toward persons with disabilities? To set the context for characterizing the relevant findings from the primary research reports that we located, we present in this chapter some general information about the population of studies that we coded and analyzed.

Numbers of Studies and Reports

As noted in Chapter 3, 273 primary research studies were coded for this integrative review. The number of reports coded was actually higher (N = 303) because any piece of research reported in multiple sources was coded as one study. For example, Donaldson's (1974) dissertation research was also reported in journal articles she authored (Donaldson, 1976) and coauthored (Donaldson & Martinson, 1977). Donaldson's study was coded only once, based on all three reports, and contributed only once to our list of 273 studies.

Multiple reports of studies were most common, as they should be, with dissertations and convention papers. Dissertations or thesis research was reported in 33 journal articles, as well as in two convention papers, one agency report, and one book. Similarly, 6 studies reported in convention papers and 3 reported in agency reports were also available in journal articles. Publication of the same research in different journal articles was, as is to be expected, infrequent. Three sets of articles in different journals were based on the same data (Dye, 1978, 1980; Donaldson, 1976; Donaldson & Martinson, 1977), including a follow-up study (Esposito & Peach, 1983; Esposito & Reed, 1986).

The 273 studies yielded 725 effect sizes (see Table 7--644+61+20), as well as 182 results for which information to compute an effect size was not available, but information on statistical significance was. A total, then, of 907 effect sizes and "no information" results were coded.

Partitions in the Data Set

Table 8 presents further information by type of comparison, with the Mainstream and the No Information comparisons omitted. That omission reflects early analysis decisions. Different categories were used to code the treatment in studies of attitude change with classroom mainstreaming (although studies which included other types of in-school contact, such as having a self-contained special education classroom in the school, were analyzed using the main coding instrument), and it was decided to analyze those effect sizes separately. Also, the No Information reports lacked basic information, but were analyzed separately, primarily to determine if studies lacking the information to compute effect sizes differed otherwise from the effect size reports.

The other categories in Tables 7 and 8 represent early coding decisions. The single-group, pre-posttest design is of dubious quality. Studies using that design were coded, however, because there seemed to be a large number of them and it was thought important not to lose the potential of garnering information from those results.

Table 7

Numbers of Attitude Modification
Effect Sizes and Studies

Effect sizes		
Type of Comparison	Number	Number of Studies
T vs. C	497	
T vs. P	49	
Pre-post	98	
Subtotal	644	200
A vs. B	61	20
Mainstream	20	10
No information results	182	53
Total	907	283(273)*

Note: T X C = treatment-control group comparison; T X P = treatment by placebo group comparison; pre-post = single-group, pretest-posttest mean comparison; A vs. B = comparison of two treatment groups, Mainstream = comparison from a study of the effects of mainstreamed classrooms; No information = statistical significance information available, but not information for computing effect size.

*Because a study could yield more than one type of comparison, the total of 283 exceeds the number of studies actually coded, i.e., 273.

Table 8

Types of Comparisons: Effect Size Ns, Means, and Standard Deviations for 705 Effect Sizes

	Effect Sizes (Ds)				
Comparison Type	N	Mean	Median	s.d.	Range
T vs. C	499	.36	.32	.57	-1.42 to 4.40
T vs. P	49	.29	.23	.70	-1.22 to 2.42
Pre-post	97	.50	.33	.72	-1.61 to 3.11
A vs. B	60	.15	.06	.71	-1.07 to 3.55
Total	705*	.36		.62	-1.61 to 4.40

*An effect size was discarded and the Comparison Type categorization of two effect sizes were changed during later analyses upon which Table 7 is based; consequently, the numbers in this and later tables will not match perfectly with those in Table 7.

By the same token, it seemed important to distinguish between "control" and "placebo" groups, terms misused by many researchers in discussing research designs. A "control" group is one which receives no treatment, while a "placebo" group (sometimes referred to as a type of control treatment) receives attention and/or materials that are assumed to be inert (i.e., are assumed to have no affect per se on the dependent variable), but in all other ways (e.g., time, amount of experimenter attention) comparable to the experimental treatment (see, e.g., Borg & Gall, 1983, pp. 221-22, 355). It seemed likely that control and placebo groups might react differently due to the differences between receiving no treatment and some, even if supposedly inert, treatment. Consequently, we coded which of the designs was involved in a treatment comparison.

Analysis Base

It was further decided that the comparison base for an attitude change treatment would, when available, be the absence of treatment--i.e., a control or placebo condition*--rather than another treatment. With data from a number of treatment comparisons, it would be difficult to illustrate and keep evident what had been compared across effect sizes; data based on comparing each treatment against an absence of treatment to obtain an effect size are more amenable to interpretation. Consequently when two treatment groups (i.e., Treatment A and B) were present in a study and each was compared with a control or placebo group, effect sizes were computed and coding conducted for the treatment versus control (T vs. C in Tables 7 and 8) or treatment

---

*The single-group, pre-post design is, of course, a weak form of this comparison with the pretest serving as an indication of attitudes in a no-treatment, control situation.

versus placebo (T vs. P) comparisons, and not for the Treatment A versus Treatment B (A vs. B) comparison. Treatment A versus Treatment B (A vs. B) effect sizes were, however, computed and coded when a T X C or T X P effect size was not available. Our later difficulties in untangling the A vs. B Ds which were coded confirmed the soundness of the decision not to code them generally.

Also, as can be seen in Table 8, A vs. B comparisons yielded Ds that were, on the average, different than those from T vs. C and T vs. P comparisons. As would be expected when comparing differing treatments rather than treatment against lack of treatment, A vs. B comparisons yielded the smallest mean D, .15, barely above the standard we set for trivial results --.12 (see Chapter 3). The lower median D of .06 is probably a better indicator of central tendency, as two outliers (Ds of 2.71 and 3.55, with the next lowest D, 1.26) distorted an otherwise near-symmetrical distribution. Also, sizeable differences between the pre-post mean effect size mean (.50) and the T vs. C (.36) and T vs. P (.29) effect size means (differences of .14 and .21, respectively) look trivial when the median is used as the measure of central tendency.

Most of the reporting of our data analyses will be based on the 644 T vs. C, T vs. P, and pre-post Ds. However, the description of the data set that follows includes A vs. B comparisons as well so as to provide a more complete picture of the available body of research on modifying attitudes toward disabled persons.

Types of Reports

One aspect of integrative reviews that has attracted interest is the types of research reports that were coded. As is evident in Table 9, our

Table 9

Ds by Type of Report
for 705 Effect Sizes

	Effect Sizes (*Ds*)			
Report Type	N	Mean	Median	S.D.
Journal Article	141	.59	.42	.69
Dissertation	430	.24	.22	.60
Thesis	13	.31	.30	.29
Convention Paper	13	.37	.38	.40
Unpublished Report	30	.37	.36	.38
Combination	78	.54	.44	.61
Total	705	.36	.36	.62

major source of effect sizes was dissertations (N = 430), with journal articles second (N = 141). Because a journal article was included in most of the "combination" sources (e.g., effect sizes available from a dissertation and a journal article), the number of journal article-based Ds is close to 200.

Of greater interest than the frequencies of Ds by type of publication are the mean Ds. They support the observation by Bryant and Wortman (1985), as well as by Bangert-Drowns (1986) and Walberg (1986)--based on Smith (1980)--that published articles have consistently higher effect sizes than do dissertations and other unpublished sources. The mean D for journal articles is .59 and for dissertations, .24; the difference of .35 yields a d of .56, using the population standard deviation of .62 (see Analysis section of Chapter 3). The mean for Ds from combined sources (which, again, are largely those Ds that came from dissertations and other unpublished reports from which articles acceptable for publication were drawn) is .54, almost identical to the mean D for articles. Dissertations have the lowest mean D, .24. That mean is within .12 (our standard for a trivial difference; see Analysis section of Chapter 3) of the thesis mean D (.31), but is barely more than .12 lower than the means for convention papers and unpublished reports (mean Ds for both = .37). The picture does not change much if the median Ds, which are less influenced by outlying Ds, are inspected. Interestingly, the dispersions are similar for journal article, combination source, and dissertation Ds (s.d. = .69, .61, and .60, respectively), with the Ds from the other three types of publications having lower and fairly similar standard deviations.

The potential for a distorted picture of research findings in a review based only on published reports is obvious from Table 9. Whether the higher

D for journal articles is a reflection of the type of designs in studies that are submitted and accepted for publication is addressed in part by Table 10. The higher mean Ds for published articles, as contrasted with dissertations, are clearly evident across all types of comparisons. Journal mean Ds are also higher than those for the other report types except for convention papers and unpublished reports based on treatment by control (T vs. C) designs.

It is noteworthy that treatment by control designs were more frequent for dissertation effect sizes than would be expected based on marginal totals, and less frequent than expected for journal article effect sizes. At the same time, single-group, pre-post designs were more frequent than expected for journal article effect sizes and less frequent than expected for dissertations. (Many of the journal articles, we found, were reports of "convenience" research in which students in a college course were pre- and posttested.) Similarly, while the mean D for T vs. C results (.49) published in journals was lower than the journal Pre-post mean (.65; 1.05 for combination sources), the reverse was true for dissertations (.31 versus .05). Particularly striking as well, given the mean overall D for A vs. B studies of .15 is the mean D for the eight journal article effect sizes--1.19. That is in sharp contrast to the mean D of -.03 for dissertation A vs. B effect sizes. The data in Table 10 suggest rather strongly either a predilection to submit research for publication only when the results are striking (i.e., a large, statistically significant difference), or to accept only such reports for publication. There does appear to be a bias against the importance of confirming the lack of an effect.

Table 10

$\underline{D}$s by Type of Report and Type of Comparison for 705 Effect Sizes

Report Type	Comparison Type: T vs. C	T vs. P	Pre-post	A vs. B	Totals
Journal Article	.49 78[a]	.68 12	.65 43[b]	1.19 8	.59 141
Dissertation	.31 332[b]	.ii 32	.05 20[a]	-.03 46	.24 430
Thesis	.31 11	.31 2	— 0	— 0	.31 13
Convention Paper	.48 5	— 0	.31 8	— 0	.37 13
Unpublished Report	.46 17	— 0	.27 13	— 0	.37 30
Combination	.46 56	.62 3	1.05 13	.14 6	.54 78
Total	.36 499	.29 49	.50 97	.15 60	.36 705

Note. The top number is the mean $\underline{D}$; the bottom number is the frequency of $\underline{D}$s.

[a]At least 10 less than expected, based on marginal totals.
[b]At least 10 more than expected, based on marginal totals.

162

Research Over Thirty Years

The amount of research in an applied field, such as the modification of attitudes toward disabled persons, is likely a function of a number of factors, including public interest, especially as reflected in federal funding for the particular research area. The growth in research in this field has been great since the 1955-60 period. As indicated by the total numbers of effect sizes we coded (see Total Effect Sizes column in Table 11), there was a 2700 percent growth from the 1955-60 period to the 1971-75 period--from 3 effect sizes in 1955-60 to 91 effect sizes in 1971-75. (With a mean of 3, and a median [and mode] of 2 effect sizes per study, that is a growth from approximately 1 or 2 reported studies to approximately 46.) That percentage of change is large because the base is so small. What catches the eye is the over 200 percent growth from 1971-75 to the next time period, 1976-80--from 91 effect sizes, or about 46 reported studies, to 286 effect sizes, or about 143 studies.

Keeping in mind the time lags between various steps in the production of applied research--expression of public concern, legislative appropriations, program announcements, funding of projects, proposal submission and approval, completion of research, and report preparation and publication--an upsurge in concern for persons with disabilities seems evident in Table 11. (Note that the overall picture of growth in research portrayed by the effect size data in Table 11 is consistent with the report data in Figure 1, Chapter 2.) Why the drop in number of effect sizes from 1976-80 to 1981-86? It may be an artifact of the difficulty in locating studies which are not yet well-cited in the reference lists of reviews and other research reports, or it may reflect a decline in interest in this area of research.

Table 11

Frequencies of Types of Comparisons by Year of Publication for 705 Effect Sizes

Year	Type of Comparison				Total Effect Sizes
	T vs. C	T vs. P	Pre-Post	A vs. B	
1955-60	2	0	1	0	3
1961-65	13	0	8	2	23
1966-70	44	0	14	5	63
1971-75	62	4	17	8	91
1976-80	228[b]	11	26[a]	21	286
1981-86	150[a]	34[b]	27	24	235
Total	499	49	93	60	701[c]

Note: Data are effect size (D) frequencies. Cramer's V = .16.

[a]At least 10 less than expected, based on marginal totals.
[b]At least 10 more than expected, based on marginal totals.
[c]Rowe & Smith, in press, had 4 effect sizes that are not included in this table.

Types of Comparisons

Given general concerns about research design, it is of interest whether the designs used in disabilities attitude modification research have changed over the years. Table 11 indicates little relationship between year of publication and type of comparison for the effect sizes in our set. A low Cramer's V (.16) reflects a relative increase in treatment-control (T vs. C) comparisons reported in 1976-78, an increase in treatment-placebo (T vs. P) comparisons reported in 1981-86, relatively fewer pre-post comparisons reported in 1976-80, and fewer treatment-control comparisons in 1981-86, than expected given marginal totals.

Attitude Modification Techniques

Have researchers concerned with attitudes toward disabled persons directed their attention to different attitude modification techniques over the years? The type of modification techniques used in the research which yielded our Ds changed somewhat over the 31-year span for which we located reports (see Table 12). Perhaps of most interest, compared to the frequencies expected based on marginal totals, are the dramatic increase in the number of information technique effect sizes for 1976-80 period, the decrease in number of contact effect sizes reported in 1981-86, the spate of vicarious experience effect sizes from the 1976-80 reports, followed by a decline in 1981-86 which was accompanied by a jump in vicarious experience plus information effect sizes in 1981-86. The rather dramatic increase in effect sizes for which a combination of techniques was used (the Other category) in 1976-86 is also of interest, perhaps indicating efforts to go beyond conventional modes of thought in regard to modifying attitudes toward persons with disabilities.

Table 12

Frequencies of Types of Modification Techniques by Year of Publications
(T X C, T X P, Pre-post, A vs. B effect sizes included)

Year	Type of Technique									
	Information	Contact	Information + Contact	Vicarious Experience	Information + Vicarious Experience	Positive Reinforcement	Persuasive Message	Systematic Desensitization	Other	Total Effect Sizes
1955-60	1	2	0	0	0	0	0	0	0	3
1961-65	11	6	4	0	0	0	0	0	2	23
1966-70	7[a]	21[b]	20[b]	9	0	0	6	0	0	63
1971-75	28	11	27[b]	1	0	3	3	4	14	91
1976-80	107[b]	34	36	35[b]	13[a]	0	22	11	28[a]	286
1981-86	77	21[a]	20[a]	13	52[b]	0	3	0	49[b]	235
Total	231	95	107	58	65	3	34	15	93[c]	701[c]

Note: Data are effect size (D) frequencies. Cramer's V = .26.

[a]At least 10 less than expected, based on marginal totals.
[b]At least 10 more than expected, based on marginal totals.
[c]Rowe & Smith, in press, had 4 effect sizes that are not included in this table.

Types of Disabilities

Another question of interest in regard to our data set is, what disabilities have been the targets of attitude modification research, and have there been changes over the years? As can be seen in Table 13, the most frequent target of the attitude change studies from which our Ds came was the general category of persons with disabilities (N = 330). Attitudes toward persons with physical disabilities were a distant second (N = 97), with attitudes toward mentally retarded persons third (N = 81), and attitudes toward the mentally ill fourth (N = 67).

There have been some changes over the years in the disabilities that have been the targets of attitude modification studies. Table 13 contains information about those changes. Particularly striking is the increase during 1976-86 in effect sizes that came from studies in which disabilities in general were the attitude target of concern. Given the evidence that attitudes do differ according to disabilities (although according to Yuker [1983], the differences may not be as uniformly stable as some [e.g., Richardson & Ronald, 1977]* have claimed), a broad focus on changing attitudes toward disabilities in general may not be a wise strategy. In contrast is the increased number of effect sizes during 1976-80 that came from studies of attitudes toward specific mental retardation levels, rather than toward mental retardation generally. However, the general drop-off in 1981-86 in effect sizes from research on attitudes toward mentally retarded persons, persons with physical disabilities, and those who are mentally ill may be perplexing to some professionals and advocates. The absence of

---

*Also, Abroms and Kodera (1979), Antonak (1980), Richardson, Goodman, Hastorf, and Dornbusch (1961), and Tringo (1970).

Table 13

Frequencies of Disabilities Toward Which Treatment Directed, by Year
(T X C, T X P, Pre-post, A vs. B effect sizes included)

Year	General[a]	Physically Disabled	Mentally Retarded General	Mentally Retarded Moderate	Mentally Retarded Severe	Mentally Ill	Emotionally Disturbed	Visually Impaired	Hearing Impaired	Learning Disabled	Para-Quadra-Plegic	Other Physical	Other	Combination	Total Effect Sizes
1955-60	1	0	0	0	0	0	0	0	0	0	0	0	2	0	3
1961-65	2	0	5	0	1	12	0	3	0	0	0	0	0	0	23
1966-70	9[b]	10	7	0	0	18[c]	5	1	0	0	0	7	0	6	63
1971-75	33[b]	18	12	0	1	19[c]	0	6	0	0	0	0	2	0	91
1976-80	141	46	6[b]	20[c]	13	18	4	6	9	2	2	2	0	17	286
1981-86	144[c]	23[b]	8	0	8	0[b]	4	4	15	4	0	2	10	13	235
Total	330	97	38	20	23	67	13	20	24	6	2	11	14	35	701[d]

Note. See conventions for Category C.5. in Appendix C for definitions of disability categories. Data are effect size *D* frequencies. Cramer's V = .33.

[a]No disability target was specified.
[b]At least 10 less than expected, based on marginal totals.
[c]At least 10 more than expected, based on marginal totals.
[d]Rowe & Smith, in press, had 4 effect sizes that are not included in this table.

research on attitudes toward persons with some disabilities, such as speech-language impairments, may also be a matter of concern. On the other hand, even the modest increase in effect sizes from research on modifying attitudes toward the hearing impaired during 1976-86 may be encouraging to others.

Educational Level

Also of interest in terms of the thrust of research in modifying attitudes toward persons with disabilities is the educational level of the subjects used in the research. Table 14 shows that nearly the same number of effect sizes came from college and university samples (N = 313; 44%) as from elementary and secondary schools combined (N = 282; 40%), with the majority of the latter group coming from the elementary grades. Little research has been done with adult groups, with a total of only 103 effect sizes (14%) coming from studies in which samples of postprofessional persons or adults not in school were used.

Assessment

What methods have been used to assess attitudes in the research on modifying attitudes toward persons with disabilities? Given the concerns commonly expressed about the validity of questionnaires for predicting the behavioral aspects of attitudes, it is disconcerting to find that 66% (N = 460) of the effect sizes we obtained were based on questionnaire data (see Table 15), with items usually of the Likert-scale type. (It will not be surprising to those familiar with the research literature which we coded to know that for 44% [N = 201] of the questionnaire-based Ds, the Attitudes Toward Disabled Persons [ATDP] scale [Yuker et al., 1970: Yuker & Block,

Table 14

Educational Level of the Ss for 704 Effect Sizes

Context	Effect Sizes N	%
Can't tell	5	1
School		
Preschool	3	0.4
Primary	30	4
Intermediate	116	16
Middle	18	3
Junior High	17	2
Senior High	19	3
Combination	74	10
Total	282	43
College		
Undergraduate	283	40
Graduate	30	4
Total	313	44
Postprofessional	53	7
Adults not in school	50	7
Other	6	1
Total	704	101

Note. Proportions in this table and the ones that follow were rounded to two decimal places before being converted to percentages, except where less than .005 to avoid rounding to zero where there was a frequency. Also, as a result of rounding, percentages will sometimes add up to slightly more or less than 100.

Table 15

Frequencies of Types of Assessment by Year (T X C, T X P, Pre-post, A vs. B effect sizes included)

	Assessment Type													
	Interview													
Year	Structured	Unstructured	Questionnaire	Sociometric	Social Distance	Systematic Observation	Semantic Differential	Telephone -Mail	Projective Technique	Sentence Completion	Adjective Checklist	Rankings	Other	Total Effect Sizes
1955-60	0	0	0	0	0	0	0	0	2	1	0	0	0	3
1961-65	0	0	17	0	0	0	3	0	0	0	2	0	1	23
1966-70	0	0	43	0	2	0	11	0	0	0	2	0	5	63
1971-75	0	0	79[b]	0	3	0	8	0	0	0	0	0	1	91
1976-80	1	0	181	4	31	2	47[b]	2	0	1	4	3	10	286
1981-86	0	2	140[a]	6	27	1	10[a]	2	0	0	24[b]	1	22	235
Total	1	2	460	10	63	3	79	4	2	2	32	4	39	701[c]

Note. Data are effect size (D) frequencies. Cramer's V = .44.

[a]At least 10 less than expected, based on marginal totals.
[b]At least 10 more than expected, based on marginal totals.
[c]Rowe & Smith, in press, had 4 effect sizes that are not included in this table.

1986] was the assessment tool.) Only 3 Ds were based on systematic observation data, and only 2 of the 4 Ds that came from telephone or mail surveys were aimed at obtaining a response to a situation removed from the research project (such as a poll of opinions toward use of university money for a Center for Disabled Students) so that Ss would not be likely to see a connection to the research and thus give biased responses (not in Table 15). It should be noted that the 39 "Other" effect sizes included those obtained with assessments of behavioral intentions (e.g., responses to questions about intent or willingness to invite a disabled person home or to volunteer to work with disabled persons). Clearly, assessment of attitudes was dominated by questionnaires, especially if all paper-and-pencil assessments, such as social distance and semantic differential results, are included with the typically Likert scale questionnaire results. Only 12 Ds (2%) came from nonpaper-and-pencil assessments (interviews, observation, telephone surveys).

Some changes in assessment are of interest. Two of these are the relative dropoff in questionnaire use in 1981-86, accompanied by an increase in use of adjective checklists. Also, the cause of the brief flare of semantic differential use in reports that became available in 1976-80, nearly 20 years after the landmark Osgood, Succi, and Tannebaum (1957) publication, raises intriguing questions about the acceptance, use, and reporting of research techniques.

Data Collection

Methods of data collection can affect study outcomes. One particularly relevant question is whether those who administered pretests and posttests were blind to the purpose of the study and to the experimental group membership of the Ss to whom they administered assessments. The convention for coding whether blinded data collection occurred was based on the assumption that researchers will usually report those aspects of their methodology that are especially valued in treatises on research design. Random sampling and assignment are such high priority procedures. Blinded data collection is another. So, if there was no mention that test administrators were blind to the purpose of the research or to the Ss' group membership, it was presumed not to have occurred.

For 663 out of 705 effect sizes (94%), "No" was coded for blinded collection. For 27 effect sizes (4%), blinded data collection was obvious, with partial blinding (information kept from coders as to group membership or whether a pre or posttest was being administered) for 5 effect sizes (nearly 1%). Enough information was provided for 9 effect sizes (1%) to make the rater unsure, and "Can't tell" was coded.

Test scoring is an essential part of data collection, and the blinded scoring of tests is desirable in cases where test scorers must draw inferences. As already noted above, paper-and-pencil, questionnaire-type assessments were the predominant mode of dependent measure. With such assessments, coding is routine. It is not surprising, therefore, that for 674 of 705 effect sizes (96%), the category on blinded test scoring was coded as "Not Applicable". For 11 effect sizes (nearly 2%) for which blinded score was pertinent, it was done; for 14 effect sizes (2%), it was not.

Reliability

The reliability of the scores obtained on dependent measures is a central concern in research, as low reliability has a negative impact on validity as well as attenuating group differences. It is surprising, therefore, that for 44 percent (N = 293) of the 705 effect sizes (see Table 16), no reliability coefficient was reported. (For 9 effect sizes, the adequacy of reliability was asserted, but no coefficient was reported.) For those effect sizes for which a reliability coefficient was reported, 187 coefficients (26% of the 705 effect sizes) fell in the range of .80 to 1; 200 coefficients (28%) in the range of .60 to .79; and 25 (3%) below .60.

Many of the researchers were apparently not mindful that reliability is an attribute of scores, not of tests, and that coefficients can vary widely by population and test administration circumstance: About 64% of the reported reliability coefficients came from other studies. (For example, when the ATDP scale was used, it was common for the authors to cite the reliability figures given in Yuker et al. [1970], and not to report a coefficient for their sample.) Another 9 percent of the reported coefficients came from pilot studies. For 4 percent of the effect sizes for which a reliability coefficient was reported, no source could be discerned. Only 24% of the coefficients reported (for 14% of the effect sizes) were computed for the samples of Ss studied.

Validity

Did test score validity fare any better in our population of studies? A crucial starting point in the consideration of the validity of measures for assessing attitudes would seem to be definition of the construct, "attitude". However, Towner (1984) lamented the lack of definitions of "attitude" in the modification studies she reviewed. In our data set, no definition of "attitude" was given in 114 of 215 studies (53%), accounting for 317 of 705 effect sizes (45%) (see Table 17). For those effect sizes for which a definition was presented, a conception of attitudes as having affective,

Table 16

Magnitude of Reliability Coefficients
for Dependent Measures in Effect Size Comparisons

Magnitude	Effect Sizes N	Effect Sizes %
None reported	293	42
.00 - .60	25	3
.60 - .79	200	28
.80 - 1.00	187	26
Total	705	99

Table 17

Definitions of "Attitude"
in Research Reports

Type of Definition	Effect Sizes N	Effect Sizes %	Reports N	Reports %
None	317	45	114	53
Affective	65	9	22	10
Cognitive	9	1	4	2
Behavioral	5	1	3	1
Affective and Cognitive	88	12	21	10
Affective and Behavioral	13	2	3	1
Affective, Cognitive, Behavioral	208	29	48	22
Total	705	99	215	99

cognitive, and behavioral components was most common (N = 208; 54% of 388 effect sizes). A definition that involved affective and cognitive components was next in frequency (N = 88; 23% of 388 effect sizes), followed by an exclusive emphasis on affect (N = 65; 17%). The other definitions (Cognitive, Behavioral, Affective and Behavioral; N = 27) constituted 7% of the effect sizes for which definitions could be coded.

Given the large proportion of effect sizes for which the object of the experimental treatment, "attitudes", was not defined, it is not surprising that for 374 out of 705 effect sizes (53%), the validity of scores on the dependent measure was not discussed*. For only 32 effect sizes (4%) was there extensive discussion of test validity. Moreover, for 94% (N = 660) of the effect sizes the dependent measure was coded as "high" in reactivity (with "low" [1%] and "moderate" [5%] the other choices). The adequacy of validity was rated as "moderate" for 573 effect sizes (81%), "low" for 122 (17%), and "high" for only 10 effect sizes (1%).

Time of Posttest

Changes in attitudes toward persons with disabilities must be sustained, not temporary, to be of consequence. Towner (1984) called for testing to determine the "longterm effects" of treatments to modify attitudes (p. 254). Our data set confirms the need for that call. For 476 effect sizes (67%), an immediate posttest was the source of data. For 90 effect sizes (13%), the posttest was not immediate, but was delayed as much as a week to obscure the connection with the treatment. Only 89 effect sizes (13%) were based on follow-up posttesting--i.e., testing that followed an initial posttest. (For

*With attitudes clearly a psychological construct, it was perplexing that several authors (e.g., Lapp, 1974; Ozyurek, 1977) referred to the "content validity" of their attitude measures.

7% of the effect sizes, the rater could not determine when the posttest was administered.)

Use of Theory

Given the scant attention paid to attitudes as a construct and to the validity of the attitude assessments used, it would have been surprising to find careful attention given to the theoretical bases for the attitude modification techniques investigated in the various studies. Towner (1984) noted that a majority of the reports of attitude modification she reviewed did not indicate a theoretical base for the approach taken to attitude change. In our data set, only 194 effect sizes out of 705 (27%) came from comparisons in which an attitude change theory was the explicit basis for the experimental treatment. The most common basis was prior research (N = 403; 57%), with the case "well developed" for 308 effect sizes (76% of 403), with "few citations of prior studies" for 91 effect sizes (23% of 403), and with prior research "mentioned but not cited" for 4 effect sizes (1%).

As Table 18 indicates, the predominant theory either used explicitly as a base for a treatment (194 effect sizes; see paragraph above) or implicit in the intervention (as judged by the rater with no direct evidence in the report; 458 effect sizes of 705, or 65%) was the consistency-equilibrium theory associated with theorists such as Festinger, Heider, Lecky, Levin, McGuire, and Newcomb (see the conventions for Category C.2.a. in Appendix C for theory definitions). The data in Table 18 must be interpreted with caution, however, in light of the large number of effect sizes for which the theoretical bases for the modification technique had to be inferred. The most apt generalization is probably that the research on modifying attitudes toward disabled persons is largely atheoretical.

Table 18

Attitude Change Theories
Underlying Experimental Treatments

Theory	Effect Sizes	
	N	%
SR, Behavioral	29	4
Conditioning	24	3
Consistency/equilibrium	518	73
Social Judgment	14	2
Functional	59	8
Combination	61	9
Total	705	99

It was, in particular, puzzling to find no attitude change procedures based on Rokeach's (1973) version of balance (consistency-equilibrium) theory. His approach to attitude change is to induce self-dissatisfaction with values, as a means to value change, attitude change, and behavioral change. The approach has been applied successfully in areas as diverse as civil rights (Rokeach, 1971, 1973; Rokeach & McLellan, 1972), teaching behavior (Greenstein, 1976), and women's rights and the environment (Ball-Rokeach, Rokeach & Grube, 1984). It merits attention by those interested in affecting not only modifications in attitudes, but changes in behavior toward persons with disabilities.

Study Populations and Samples

There is some evidence that educational researchers do not often address specifically in their reports the nature of their target or accessible population, nor draw random samples from their accessible populations, make random assignments to treatments, or replicate their results to establish their stability and generalizability (Shaver & Norton, 1980a, b). Is that statement applicable to the body of research on modifying attitudes toward persons with disabilities?

Table 19 indicates that those doing research in this area have addressed population issues even less often than those who have published in two social studies journals and in ten years of the American Educational Research Journal (AERJ). For a majority of the effect sizes in the reports coded for this review, there was no mention of the groups to which the authors hoped their results would be generalizable (target population--73%) or from which their samples came (accessible population--61%). In fact, few authors even used that terminology. For 7 effect sizes (1%), the term "target population"

Table 19

Treatment of Target and
Accessible Populations
for 705 Effect Sizes

	Target Population				Accessible Population			
	Effect Sizes		% Reports[a]		Effect Sizes		% Reports[a]	
Category	N	%	Social Studies	AERJ	N	%	Social Studies	AERJ
Not mentioned	512	73	45	67	432	61	17	49
Term used	0	0	0	1	3	0.4	0	1
Defined	186	26	55	32	193	27	72	41
Described	0	0	0	0	0	0	11	8
Term used and Population defined	7	1	--	--	77	11	--	--
Total	705	100	100	100	705	99.4	100	100

[a]Percentages from Table 2 in Shaver and Norton (1980a), based on 53 research reports in all issues of two social studies journals through 1978 and 151 reports in the *American Educational Research Journal* (*AERJ*) for ten years, 1968-77.

was used and the population was defined. For 3 effect sizes (.4%) the term "accessible population" was used, and for 77 other effect sizes (11%), the term was used and the population defined in at least rudimentary terms.

By the same token, as Table 20 shows, random sampling of individual Ss was rare. It was the basis for sample selection for 31 effect sizes (4%)*. The random selection of groups provided the Ss for 31 effect sizes (4%). The use of intact groups was the most common means of obtaining a sample (N = 327 effect sizes; 46%). The use of volunteers was common (N = 237 effect sizes; 34%), and greater than for Shaver and Norton's (1980a) sample of AERJ reports (9%) and social studies reports (24%).

Table 21 presents information on assignment to groups, with the 60 Treatment A vs. B effect sizes not included. Random assignment of the individuals or groups used as the unit of analysis was done for 35% of the effect sizes (N = 227), including 21 (3%) instances of matching followed by random assignment. This is almost identical to the 35% of reports of random assignment in Shaver and Norton's 10-year AERJ sample and considerably above the 9% for the reports in their social studies research sample (Shaver & Norton, 1980b)**.

Replications

Related to the task of defining the populations from which samples are drawn and to which one wants to generalize is the matter of replication, as

---

*This compares to 15% and 19%, respectively, for the samples of reports from two social studies journals and AERJ reported by Shaver and Norton (1980a). The Shaver and Norton data are not reported fully in Table 20 because different categories were used.

**Information from Shaver and Norton (1980a) was not included in Table 21 because different categories were used.

Table 20

Sample Selection for
705 Effect Sizes

	Effect Sizes	
Category	N	%
Can't tell	43	6
Random—Individuals	31	4
Random—Groups	31	4
Volunteer	237	34
Intact Groups	327	46
Other	36	5
Total	705	99

Table 21

Assignment to Treatment Groups for 644 Effect Sizes

Category	Effect Sizes N	Effect Sizes %
Can't tell	23	4
Random	206	32
Match-random	21	3
Select controls randomly or matched	3	0.5
Intact groups--randomly[a]	130	20
Convenience	154	24
Other	23	4
Not applicable[b]	84	13
Total	644	100.5

[a]Intact groups assigned randomly, but not used as unit of analysis. If assigned randomly and used as unit of analysis, coded as "random".

[b]Single-group studies.

it is often argued to be the basic scientific means of establishing the reliability and generalizability of results (e.g., Shaver, 1979a). As inspection of Table 22 reveals, replications have not been a common feature of studies in modifying attitudes toward disabled persons. About 1.5 percent of the effect sizes came from efforts to replicate other studies. About 12 percent of the effect sizes came from within-study replications; however, almost one-fourth of those 87 effect sizes were "quasi-replications"—effect sizes based on data gathered from different samples or in different settings in the study and coded separately even though the researchers did not recognize them as replications.

Replicability

It is noteworthy as well that for 290 of the 705 effect sizes (41%), the description of the treatment variable was not deemed adequate to allow another researcher to replicate the study (Category C.8.e.). For 111 effect sizes (16%), description was deemed adequate for replication; and for 304 (43%), description was judged to be "somewhat" adequate.

A treatment must first be implemented to be replicated later. However, for only 37 effect sizes (5%) was the actual implementation of treatment rated as "complete" (Category C.8.d.). For 630 effect sizes (89%), implementation was judged to be "mostly" complete, and for 38 effect sizes (5%), the treatment was rated as implemented "only in part". At the same time, for 639 effect sizes of 705 (91%), no report was made of an effort to verify that the treatment had been implemented as intended.

Qualification of Results

In their critiques of research reporting, Shaver and Norton (1980a, b) gathered data on the extent to which the authors of research reports

Table 22

Replications Among 705 Effect Sizes

Type of Replication	Effect Sizes N	Effect Sizes %
Other Research		
None	696	99
Direct	3	0.4
Systematic	6	1
Total	705	100.4
Within Study		
None	618	88
Direct	7	1
Systematic	80[a]	11
Total	705	100

[a]Includes 21 "quasi-replications"--that is, studies in which the treatment was repeated on different samples or in different settings and the results were coded separately, even though not treated as a replication by the researchers.

188

restricted their conclusions in terms of the shortcomings in their accessible populations (as related to the target population) or samples. They found conclusions tempered by accessible population deficiencies in 15 percent of their social studies research reports and in 7 percent of their AERJ articles; 23 percent of the social studies reports took sampling inadequacies into account in drawing conclusions, while 11 percent of the AERJ articles did. We broadened the question for this review, asking if conclusions were qualified by reference to sampling or design problems, possible interactions of personological or ecological variables with the experimental treatment, the assessments used, the need for replication, or "other" considerations.

Table 23 presents the results. As can be noted, for 66% of the effect sizes, the authors provided some limitation on their conclusions about the effectiveness of the technique for attitude modification. The 14 percent of qualifications based on the sampling process is close to those of the Shaver and Norton study (see preceding paragraph). Interestingly, the largest percentage of qualifications (for 260 effect sizes, 37%) took into account combinations of factors. That is encouraging, although the 34% with no qualifications is an offsetting concern.

Bases for Effect Sizes

A relevant methodological matter for those interested in meta-analysis as an approach to integrative literature reviews is the type of information that is available for computing the effect sizes--in our review, Ds--which are the center of attention. As noted in the discussion of our coding instrument in Chapter 3, we gathered information on the source of each effect size, and on the scale of mean difference and the standard deviation used on computing each one. Tables 24-26 present data on the bases for the 645

Table 23

Limitations on Conclusions About Treatment Effects for 705 Effect Sizes

Limitation	Effect Sizes N	%
None	242	34
Sampling	100	14
Design	66	9
Measures	21	3
Interactions	5	1
Need for replication	2	0.3
Other	9	1
Combination	260	37
Total	705	99.3

effect sizes for our treatment versus control, treatment versus placebo, and single-group, pre-posttest comparisons.

The major source of effect sizes was the calculation of Ds from data available in the research reports (see Table 24). Approximately 90% (N = 578) of our Ds were obtained in that way. The other 10% were estimated using statistics such as t-ratios and F-ratios. (When a D was estimated from a correlated t or an F from an analysis of covariance, but the pre-post or covariate-dependent measure coefficient was not available, a coefficient of .50 was used. See Appendix B for effect size computation conventions.) Although a subcategory was included on the instrument ("Available" in Table 24) for coding that a standardized mean difference was available in the report, no use was made of it. It is interesting as well that in every use of statistics to estimate Ds, except for the 9 Ds computed from analysis of covariance F-ratios using .50 as the estimated coefficient, the mean D was higher than the mean for Ds computed directly from the data. Fortunately, with 90% of the Ds computed directly that bias had little effect on the analysis of data.

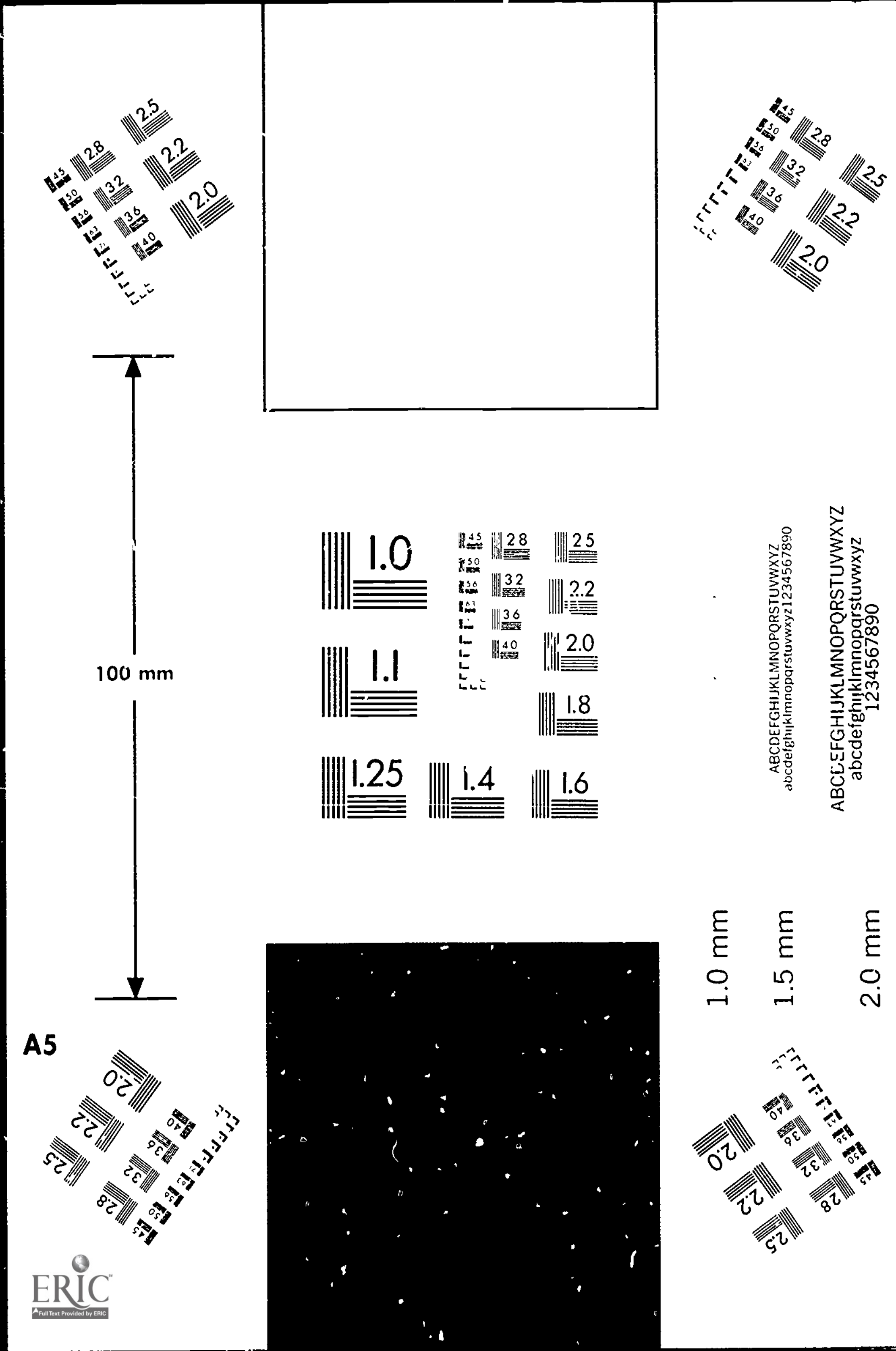
Another methodological feature of our data set having to do with the basis of our effect sizes is the information available for scaling mean differences. It was decided in developing the coding instrument that the preferred scale would be raw mean gain scores, because they are a scale of the same units as final status scores (Glass et al., 1981, p. 116) and they avoid the problems of covariance-adjusted scores when the assumption of homogeneous regression lines is not met. If raw gain means were not available, raw posttest means were the next choice, followed by covariance adjusted means, and then by residual gain means.

Table 24

Sources of 644 Effect Sizes (*Ds*)

	Effect Sizes (*Ds*)		
Source	N	Mean	SD
Availability	—	—	—
Calculated	577	.36	.60
t or ANOVA F	18	.46	.84
Correlated t	6	.40	.24
Correlated t (.50)	15	.58	.52
N-Way ANOVA	3	.42	.46
COVAR (.50)	9	.15	.20
Proportions--χ^2	12	.74	1.00
Significance level	2	1.23	.00
Other	2	.90	.01
Total	644	.37	.61

Note. $E^2 = .02$.


100 mm
A5
1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
1.0 mm
1.5 mm
2.0 mm
ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz1234567890
ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz
1234567890
ERIC
Full Text Provided by ERIC

As can be seen in Table 25, 387 of the Ds (60%) in our data set were based on raw mean gains. Raw posttest means were the basis for 163 Ds (25%) and covariance-adjusted means for 28 Ds (4%). No residual gain means were used. (The Not Applicable category [N = 66] encompasses the effect sizes not calculated, but estimated from available statistics.) The proportion of variance in Ds associated with the type of mean difference scale (Eta^2) is only .01. The Not Applicable mean D is the one discrepant value; and, again, with only 10% of the Ds falling in that category, the effect of the bias is minor.

For calculating standardized mean differences, as noted in Chapter 3, a standard deviation based on as much data as possible that was free of treatment effects was preferred. Pooled pre and posttest standard deviations (see Table 26) were obtained for computing 248 Ds (38%), pretest standard deviations were used for 108 Ds (17%), and posttest control or placebo group standard deviations were used for 140 Ds (22%). Estimates of standard deviations from various sources (within-group variances, or gain or covariance adjusted scores) and pooled posttest variances were used in computing 86 (13%) of the Ds. Although the mean Ds based on these estimated standard deviations tend to be higher than those from sources free of treatment effects (but similar to the Not Applicable mean D), the Eta^2 is only .02. The small proportion of variance in the Ds that was associated with the source of standard deviations is a reflection of the small numbers of standard deviations that come from within-group variances.

Frequency of Can't Tell Scoring

One other methodological aspect of the data set that may have general interest is the frequency with which Can't Tell was coded for the various threats to treatment validity and internal validity (see discussion in

Table 25

Scale of Mean Differences
For 644 Effect Sizes (Ds)

Scale	Effect Sizes (Ds)		
	N	Mean	SD
Not Applicable	66	.51	.69
Raw Post	163	.29	.68
Raw Gain	387	.39	.57
COVAR-Adjusted	28	.36	.39
Total	644	.37	.61

Note. $E^2 = .01$.

Table 26

Standard Deviations for 644 Effect Sizes (Ds)

Source	Effect Sizes (Ds)		
	N	Mean	SD
Not Applicable	62	.53	.70
Post Control/Placebo	140	.24	.67
Pretest	108	.34	.76
Post & Pre pooled	248	.37	.47
1-way ANOVA	15	.53	.75
N-way ANOVA	21	.45	.44
1-way COVAR	12	.47	.58
N-way COVAR	28	.61	.47
Adjusted COVAR	1	.51	.00
Adjusted gain	4	.58	.46
Pooled post	5	.43	.54
Total	644	.37	.61

Note. $E^2 = .02$.

Chapter 3). Tables 27 and 28 contain that information for the 644 effect sizes in our main analyses (see Chapter 5). As can be seen, the availability of information necessary to decide if a threat was present varied greatly among the threats. It is also clear that forced judgments about the presence of threats would have been largely speculative for a large number of effect sizes. It was our judgment in coding, supported by the data in Tables 27 and 28, that essential information is frequently missing from reports for determining whether threats to treatment and internal validity existed in the research. We recommend that others doing meta-analyses use categories to pick up the extent of missing information, as our Can't Tell scoring did.

Summary

Although this chapter contains discussions of some coding and analysis decisions in this chapter, the primary purpose was to sketch some dimensions of the available body of research on modifying attitudes towards persons with disabilities.* From all of the data available, some were selected to provide a flavor of the volume of research, where it has been reported, the types of attitude modification techniques investigated, with what groups of people, and to change attitudes toward what disabilities. Some methodological attributes were described as well--types of treatment comparisons, the adequacy of assessment and the use of theory, sample selection and assignment, the use of replications and the implementation and verification of treatments, the extent to which conclusions took into account study limitations, the bases for computing Ds, and the extent of "Can't Tell"

*As noted at the end of Chapter 3, the entire data set is available on magnetic tape and can be obtained at cost by anyone who wishes to explore further the characteristics of the reported research in this field.

Table 27

Threats to Treatment Validity for 644 Effect Sizes

	Category																	
	Can't Tell			Not Plausible			Minor			Substantial			Major			Not Applicable		
Threat	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Implementation	—	—	—	33	.47	.63	582	.38	.61	30	.18	.58	—	—	—	—	—	—
Hawthorne	160	.35	.58	93	.29	.62	234	.42	.65	157	.39	.57	—	—	—	—	—	—
John Henry	254	.40	.59	262	.36	.64	30	.22	.32	13	.11	.39	—	—	—	85	.42	.66
Treatment Diffusion	239	.36	.71	142	.29	.50	131	.46	.55	45	.35	.41	5	1.06	.87	82	.38	.62
Dissatisfaction/ Resentment	246	.31	.53	303	.39	.64	44	.53	.71	5	-.25	.53	5	.27	.25	41	.52	.74
Novelty/Disruption	183	.41	.58	382	.36	.59	52	.42	.92	7	.03	.11	1	.41	.00	19	.26	.35
Experimenter Effect/ Expectations	218	.37	.56	23	.53	.63	130	.35	.70	241	.37	.59	31	.36	.76	1	.75	.00
Treatment-experimenter Confounded	82	.36	.51	37	.34	.61	98	.36	.75	221	.52	.64	205	.23	.50	1	.75	.00
Testing by treatment interaction	24	.54	.76	10	.61	.81	151	.37	.57	426	.37	.62	33	.27	.38	—	—	—
Multiple treatment interference	573	.37	.61	62	.50	.65	1	.52	.00	8	-.10	.28	—	—	—	—	—	—

Note. Means and standard deviations are for *D*s.

Table 28

Threats to Internal Validity for 644 Effect Sizes

	Category														
	Can't Tell			Not Plausible			Minor			Substantial			Major		
Threat	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Maturation	326	.43	.63	272	.29	.58	9	.45	.45	37	.46	.65	--	--	--
History	456	.32	.55	66	.47	.69	56	.38	.66	62	.64	.79	4	.60	.36
Testing	74	.32	.50	433	.34	.60	38	.53	.56	60	.43	.56	39	.62	.92
Instrumentation	35	.58	.93	29	.29	.43	566	.36	.59	12	.33	.42	2	1.66	.23
Statistical Regression	25	.68	.84	592	.35	.59	18	.55	.60	9	1.02	.65	--	--	--
Selection	52	.38	.58	197	.41	.70	65	.36	.46	236	.38	.61	94	.28	.53
Experimental Mortality	153	.30	.46	303	.44	.66	97	.38	.70	66	.28	.56	25	.25	.41

Note. Means and standard deviations are for *D*s.

coding for treatment and internal validity. The intent was to set a context for the reporting of our analyses to determine the effects of attitude modification techniques as they can be discerned from the available literature. That is the subject of the next chapter.

CHAPTER 5

ATTITUDE MODIFICATION TECHNIQUES: THE RESULTS

The sketch of our data set in the prior chapter does not address the central objective of this review of literature: to determine whether different techniques for modifying attitudes toward persons with disabilities have yielded different effect sizes. Before turning to analyses that address that item more specifically, some prefatory comments are appropriate in regard to differences between data analysis for an integrative review as compared to the typical data analysis approach for primary research studies.

Perspective on Analysis

Once the data are collected for a quantitative, integrative review of the research literature, the next steps, as in a primary research study, are to analyze the data and report the results. Data analysis is, however, not so straightforward for a quantitative review as for most primary experimental research. In conducting primary experimental research, there are usually few variables to take into account--a treatment variable with two or three levels, one or perhaps two dependent measures, and maybe a covariate, or classification variable, or two. The analysis can be pretty well specified beforehand, with the only open question likely to be whether the correlation of a covariate with a dependent measure is sufficiently high to justify an analysis of covariance. Beyond that, the researcher conducts an analysis (using a technique such as ANOVA or COVAR), checks for statistical significance, does any contrasts of pairs of means that are appropriate, and reports the results in a table or two accompanied by a brief discussion of whether statistical significance (usually the conventional .05 level) was attained. Occasionally now, the researcher also reports an effect size (standardized mean difference or correlation coefficient) to back up the inferential statistics results.

All of the above is accomplished via a relatively noncircuitous route; once the preplanned analysis is set in gear, few decisions are necessary about the paths to follow. The decisions to be made, rather automatic ones, are such as whether to include a covariate, whether a finding attained statistical significance, whether post hoc contrasts should be made, and, on occasion, whether the results from analyses with inferential statistics square with those from effect size computations.

This admittedly somewhat overdrawn picture of the simplicity of analysis in primary research is in marked contrast to the analysis in an integrative review such as reported here. To begin with, there is a greater number of variables to take into account. The number of what might be termed "treatment levels" may not appear large at first glance. For example, we coded six attitude change techniques, plus combinations (Category C.4.). But rather than a defined treatment (e.g., "information" or "contact") applied in a particular setting, as in a typical research project, the review analyst has a number of applications of the treatment with differing characteristics across applications. As a result, the number of possible treatment variations is large. The number becomes larger when variations in sample attributes and perhaps even ecological variables are taken into account. Of course, the complexity is further compounded when different dependent measures are used across primary research studies, and even further by the consideration of research quality variables. With some 140 categories for coding the research reports, as in this review, the number of possible combinations is astronomical.

One solution would be to use a much smaller number of categories to code the primary research reports, but that would be akin to the dysfunctional primary research strategy of not attending to interactions between treatment and personological and ecological variables because they make interpretation difficult. Instead, we were guided in setting up the coding instrument (as discussed in Chapter 3) by the admonition to include "all characteristics of the primary studies that are strongly suspected of affecting the findings . . ." (Jackson, 1978, p. 57). The upshot is a complex analysis process with difficult decisions about what to report and how. One major issue is how to handle data on the methodological quality of the studies in a data set.

Quality of Research

The methodological quality of the studies from which effect sizes are collected has been a source of concern since Glass (1976) first proposed the use of the meta-analytic approach to integrative reviews. Glass's proposal to include studies regardless of quality and analyze for the effects of quality (discussed in Chapter 3) drew a "garbage in-garbage out" criticism from Eysenck (1978) that is often cited in discussions of meta-analysis. Although the concept of analyzing for quality effects is still controversial (Bangert-Drowns, 1986), our stance in planning the procedures for this review was the same as Glass's: That is, include all studies, code for quality, and determine if effect sizes covary with study quality.

Quality Indicators

Although a number of our coding categories are related to quality of study, three global categories are particularly appropriate indicators of methodological soundness: general treatment validity, general internal validity, and adequacy of test validity. Each is widely regarded by researchers to be central to the validity of experimental results, and each is based on information from other categories.

Internal validity is, for example, in part a function of how Ss are assigned to treatments. In fact, some researchers regard studies with random assignment to be, ipso facto, of high internal validity. However, as we noted in Chapter 3, that is not the case. Random assignment may be a sine qua non of excellent internal validity (as indicated by the fact that random assignment was used for all but two of the effect sizes coded in our review as coming from studies with high internal validity), but it does not guarantee validity. Even selection as a threat to internal validity is controlled by random assignment only within chance limits. That is, groups with quite different antecedent characteristics can, and do, occur even with random assignment, as statistical theory indicates should be expected. So, rather then consider the absence or presence of random sampling per se as an indicator of study quality, we coded method of assignment and took that into account in coding the various threats to internal validity (see Table 28 in Chapter 4) that were weighed in coming to a judgment about the internal validity of each study.*

Results

Summary statistics for the three global indicators of quality are presented in Table 29. Two attributes of the data are striking: First, few studies received high or excellent ratings on any of the three types of

---

*The Cramer's V for internal validity and type of assignment to groups was .44, a moderately high relationship, as would be expected.

177

Table 29

Quality of Study Indicators

General Treatment Validity

Quality	Effect Sizes (Ds) N	Mean	SD
Excellent	4	.25	.30
Fair	245	.45	.68
Poor	395	.33	.56
Total	644	.37	.61

Note. $Eta^2 = .01$

General Internal Validity

Level	Effect Sizes (Ds) N	Mean	SD
High	15	.89	.87
Medium	211	.3[illegible]	.58
Low	418	.38	.61
Total	644	.37	.61

Note. $Eta^2 = .02$

Adequacy of Test Validity

Adequacy	Effect Sizes (Ds) N	Mean	SD
High	9	1.13	.69
Moderate	520	.36	.62
Low	115	.40	.55
Total	644	.37	.61

Note. $Eta^2 = .02$

global validity. Second, none of the ratings of validity explain much of the variability in effect sizes (as indicated by the Eta^2s of .01, .02, and .03). The low correlation between quality ratings and Ds is at least in part a function of the lack of variability in the former: few effect sizes came from studies with excellent or high ratings.

To determine the association between Ds and membership in the higher frequency medium and low quality categories, point biserial coefficients were computed. The squared coefficients are .004, .01, and .001 for treatment, internal, and test validity, respectively, again indicating that very little variance in Ds was associated with quality ratings.

Why Few High Ratings?

The lack of studies rated to be high on the quality indicators could be a function of the actual quality of research in the field or of invalid ratings. In regard to the latter possibility, it is worth noting that the raters were able to apply the research quality conventions reliably; readers will have to peruse our discussion of our codin strument in Chapter 3 and the conventions (Appendix C) to determine if we were over-rigorous in our definitions of excellent or high validity. We think not.

After reading the reports of the 273 studies that made up our population, it is our conclusion that the quality is actually not very high in general, due to two factors. The first is that attitude research is difficult to conduct, especially in applied settings (e.g., in elementary schools) rather than laboratories. There are major impediments to gaining adequate control over subject and ecological variables in applied settings. Moreover, major issues remain to be addressed before adequate instruments for assessing attitudes toward persons with disabilities will be readily available (see, e.g., Makas, 1985). Even careful researchers are sorely tested to produce studies that have no flaws in internal validity when faced with the challenges of applied psychological research.

Beyond the difficulty in conducting attitude change research, however, another reason for the lack of high quality ratings accorded to the studies in our sample is simply poor design and execution (as well as inadequate reporting, if better methodology was used than we were able to discern). Some examples from our data illustrate the point: For some 65% of the effect sizes, randomization of Ss was not reported; in 24%, conveniently available groups were used for the treatment and control conditions. For only about 4% of the effect sizes were data collectors either fully or partially blinded. For 41% of the effect sizes, there was no mention of a reliability coefficient for the dependent measure scores; and, for the 47% of the effect sizes for which reliability coefficients were reported, only 24% had been computed on the study sample (64% came from other reports of research). For 93% of the effect sizes, the reactivity of the attitude assessment was judged to be high. And, for 83% of the effect sizes, the research reports contained no mention of any effort to verify implementation of the treatment independent variable. It would be difficult to argue that the available body of research on modifying attitudes toward persons with disabilities is exemplary in methodology.

A certain irony is hinted at in the data reported in Table 29. Frequently, internal validity is construed as contributing to large group differences that are inadvertently attributed to treatment effects. Although it is recognized that low test score reliability attenuates mean differences (as well as correlation coefficients), obscuring true treatment-outcome relationships, it is often not recognized that other threats to experimental validity can have the same effect. To the extent that studies are poorly designed and executed, treatments may be less powerful or differences otherwise obscured by threats to validity. While the number of effects from studies with "excellent" treatment validity was too small (N = 4) in our data set to be interpretable, the results from the somewhat larger numbers of effect sizes for high internal and test validity (15 and 9) indicate that higher mean Ds may come from high quality studies. Even in those two cases, however, the numbers are so small that it is difficult to place confidence in the stability of the findings. Nevertheless, they raise interesting implications for validity-outcome relationships.

A "Best-evidence" Approach

In any event, the results with treatment, internal, and test validity pose a quandary. On the one hand, there appears to be little association in our data set between the magnitude of Ds and the quality of the studies from which they come, at least as assessed via these global indicators. The logical conclusion is, therefore, that quality of study need not be a consideration in analyzing the data; that is, it did not produce much of the variability in results for different attitude modification techniques which can be seen in Table 34. On the other hand, it can be argued (see, e.g., Bangert-Drowns, 1986, p. 392) that unless the studies being reviewed vary widely in methodological rigor, it makes little sense to examine study quality-outcome relationships. In particular, without a sufficient number of high-quality studies, it can be contended, one lacks an adequate comparative base for determining whether or how study quality affected results. For our data set, we only have a hint that the Ds might have been higher if the studies from which they came had been carried out with greater methodological rigor.

If this review were being conducted from a non-Glassian stance that studies with methodological flaws should be excluded from the analysis, our data set would shrink appreciably. Moreover, given the lack of association between effect sizes and moderate or low quality categorization, there would be little reason to discard the low quality studies and analyze only effect sizes from the medium quality ones.

What might seem like a readily apparent solution to some--that is, not attempt any integrative review--is argued against by Slavin (1986) in his proposal for "best evidence" research syntheses. Clearly, if high quality studies are available, they should be relied on in an integrative review. But if such studies do not exist, it is appropriate to "cautiously examine the less well designed studies to see if there is adequate unbiased information to come to any conclusion" (p. 6).

Slavin argues that the application of a prior criteria in selecting "best evidence" studies (rather than Glassian exhaustive inclusion, followed by quality-outcome analyses) is at the heart of the "best evidence" synthesis approach (p. 6). He does not, however, present any compelling reason to reject as "best evidence" studies that have been selected because they are topic-relevant but turn out not to yield an adequate basis for determining the association between study quality and results. The biggest impediment in such a best-evidence review would be exercising the necessary caution without any basis for deciding which information is "unbiased" and, therefore, legitimate to use in drawing conclusions.

We have proceeded with our analysis in a form of "best-evidence" review which Slavin obviously did not intend to support. As Bangert-Drowns (1986) has pointed out, such a decision depends in large part on the purpose of the integrative review. An appropriate goal, rather than to construct theory or estimate treatment parameters, is to characterize the available research as a basis not only for insights into treatment effectiveness, but for decisions about further research. With those objectives, careful summarization of past research has a significant place, even if only to make evident that which remains to be done.

Overall Effects

In Chapter 4, we discussed the decision to focus our analysis on effect sizes that came from comparisons of attitude modification techniques against the absence of treatment (i.e., a control or placebo group, or pre-treatment scores). Unless specifically noted, the reports of effect sizes that follow come from treatment versus control (T vs. C), treatment versus placebo (T vs. P), or single-group, pre-posttest (Pre-post) comparisons.

Some preliminary information is pertinent to the general perception of the treatment effect sizes. One such bit of information is the research report authors' own views of the effectiveness of their attitude change treatments. Table 30 indicates nearly a balance between the number of effect sizes for which the authors concluded the treatment was effective (N = 285; 44%) and those for which the treatment was deemed not to have had an effect (N = 259; 40%). Considered with the 40 effect sizes (6%) for which the results were considered equivocal, along with the 19 (3%) effect sizes for which it was concluded that the effect was negative, the summary of conclusions suggests that it should not be easily assumed that the use of just any modification technique will lead to a positive effect. Note, too, that the mean Ds parallel closely the conclusions drawn, with means of .74

Table 30

Research Report Authors'
Conclusions *re* Treatment Effectiveness

	Effect Sizes (*Ds*)			
Conclusion	N	%	Mean	SD
None stated	42	6	.34	.41
No effect	258	40	.03	.32
Equivocal	40	6	.51	.49
Produced effect	284	44	.74	.61
Negative effect	20	3	-.63	.36
Total	644	99[a]	.37	.61

Note. $Eta^2 = .37$.

[a]On this and later tables, percentages may not always add up to 100 because of rounding error.

212

for "produced effect", .51 for "equivocal", .03 for "no effect", and -.64 for "negative effect". The Eta^2 of .37 indicates a moderate association between the conclusions drawn and the magnitude of effect sizes.

Along the same lines, there was a small association ($Eta^2 = .22$) between the availability and level of statistical significance and the Ds obtained (see Table 31). The 173 effect sizes (27%) for which statistical significance was not available yielded a mean D of .41. As expected, the mean D for statistically significant effect sizes was much higher than that for nonstatistically significant ones, .73 versus .06. The point biserial coefficient for magnitude of Ds and membership in the "not significant" and "significant at .05" groups is .47 ($r^2_{pb} = .22$).

Relevant to the number of conclusions about negative effects (Table 30) are the number of effect sizes for which an attitude modification treatment group showed a negative change. As Table 32 reveals, for 12 percent of the effect sizes (N = 77) the treatment group's posttest mean was lower than its pretest mean, producing a mean D of -.19. It should be noted that a "negative change" by a treatment group is not the same as a "negative D", except for single-group, pre-posttest comparisons. If the control or placebo group had a greater negative change, the D would be positive; also, the treatment group might have had a positive change that was less than the positive change of the control or placebo group, yielding a negative D.

Comparisons of Experimental Treatments

What about the outcomes of the comparisons of experimental treatment groups against control or placebo groups or pretest scores? The various treatment techniques and combinations of techniques are briefly described in Table 33. They are arranged in rank order in Table 34, according to the

Table 31

Numbers of Effect Sizes for Which Statistical Significance Information was Available

Statistical Significance	Effect Sizes (Ds) N	Mean	SD
Not Available	173	.41	.46
Not Significant	258	.06	.34
Significant at .05	213	.73	.75
Total	644	.37	.61

Note. $Eta^2 = .22$.

Table 32

Changes by Attitude Modification Treatment Groups

Direction of Experimental Group Change	Effect Sizes (Ds) N	%	Mean	SD
Positive	567	88	.45	.60
Negative	77	12	-.19	.35
Total	644	100	.37	.61

Table 33

Brief Descriptions of Attitude Modification Techniques as Coded

Technique	Description
Information	Information on disabilities (e.g., etiology, characteristics, problems, similarities with nondisabled, prostheses) provided by means such as speakers, films, and books
Direct Contact	Ss in situation where they observe or interact with persons with disabilities
Vicarious Experience	Ss put in situations to help them experience what it is like to have disabilities
Persuasive Message	An argument presented via persons or printed or electronic media to convince Ss that they should have positive attitudes toward persons with disabilities
Persuasive Message, Contrast	Different messages or media used with treatment groups to investigate relative effectiveness
Systematic Desensitization	Thinking about disabled persons in relaxed, nonthreatening settings to extinguish negative attitudes
Positive Reinforcement	Use of classical or operant conditioning to modify behavior assumed to reflect attitudes
Other	Any combination of techniques other than Information Plus Direct Contact or Information Plus Vicarious Experience, which were coded separately

Table 34

Effect Sizes for Attitude Modification Techniques

Rank	Technique	Effect Sizes (Ds) N	Mean	SD	Differences Between Means[c] 2	3	4	5	6	7	8	9
1	Persuasive Message	23	.67	.56	.16	.24	.27	.28	.35	.38	.47	.54
2	Information Plus Contact	100	.51	.66		.08	.11	.12	.19	.22	.31	.38
3	Direct Contact	93	.43	.73			.03	.04	.11	.14	.23	.30
4	Vicarious Experience	58	.40	.76				.01	.08	.11	.20	.27
5	Other	71	.39	.64					.07	.10	.19	.26
6	Systematic Desensitization	21	.32	.44						.03	.12	.19
7	Information	203	.29	.51							.09	.16
8	Information Plus Vicarious	62	.20	.36								.07
9	Persuasive Message, Contrast	11	.13[a]	.33								
	Positive Reinforcement[b]	2	(1.74)	(.01)								
	Total	644	.37	.61								

[a]Ten of 11 Ds came from one study.
[b]Too few effect sizes (less than 10) to be interpretable, and so not ranked.
[c]Numbers correspond to those for ranks of techniques. For example, the difference between the Persuasive Message mean (1) and the Information Plus Contact mean (2) is .15 (.67 - .52).

magnitude of mean Ds. The mean effect sizes (Ds) for the attitude modification techniques can be viewed from two perspectives: (1) What does the average D for each treatment technique indicate about its effects as compared to no treatment? (2) What is indicated about the relative effectiveness of the different techniques?

Although we have cautioned against the use of conventions to judge the magnitude of effect sizes when the standards are arbitrary because there is no basis by which to judge the importance of variations in outcomes, it is difficult to discuss results with no criteria in mind. Lacking more firmly grounded conventions, Cohen's (1977) criteria for small (d = .2), medium (d = .5), and large (d = .8) effect sizes provide a useful frame, if applied with caution.

From that perspective, it is worth noting that none of the mean Ds reach the .8 criterion, although the mean D for the Persuasive Message studies is .67, closer to the large effect size criterion (.8) than to the medium one (.5). The differences between the Persuasive Messages mean D and the mean Ds for the other attitude modification techniques are all above the arbitrary standard for trivial differences (.12) which we set in Chapter 3. Moreover, in three cases the difference is greater than the standard for a medium difference (.31), approaching the standard for a large difference (.50) in one instance.

That messages developed purposely with an argument to sway attitudes would have the largest effect size on the average makes sense. It also may be of significance that 78% of the 23 Persuasive Message effect sizes come from studies in which the theory base (S-R/behavioral for 11, congruity/equilibrium for 6, and social judgment for 6) was explicit and the relationship to the treatment well-developed. (For "explicit theory base", the closest percentage was Systematic Desensitization with 76%, dropping then to Information Plus Vicarious Experience with 31%; for "explicit relationship to treatment", the same relationship held except that "Other" was third highest, with 34%.)

The Information Plus Contact studies produced the next largest mean D, .51, falling right at the arbitrary criterion for a medium effect size. Note again that the Information Plus Contact mean D is .16 below that for Persuasive Messages, barely larger than the arbitrary standard for trivial differences which we set in Chapter 3. At the same time, the differences between Information Plus Contact, on the one hand, and Contact and Vicarious Experience, on the other, .08 and .11, are both less than the .12 trivial difference standard; but the Information Plus Contact mean D equals or exceeds the .12 criterion for all other comparisons, equaling or exceeding the criterion for a moderate difference (.31) in two instances.

The next three mean Ds are clustered closely together--.43 for Contact, .40 for Vicarious Experiences, and .39 for Other (combinations of techniques other than the two in Table 34)--with Ds that fall at the midpoint of Cohen's criteria for small and medium effect sizes (.2 and .5). The differences between mean Ds that are lower in the ranking are non-trivial only for the Information Plus Vicarious Experience and Persuasive Message, Contrast techniques, which yielded small Ds (.20 and .13). The two remaining Ds--for Systematic Desensitization (.32) and Information (.29) are somewhat larger than the .20 small effect size standard, and only slightly higher than the means below them.

To sum up, although the mean Ds for the various techniques range from .67 to .13, clearly a broad range, there are no clear demarcations or groupings of techniques. In only one case (Persuasive Message versus Information Plus Contact) is the difference between contiguous means greater than our index of triviality (.12). The use of Persuasive Messages seems clearly to have resulted in larger Ds on the average than any other technique. Contact Plus Information runs a close second, and its use seems clearly to have produced larger Ds on the average than the use of Systematic Desensitization and the techniques ranked below it. At the other end of the scale, the effect sizes that came from efforts to investigate different persuasive messages or media for delivering them indicate general ineffectiveness, not even yielding a "small" (.20) mean D.

Treatment Variability--Homogeneity of Ds

While it might be tempting to look at the rankings in Table 34 as an index of effectiveness to be used in a singular fashion in selecting a technique to modify attitudes toward those with disabilities, that would obviously be too simplistic an interpretation of a complex set of data. To begin with, the standard deviations associated with each mean D serve as a reminder that the effects of each technique are not homogeneous; obviously, there is considerable overlap among the distributions of Ds for the various techniques. Moreover, it is important to remember that included in the Ds summarized by the means in Table 34 are negative values, indicating that, relative to the comparison group, a treatment had a negative rather than positive effect.

Table 35 presents a summary of the 150 negative effect sizes. Two things are worth noting: First, the percentage of negative effect sizes for each technique is roughly proportional to the percentage of effect sizes contributed to the total 644. No one technique contributed a markedly

Table 35

Negative Effect Sizes (Ds) for the Attitude Modification Techniques

Technique	Negative Effect Sizes (Ds) N[a]	%[b]	Mean	SD	% of Negative Technique Ds[c]
Persuasive Message	1/23	1/4	(-.36)[d]	(.00)[d]	(.04)[d]
Information Plus Contact	19/100	13/15	-.29	.29	19
Direct Contact	18/93	12/14	-.20	.17	19
Vicarious Experience	17/58	11/9	-.36	.42	29
Other	18/71	12/11	-.38	.31	25
Systematic Desensitization	4/21	3/3	(-.27)[d]	(.29)[d]	(19)[d]
Information	53/203	35/31	-.30	.32	26
Information Plus Vicarious	16/62	11/10	-.24	.19	26
Persuasive Message, Contrast	4/11	3/2	(-.14)[d]	(.10)[d]	(36)[d]
Positive Reinforcement	0/2	0/.3	—	—	—
Total	150/644	101/99.3	-.29	.30	23

[a]For N, the first figure is the number of negative effect sizes. The second figure is the total number of effect sizes.
[b]For %, the first figure is the percentage of the 150 negative effect sizes; the second figure is the percentage of the total 644 effect sizes.
[c]% of Negative Technique Ds is the percentage of the number of the Ds for a technique that were negative. E.g., 19% of the Information Plus Contact Ds were negative.
[d]Too few effect sizes (less than 10) to be interpretable.

221

disproportionate number, or percentage, of negative Ds. That is also made clear by the last column. But, second, it is remarkable that 23 percent (N = 150) of the 644 Ds were negative. That figure not only highlights the cautionary note in regard to keeping variability in mind, but raises serious questions about the adequacy of the bases for the attitude modification treatments that were investigated. It also suggests that the treatments grouped under each technique label were not necessarily perfectly alike, even though quite different from those grouped under other labels.

Variation in Treatment Features

Some differences in treatment features within techniques are worth consideration. For example, in Tables 36 and 37, it can be noted that there was considerable variability in both the types of information and the modes of presenting it in the studies of the Information approach to attitude modification. The large number of Combination ratings for both types of information and delivery mode also suggest further variability, in the way that individual components were put together.

Variability within treatment categories is also evident for the Vicarious Experience and Persuasive Message studies (see Tables 38 and 39). It is interesting, as well, that while variations in type of information and mode of delivery accounted for about 6 to 7 percent of the variance in Information Ds, the percentage of variance attributable to treatment variations is much larger for Vicarious Experience and Persuasive Message Ds--20% and 28%, respectively--suggesting that choice of technique features could be more important there.

Information gain. The studies of Information as an attitude change technique evidenced another type of variability. As noted in Chapter 3,

Table 36

Types of Information Presented in Information Treatment Technique Studies

	Effect Sizes (Ds)		
Information	N	Mean	SD
Characteristics of disabled persons	2	(.52)[a]	(.21)[a]
Problems of being disabled	1	(-.08)[a]	(.00)[a]
Similarities with nondisabled	11	.11	.60
Managing disabled children	1	(-.17)[a]	(.00)[a]
How nondisabled react	11	.39	.20
How to relate in social situations	2	(-.77)[a]	(.28)[a]
Other	15	.22	.51
Combination	160	.31	.52
Total	203	.29	.51

Note. $Eta^2 = .06$.

[a]To, few effect sizes (less than 10) to be interpretable.

Table 37

Information Delivery Modes Used in the Information Treatment Technique Studies

	Effect Sizes ($\underline{D}$s)		
Delivery Mode	N	Mean	SD
Lecture	7	(.55)	(.71)
Discussion	11	.18	.27
Lecture-discussion	3	(.02)	(.25)
Print	23	.22	.44
Panel-disabled	1	(.80)	(.00)
Speaker-disabled	4	(.22)	(.17)
Film, video	21	.40	.58
Picture, filmstrip	4	(-.02)	(.63)
Audio	7	(.74)	(.59)
Simulations	1	(-.08)	(.00)
Regular course	24	.32	.74
Regular program	23	.18	.44
Other	7	(.27)	(.37)
Combination	67	.27	.46
Total	203	.29	.51

Note. For mean $\underline{D}$s and standard deviations in parentheses, the number of effect sizes is less than 10 and too few to interpret.

$Eta^2 = .07$.

Table 38

Types of Experience in Vicarious Experience Treatment Technique Studies

Experience	N	Effect Sizes (Ds) Mean	SD
Role play	7	(.34)	(.43)
Simulation	26	.60	.72
Observe role play or simulation	2	(-.95)	(.46)
Video, films	9	(.05)	(.22)
Print, fiction or biography	2	(.05)	(.C7)
Other	1	(-.09)	(.00)
Combination	11	.59	1.06
Total	58	.40	.76

Note. For mean Ds and standard deviations in parentheses, the number of effect sizes is less than 10 and too few to interpret.

$Eta^2 = .20$.

225

Table 39

Types of Persuasive Messages Presentations in Persuasive Message Treatment Technique Studies

	Effect Sizes (Ds)		
Presentation	N[a]	Mean	SD
Video, film	3	(.52)	(.12)
Audio	3	(.31)	(.10)
Expert	8	(.48)	(.40)
Expert, disabled	1	(1.32)	(.00)
Other	8	(.99)	(.74)
Total	23	.67	.56

Note. $Eta^2 = .28$.

[a]All mean Ds and standard deviations are in parentheses because the number of effect sizes is less than 10 and too few to interpret.

there are two pertinent questions to be asked of studies in which purveying information is the attitude modification technique: Did the Ss learn the intended content? And, to what extent were gains in knowledge of that content related to attitude outcomes? The answer to the first question can be construed as speaking to the extent to which the treatment independent variable was implemented, the second as getting at the relationship between different values of the independent variable and scores on the dependent attitude measure.

Surprisingly few of the 203 Information effect sizes come from studies in which Ss' gains in knowledge of content were even assessed and reported— 43 effect sizes out of 203 (21%), as is indicated in Table 40. Of those, sufficient data were available for 37 effect sizes (18%) to compute Information Gain Ds and to obtain the correlation between Information Gain and Attitude Ds.

As might be expected, whether or not the authors concluded there was No Gain or a Clear Gain in information was highly associated with Information Gain Ds (r_{pb} = .38). Authors' conclusions in regard to information gain were not, however, highly associated with attitude Ds, in large part because so many (N = 160) came from studies in which attention was not paid to determining whether Ss actually learned the information which was to have an effect on their attitudes. For those instances in which information gain was assessed so that Ds could be computed, the relation with Ds for attitude dependent measures was moderately high (r = .53; r^2 = .28)*. Given the

*Inspection of the scatter-diagram indicated a rectilinear regression line was a good fit.

Table 40

Information Gain and Attitudes in Information Modification Technique Studies

Type of Report on Information Gain	N
None	160
Narrative	6
Descriptive Statistics	3
Statistical Significance	5
Descriptive Statistics and Statistical Significance	29
Total	203

Conclusion re Information Gain	Effect Sizes (Ds): Information N	Information Mean	Information SD	Attitude N	Attitude Mean	Attitude SD
Gain Not Assessed	—	—	—	160	.27	.50
Clear Gain	25	1.40	1.05	25	.29	.64
No Gain	12	-.07	.44	12	.23	.44
Can't Tell, Inconclusive[a]	6	—	—	(6)[b]	.71	.16
Total	43	.91	1.13	203	.29	.51

Note. r^2_{pb} for Information Gain Ds = .38; Eta^2 for Attitude Ds = .02. The overall Pearson product-moment correlation coefficient for the 37 effect sizes for which both Information Gain and Attitude Ds were available = .53; r^2 = 28.

[a]Narrative reports of information gain, so information Ds could not be computed.
[b]Too few effect sizes (less than 10) to be interpretable.

stability often attributed to attitudes and the attendant difficulty in modifying them, and the likely other ways in which the studies from which the Ds came varied, that correlation is fairly sizeable.

It is unfortunate that more researchers did not attend to verifying that increased knowledge was the result of their efforts, or argue directly that exposure to the information, not learning it, was deemed to be the important variable. In any event, not only the assessment of information gain but the extent of gain were other types of variability among studies in the Information attitude modification technique category.

Contact variability. There was also variability in the contact situations used in Direct Contact studies (see Table 41) and in the disabilities with which Ss were in contact. About 12 percent of ($Eta^2 = .12$) the variance in Contact Ds was associated with situation differences, and about 10% with differences in disabilities. However, the Ns upon which most of the variance Ds in Tables 41 and 42 are based are so small as to make interpretation untenable. The lack of interpretability is compounded by the fact that of the two out of three Ds with sufficient Ns in Table 41, one is a Combination category and the other is an amorphous "Other" category. Consequently, while Table 41 suggests diversity of Contact studies, it tells us little about the effects of different types of content. By the same token, the category in Table 42 with the largest N is "Combination", and only two other categories have more than 10 effect sizes in them.

Of particular interest, as noted in Chapters 2 and 3, are differences in contact that might, according to theory, be related to attitude outcomes. Along those lines, it is relevant that only 12 (13%) of the Direct Contact effect sizes were coded as coming from studies in which there was an explicit

Table 41

Contact Situations for the
Direct Contact Treatment Technique Studies

Contact	Effect Sizes (Ds) N	Mean	SD
As companion	8	(.42)	(.42)
As peer tutor	2	(.74)	(.29)
In cooperative learning group	3	(.22)	(.31)
As classmates	8	(1.13)	1.81
Practice teaching	4	(.23)	(.52)
In recreation program	4	(.52)	(.24)
Guest speaker	18	.24	.29
As teacher or counselor	8	(.13)	(.58)
Other	28	.50	.72
Combination	10	.33	.19
Total	93	.43	.73

Note. For mean *D*s and standard deviations in parentheses, the number of effect sizes is less than 10 and too few to interpret.

$Eta^2 = .12$.

Table 42

Characteristics of Disabled Persons
in Contact Studies--Disabilities

Disability	Effect Sizes (Ds) N	Mean	SD
Combination	20	.35	.52
Mentally Ill	15	.56	.53
MR--Mild/Moderate	14	.63	1.50
MR--General	6	(.51)	(.43)
MR--Can't Tell	5	(.23)	(.37)
Severe Multiple	5	(.36)	(.28)
Emotionally Disturbed	4	(.24)	(.27)
MR--Severe/Profound	3	(.50)	(.47)
Deaf	3	(-.16)	(.35)
Multiple Disabilities	2	(.67)	(1.61)
Physical--General	2	(.51)	(.03)
Wheelchair	2	(.42)	(.57)
Paraplegic	2	(.17)	(.18)
Blind	2	(.38)	(.35)
Hearing Impaired	1	(.75)	(.00)
Learning Disabled	1	(1.88)	(.00)
Can't Tell	6	(.18)	(.23)
Total	93	.43	.73

Note. For mean Ds and standard deviations in parentheses, the number of effect sizes is less than 10 and too few to interpret.

$Eta^2 = .10$.

231

theoretical basis for the treatment, and only 9 (10%) of the Ds came from comparisons in which the theoretical base was rated as well-developed.

Data from the categories on the Contact Coding Sheet (see Appendix B), constructed (as noted in Chapter 3) to reflect theoretically important attitude change elements of contact, give some indication of diversity in Contact treatment and associated outcomes. For example, one factor considered in contact theory is the relative status of the interacting disabled and nondisabled persons, with equal status deemed important to the development of positive values. Table 43 contains data on three indicators of status as well as for an overall status rating.

Clearly, there was considerable diversity among the studies in our population in regard to the relative status of nondisabled Ss and the disabled persons with whom they had contact. Moreover, some of the mean Ds are perplexing. As expected, the mean D for "same age" contact (.61) is higher than that for contact in which the persons with disabilities are younger (.35). However, when the persons with disabilities were older than the nondisabled Ss, the mean is lower (.23) than the "same age" mean. A similar overall pattern holds for vocational-educational prestige, suggesting that, as would be expected, that element of prestige and age are not independent. In terms of "helping relationships" there is diversity, too. Unfortunately, there were only 9 Ds for the professional help category; the mean for that category does, however, reflect the observation in the literature that professionals who have worked with disabled persons often develop (or are assessed as having) more negative attitudes. The most perplexing finding is for overall status. The comparisons in which the persons with disabilities were rated as lower in status produced the highest

Table 43

Status Factors and Overall Status in Contact Studies

Age

Relative Age	Effect Sizes (Ds) N	Mean	SD
Can't Tell	21	.47	.50
Disabled Younger	23	.35	.70
Same	24	.61	1.15
Disabled Older	17	.28	.29
Variety	8	(.36)[a]	(.31)[a]
Total	93	.43	.73

Note. Eta^2 = .03.

[a]Too few effect sizes (less than 10) to be interpretable.

Vocational-Educational Prestige

Relative Vocational-Educational Prestige	Effect Sizes (Ds) N	Mean	SD
Can't Tell	21	.52	.59
Disabled Lower	35	.34	.59
Same	20	.64	1.21
Disabled Higher	17	.28	.29
Total	93	.43	.73

Note. Eta^2 = .03.

Helping Relationship

Helping Relationship	Effect Sizes (Ds) N	Mean	SD
None	47	.38	.84
Professional	9	(.08)[a]	(.42)[a]
Preprofessional	18	.73	.79
Nonprofessional	19	.46	.38
Total	93	.43	.73

Note. Eta^2 = .06.

[a]Too few effect sizes (less than 10) to be interpretable.

Overall Status

Relative Status	Effect Sizes (Ds) N	Mean	SD
Can't Tell	2	(.03)[a]	(.16)[a]
Disabled Lower	59	.54	.88
Equal	16	.24	.30
Disabled Higher	16	.26	.30
Total	93	.43	.73

Note. Eta^2 = .04.

[a]Too few effect sizes (less than 10) to be interpretable.

233

mean D, .54, versus .24 and .26 for equal status and higher status for the disabled, respectively.

Variability in Contact treatments is also evident in Table 44. Only 17% of the Ss voluntarily initiated contact, while 53% (N = 49) were assigned to a contact situation. The volunteers, as would be predicted based on theory, did have a higher mean D (.48) than those assigned to contact (mean = .27). Of course, most of the "role choice" Ss, for whom the mean D was highest (.60), were volunteers in that they chose coursework or employment knowing that it would involve contact with persons who had disabilities.

There was less variability in intimacy of contact, with 70% (N = 65) of the effect sizes coming from studies in which the contact was rated as casual. Not much is revealed, therefore, with theoretical implications.

There was little variability in the nature of the cooperation involved in Contact studies, with 84% of the effect sizes (N = 78) concentrated in two categories--in which cooperation was rated as either "not necessary" (N = 21) or "implicit" (N = 57). Consistent with theory, the implicit cooperation mean D (.49) is higher than that for the "not necessary" D (.36), although the difference is just greater than the triviality criterion.

The results for three categories developed to describe types of contact are not presented because of the lack of information in reports for coding them. Little can be said in regard to the pleasantness of contact, another theoretically important factor, because there was sufficient information for coding that category for only 10% of the effect sizes. By the same token, it could not be determined from the reports for 99% of the Contact effect sizes whether the Ss received explicit reinforcement or shared internal or external reinforcement with those who had disabilities. For 95% of the effect sizes,

Table 44

Type of Contact in Contact Studies

Basis for Contact Initiation

	Effect Sizes (Ds)		
Basis	N	Mean	SD
Assigned	49	.27	.59
Role choice	23	.60	.75
Voluntary	16	.48	.42
Varied	5	(1.06)[a]	(1.88)[a]
Total	93	.43	.73

Note. $Eta^2 = .08$.

[a]Too few effect sizes (less than 10) to be interpreted.

Intimacy of Contact

	Effect Sizes (Ds)		
Intimacy	N	Mean	SD
No Interaction	3	(1.24)[a]	(2.10)[a]
Casual	65	.38	.55
Close	4	(.39)[a]	(.49)[a]
Varied	13	.31	.43
Potential Contact	8	(.75)[a]	(1.49)[a]
Total	93	.43	.73

Note. $Eta^2 = .06$.

[a]Too few effect sizes (less than 10. to be interpre able.

Extent of Cooperation and Competition

	Effect Sizes (Ds)		
Cooperation--Competition	N	Mean	SD
Can't tell	7	(.17)[a]	(.32)[a]
No opportunity	3	(1.24)[a]	(2.10)[a]
Not Necessary	21	.36	.55
Implicit cooperation	57	.49	.73
Explicit cooperation	2	(.07)[a]	(.25)[a]
Implicit competition	1	(-.47)[a]	(.00)[a]
Combination	2	(.17)[a]	(.18)[a]
Total	93	.43	.73

Note. $Eta^2 = .08$.

[a]Too few effect sizes (less than 10) to be interpretable.

the same lack of information was encountered in regard to the modeling of positive interactions.

Another theoretical important contact variable is the degree of support for positive attitudes from institutional norms, persons in authority, and peers. For 76% of the effect sizes (N = 71), there was not enough information in the report to code "yes" or "no" (see Table 45). Nevertheless, the higher mean D (.61 versus .38) when support could be identified, with a difference of .24, is consistent with contact theory.

The characteristics of the disabled persons with whom Ss have contact are also important, according to the theory. In Table 42, we presented the frequencies of types of disabilities. Table 46 contains the information that could be garnered on two other attributes--the extent to which the persons with disabilities acted in ways to reinforce negative stereotypes and the extent to which they were likely viewed as competent by the Ss or to which they openly acknowledged and accepted any lack of competence due to their disability. There was variety among studies, in that for 18 effect sizes (19%) it could be discerned that negative stereotypes likely were present, but for 23 (25%) they were not. However, there was little difference in the mean Ds (.33 versus .24). And, for 52 (56%) of the Ds, Can't Tell was coded.

Again, for competence there was some variety among the studies from which the Contact effect sizes came. The result with the greatest potential interest is for "Competent", with the high mean D of .75. However, with only 9 effect sizes, that result and the .39 mean difference with "Lacked" competence, must be treated with caution. The low mean D (.24) for the "Acknowledged/accepted" category is also of interest because, theoretically, acknowledging and accepting one's disability should have a positive effect on the attitudes of nondisabled persons with whom one has contact.

Table 45

Extent of Institutional Support for Attitude Change in Contact Studies

	Effect Sizes ($\underline{D}$s)		
Support	N	Mean	SD
Can't Tell	71	.38	.73
Yes	22	.61	.73
Total	93	.43	.73

Note. $Eta^2 = .02$.

Table 46

Characteristics of Persons with Disabilities
in Contact Studies--Negative Stereotypes and Competence

Negative Stereotype Reinforced

	Effect Sizes (Ds)		
Stereotype	N	Mean	SD
Can't Tell	52	.55	.93
Yes	18	.33	.36
No	23	.24	.29
Total	93	.43	.73

Note. $Eta^2 = .04$.

Competence of Person with Disability

	Effect Sizes (Ds)		
Competence	N	Mean	SD
Can't Tell	16	.68	1.19
Lacked	50	.36	.54
Acknowledged/accepted	18	.25	.27
Competent	9	(.75)[a]	(1.14)[a]
Total	93	.43	.73

Note. $Eta^2 = .05$.

[a]Too few effect sizes (less than 10) to be interpretable.

Finally, the characteristics of the nondisabled Ss, in particular their personality traits and their prior attitudes, are theorized to be of importance to attitude change. We could find little attention to either in the literature. For 83 Contact effect sizes (89%), personality traits were not assessed; for none were the relationships between personality variables and attitude change from Contact analyzed. Pre-treatment attitudes were assessed, as one would expect given popular research designs for applied settings, for 90 (86%) of the effect sizes; however, the relationship of antecedent attitudes to the effects of contact on attitudes was analyzed for only one effect size.

Contact Summary

Contact treatments varied widely, as did Information and Vicarious Experience as attitude modification techniques. We have discussed the differences in Contact in the context of theory as the effects of contact are of particular interest in the field. Although approached from a somewhat different perspective, our data largely support Makas's (1986) conclusion that the inadequate design of studies has precluded the productive testing of the hypotheses of contact theorists such as Allport (1954) and Amir (1976). The inadequate reporting of studies is, as well, a barrier to post hoc efforts to check results against theory. Further analyses of our data will be conducted to attempt to discern the effects of combinations of factors (such as contact which is both voluntary and with nonstereotypic disabled persons), but the potential fruitfulness is limited by the lack of attention to theoretically important variables, as evidenced by the failure to report the information necessary to code them.

Attitudes toward . . . ? An important treatment feature is the disability toward which the attitude modification efforts were directed. As Table 47 indicates, 44 percent (N = 286) of the effect sizes came from studies in which a target disability was not specified, but efforts were directed at changing attitudes toward an amorphous category of "disabled persons in general". The next most frequent change target, attitudes toward general physical disabilities (or, put differently, unspecified physical disabilities), was a distant second with 15 percent (N = 97) of the effect sizes. From there, the number of effect sizes for disability targets drops off rapidly to 65 (10%) for Mentally Ill, to 37 and 36 (6% each) for Mentally Retarded, General (i.e., level of retardation not specified) and Combination (i.e., more than one disability target specified). Each of the other disability targets accounts for 4% or less of the 644 effect sizes.

Two types of treatment variability are evident in Table 47. First, the effects of each attitude modification approach have been investigated with several disability targets. Secondly, however, there is some clustering of disability targets within treatments. For example, Contact effect sizes have only come in substantial numbers (N of 10 or more) from studies directed at changing attitudes toward disabled persons in general, the mentally ill, and the mentally retarded in general*. Conversely, substantial numbers of effect sizes for the mentally ill as an attitude change target came from studies that investigated either Direct Contact or Information Plus Contact.

Moreover, not only the numbers but the effects are not consistent within disabilities or treatments. That is, not only are there differences in total

*It does not help interpretation that none of the target disability effect sizes for Persuasive Message as a technique, which had the highest overall mean D (.67), is based on an N of 10 or more.

Table 47

Disabilities Toward Which Modification Techniques Were Directed

Technique	Disabled General	Physical General	Mentally Ill	Mentally Retarded General	Combination	Hearing Impaired	Moderately Retarded	Severely Retarded	Visually Impaired	Other	Physically Impaired, Other	Emotionally Disturbed	Learning Disabled	Total
	Disability													
Persuasive Message	(.70) (.05) 3	(.49) (.34) 9	(.48) (.40) 8	(1.71) (.75) 3	— — —	— — —	— — —	— — —	— — —	— — —	— — —	— — —	— — —	.67 .56 23
Information Plus Contact	.53 .47 31*	(.66) (.72) 9	.20 .52 22**	(.65) (.93) 8	(.52) (1.14) 8	— — —	(.91) (.96) 6	.50 .32 10	(.71) (.23) 2	(.56) (.98) 2	— — —	(1.20) (.25) 2	— — —	.51 .66 100
Direct Contact	.41 .59 33	(.26) (.26) 4	.56 .53 15	.20 .30 11	(.74) (.29) 2	(.07) (.53) 4	(.91) (1.84) 9	(.50) (.47) 3	— — —	(.29) (.31) 7	(.83) (.00) 1	(.24) (.27) 4	— — —	.43 .73 93
Vicarious Experience	.27 .84 29	(.61) (.50) 7	(.41) (.30) 4	(.30) — 1	(.79) (.51) 4	(1.47) (1.48) 3	— — —	— — —	(.52) (.27) 3	— — —	(-.01) (.17) 7	— — —	— — —	.40 .76 58
Other	.41 .47 30	.64 .40 11	(.34) (.41) 4	(1.04) (.88) 6	(.37) (.66) 3	-.30 .49 13**	(.67) (.00) 1	— — —	— — —	(1.67) (.73) 2	(.04) (.00) 1	— — —	— — —	.40 .64 71
Systematic Desensitization	(.13) (.00) 1	(.30) (.55) 5	(-.10) (.49) 4	— — —	(.71) (.20) 6	— — —	— — —	— — —	(.25) (.13) 5	— — —	— — —	— — —	— — —	.32 .44 21
Information	.23 .51 104**	.36 .51 43**	(.17) (.41) 8*	(.19) (.28) 6	.44 .71 13	(.92) (.59) 4	(.22) (.17) 4	(.12) (.41) 7	(.25) (.59) 8	(.42) (.18) 2	— — —	(.95) (.36) 2	(.18) (.30) 2	.29 .51 203
Information Plus Vicarious Experience	.15 .35 44**	(.18) (.18) 7	— — —	(.17) (.20) 2	— — —	— — —	— — —	— — —	(.55) (.21) 2	— — —	(.10) (.33) 2	(.87) (.00) 1	(.59) (.47) 4	.20 .36 62
Persuasive Message, Contrast	.13 .33 11	— — —	— — —	— — —	— — —	— — —	— — —	— — —	— — —	— — —	— — —	— — —	— — —	.13 .33 11
Positive Reinforcement	— — —	(1.74) (.01) 2	— — —	— — —	— — —	— — —	— — —	— — —	— — —	— — —	— — —	— — —	— — —	1.74 .01 2
Total	.29 .53 286	.46 .52 97	.31 .50 65	.56 .75 37	.55 .71 36	.19 .92 24	.76 1.32 20	.36 .40 20	.37 .42 20	.56 .62 13	.09 .30 11	(.68) (.49) 9	(.46) (.44) 6	.37 .61 644

Note. The first number in each cell is the mean D, the second is the standard deviation, and the third the number of cases.

Means and standard deviations in parentheses are based on fewer than 10 cases.

ast 10 fewer cases than expected, based on marginal frequencies.
ast 10 more cases than expected, based on marginal frequencies.

mean Ds between disability targets (the Eta^2 for disability target and Ds is .05), but between mean Ds within disability categories as well. For example, there is a difference of .47 between the mean D for Disabled General (.29) and Moderately Retarded (.76); yet, within Disabled General, the range of mean Ds is from .15 (for Information Plus Vicarious Experience, ignoring the .13 for Persuasive Message, Contrast because 10 of the Ds came from the same study) to .53 (for Information Plus Contact), a difference of .38. By the same token, there is considerable variance in mean Ds within treatment categories. For Information Plus Contact the mean Ds range from .20 (Mentally Ill) to .66 (Physical General), a difference of .46. Although the effectiveness of treatments might appear to be largely a function of interactions with disability attitude targets, the disparities in Ns for cells, as well as the large number of empty cells (not to mention the potential underlying interactions with other factors such as age of Ss), preclude such a conclusion--or even the use of analysis of variance to determine the proportion of the variance in Ds attributable to the treatment by disability interaction.

Implications. Variations in the features of treatment variables with similar labels, such as described in the previous sections, are at the heart of the "apple and oranges" criticism of meta-analyses--that is, the objection that lumping studies for quantitative analysis obscures important differences within groups of treatments (see, e.g., Bangert-Drowns, 1986, pp. 389-90, 392). That objection to meta-analysis raises a dilemma. One horn is the effects of grouping studies. The other horn has to do with adequacy of numbers of cases to sort out the apples and oranges. If the sets of effect sizes for treatment groups are broken down to make fine, or even fairly gross, within-treatment distinctions, the number in many cells is likely to be so small as to yield findings in which little confidence can be placed. The problem is exacerbated even further if one takes into account not only treatment differences (those coded as well as the large number of uncoded possibilities), but other study characteristics (to be discussed shortly for our data set).

Rather than using the apple and oranges metaphor to refer to studies, or effect sizes, within a treatment type, the metaphor might be more aptly applied to refer to different types of treatment. That is, if keeping the doctor away is the desired outcome, it is reasonable to ask whether different fruits have different effects. However, it is especially important when there is considerable variability in outcomes within fruit types to keep in mind that all of the apples (or oranges, pears, bananas, etc.) were not the same nor were the people who ate them or the conditions under which they were eaten. By the same token, asking if there are differences in outcomes among categories of attitude modification techniques is legitimate as long as one does not lose sight of the variability in study characteristics and outcomes within technique categories, and of the extent to which size of outcomes is not neatly clustered within those categories.

Other Study Characteristics

As reported earlier in this chapter, the global indicators of methodological quality are not related to outcomes in our data. Variations in treatment techniques tend to be. Are there other study characteristics that are? What other reservations might be necessary in drawing conclusions about the attitude modification data presented in Table 34? One such issue is whether effect sizes from treatment versus control (T vs. C), treatment versus placebo (T vs. P), and single-group, pre-posttest (Pre-post) comparisons should have been pooled for analysis.

Type of Comparison

As can be seen in the first three columns of Table 48, the overall means for treatment versus control (T vs. C) and treatment versus placebo (T vs. P) comparisons (.36 and .29, respectively) were close to one another, but the difference between each and the single-group, pre-posttest (Pre-post) mean D (.49) was .13 and .20, respectively. Yet, the Eta^2 for the relationship between comparison type and magnitude of D is only .01. The small Eta^2 reflects in part the small numbers of T vs. P (N = 49; 7%) and Pre-post (N = 97; 15%) comparisons. With 77 percent (N = 498) of the Ds in the T vs. C category, there was little variability in comparison type.

Another way to approach the issue is to ask whether excluding the Pre-post Ds from the mean Ds for the treatment techniques would have affected the picture portrayed. As can be seen in the fourth and fifth columns in Table 48, the rankings and relative magnitudes of the mean Ds remain essentially the same when the Pre-post means are excluded and the T vs. C and T vs. P means are pooled. The only changes in ranking are for means which are nearly identical in both the T vs. C plus T vs. P column and the Total column, with mean differences so small (.01 to .03) that differences in ranks are basically meaningless.

It is worth noting, in comparing the first three columns of Table 48 (T vs. C, T vs. P, Pre-post), that Pre-post comparisons yielded higher mean Ds for Information Plus Contact and Direct Contact, types of techniques likely to be used in college courses where pre-posttest data are often gathered. As a matter of fact, 63% (N = 61) of the Pre-post Ds came from course

Table 48

Treatment Technique Effect Sizes (Ds) by Type of Comparison

Technique	Effect Sizes (Ds)									
	T vs. C		T vs. P		Pre-post		T vs. C plus T vs. P		Total	
	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
Persuasive Message	18	.53	3	(1.71)	2	(.41)	21	.70	23	.67
Information Plus Contact	79	.46	1	(.30)	20	.75	80	.46	100	.52
Direct Contact	72	.35	3	(.35)	18	.76	75	.35	93	.43
Vicarious Experience	52	.32	2	(.47)	4	(1.41)	54	.33	58	.40
Other	53	.45	10	-.27[a]	8	(.84)	63	.34	71	.39
Systematic Desensitization	21	.32	—	—	—	—	21	.32	21	.32
Information	152	.33	24	.30	27	.03	176	.33	203	.29
Information Plus Vicarious	37	.14	6	(.36)	19	.28	43	.17	62	.20
Persuasive Message, Contrast	11	.13[b]	—	—	—	—	11	.13	11	.13
Positive Reinforcement	2	(1.74)	—	—	—	—	2	1.74	2	(1.74)
Total	497	.36	49	.29	98	.49	546	.35	644	.37

Note. Mean Ds based on fewer than 10 effect sizes are considered too unstable to interpret. They are in parentheses.

Eta^2 for Comparison Type (T vs. C, T vs. P, and Pre-post) and magnitude of D is .01.

[a] Nine of 10 Ds from the same study.
[b] Ten of 11 Ds from the same study.

246

evaluations (N = 39; 40%) and program evaluations (N = 22; 23%), and nearly 60 percent of the course and program evaluation effect sizes came from samples of college and university students. So, even though pooling mean Ds from the three types of comparisons did not have a significant impact on the relative size or rankings of treatment technique means, it must be kept in mind that single-group, pre-posttest comparisons contributed heavily to the mean Ds for certain attitude modification techniques used with certain samples.

Time of Posttest

In Chapter 4, the small percentage (13%) of effect sizes that came from follow-up posttesting in the total data set was noted. The question of concern at this point is, was the magnitude of Ds associated with time of posttest? The answer to that question bears on the long term effects of the attitude modification treatments. The answer also speaks to whether it was appropriate to pool Ds from different times of posttesting for analysis, as we have done.

In coding time of posttesting, two types of information were entered: (1) the number of weeks between the end of treatment and administration of the posttest; and, (2) a code to identify whether the posttesting was immediate, delayed, or follow-up. An "immediate" posttest was an initial assessment administered very soon after the end of the experimental treatment--i.e., within a day or at the next available meeting of the group. A "delayed" posttest was also an initial posttest, but was not administered immediately, sometimes to obscure the relationship between the treatment and the test. A "follow-up" posttest followed an initial posttest, or it may have followed a prior follow-up posttest.

As would be anticipated, the statistics for number of weeks from treatment end to posttesting varied for the three time-of-posttesting categories. For immediate posttests, the mean number of weeks to testing was .03, with a standard deviation of .24 (and a mode and median of 0.00, and a range of 0.00 to 4.3 weeks); for delayed posttests, the mean number of weeks to testing was 1.7, with a standard deviation of 1.9 (and a mode and median of 1, and a range of .20 to 8.00 weeks); for follow-up posttests, the mean was 10.4, with a standard deviation of 18.4 (and a mode and median of 6, with a range of 1.3 to over 100 weeks).

As can be noted from the mean Ds and the Eta^2 (.003) in Table 49, there is little relationship between time of posttesting and average D. In addition, when a correlation was run between the number of weeks after the end of treatment when the posttest (immediate, delayed, or follow-up) was administered and the Ds for the 586 effect sizes for which that information was available, $r = .05$ ($r^2 = .002$)—a very small relationship.

Was there a systematic time-of-posttest effect among treatment techniques? Table 50 has information for the three treatments for which the N for follow-up posttesting was sufficient (i.e., at least 10) for interpretation. No pattern is evident, although there is one nontrivial (by our .12 criterion) difference between a posttest and follow-up mean. For Other, the follow-up posttest mean D (.21) was .25 less than the posttest mean D (.46).

To sum up, little evidence about "long term" effects is available from our data. What information there is indicates little relationship between time of posttesting and magnitude of D. And, there appeared to be little reason not to pool the total 644 effect sizes for analysis.

Table 49

Effect Sizes (Ds) for Different Posttesting Times

	Effect Sizes (Ds)		
Time of Posttest	N	Mean	SD
Can't Tell	50	.28	.54
Immediate	427	.38	.63
Delayed	86	.42	.60
(Immediate Plus Delayed)[a]	(513)	(.39)	(.62)
Follow-up	81	.38	.57
Total	644	.37	.61

Note. $Eta^2 = .003$.

[a]The information from Immediate and Delayed (within one week of end of treatment) posttesting is presented pooled, as well as separately, because posttests that close together would typically not be discriminated in research reports, and the mean Ds were nearly identical.

Table 50

Follow-up Posttest Effect Sizes (Ds)

	Posttest[b]		Follow-up	
Technique[a]	N	Mean	N	Mean
Information	167	.27	23	.37
Information Plus Contact	74	.54	13	.42
Other	48	.46	14	.21

[a]Techniques included only if the number of effect sizes for either posttesting or follow-up posttesting was at least 10.
[b]Immediate and delayed posttest data combined.

Type of Dependent Measure

Another aspect of testing that introduces study variability is the instruments used to assess attitudes. As Bangert-Drowns (1986) has noted, the "apples and oranges" criticism of meta-analysis has two components. The first, discussed above, is the concern about grouping together under similar labels studies with different features. The second involves apprehension about the effects of lumping together for analysis effect sizes that come from different dependent measures.

One way of considering similarities or differences in dependent measures is to examine the assessment approaches used. Information in that regard is presented in Table 51. It is clear that the assessment of attitudes was dominated by questionnaires (N = 425; 66%), most of which were made up of Likert-type items. The next highest type of instrument, the semantic differential, is a distant second--N = 73; 11%. Only three other assessment types yielded data for at least 10 effect sizes: social distance scales (N = 58; 9%); adjective checklists (N = 32; 5%); and, a composite category of tests that didn't fit in any of the major categories, "Other", with 35 effect sizes (5%). The Eta^2 for Ds and type of assessment is .05. The association is not large, with assessment clearly dominated by Likert-type scales.

A perplexing piece of information in Table 51 is the relatively low mean D for social distance dependent measures (D = .16). With that low mean D, it is relevant to inquire whether the few social distance scale assessments were associated predominantly with any attitude modification technique.

One source of evidence relevant to that question is the overall association between treatment technique and type of assessment. The Cramer's V is only .18. And, that value reflects primarily a greater frequency of

Table 51

Types of Dependent Measures

	Effect Sizes (Ds)		
Assessment Type	N	Mean	SD
Questionnaires (Likert-type)	425	.38	.58
Semantic Differential	73	.35	.61
Social Distance	57	.16	.41
Adjective Checklist	32	.48	.57
Sociometric	9	(1.11)	(1.71)
Telephone-Mail Request	4	(.86)	(.44)
Projective Test, Pictures	2	(.03)	(.27)
Sentence Completion	2	(.08)	(.45)
Interview-Nonstructured	2	(1.66)	(.23)
Systematic Observation	1	(.44)	(.00)
Rankings	1	(.23)	(.00)
Other	36	.33	.62
Total	644	.37	.61

Note. Means and standard deviations based on fewer than 10 Ds are in parentheses.

$Eta^2 = .05$.

questionnaire testing with the Information and Information Plus Contact techniques than expected based on marginal totals (148 versus 134 expected and 80 versus 67 expected, respectively) and fewer or that type of assessment than expected with Direct Contact studies (51 versus 61 expected) and Information Plus Vicarious Experience studies (19 versus 41 expected). There was not a great disparity in the expected and actual uses of social distance scales, based on marginal frequencies nor, as follows, a great concentration of use by treatment technique--as indicated in Table 52. Moreover, the differences between the mean Ds based on sufficient numbers of effect sizes to be interpretable are slight. The low mean D for Information Plus Vicarious Experience (.07) is a trivial difference from the mean D for Information (.14) and for Direct Contact (.11).

There do not appear to be systematic differences in dependent measures among studies of different attitude modification techniques that would have major effects on outcomes. This is, of course, due in part to the lack of variability in types of assessments--i.e., the prevalent use of questionnaires to assess attitudes. Perplexing questions of construct validity are raised by that use, and the perplexity is piqued by the low mean D for social distance scale assessments.

The meaning of mean Ds based largely on reactive paper-and-pencil assessments, with so few coming from indirect, behavioral methods of assessment (Antonak, 1986), is a greater concern with this data set than is the "apples and oranges" concern about intermingling effect sizes from differing dependent measures. Rokeach (1970) commented similarly on the state of assessment in attitude research in general: Typically, one posttest is administered shortly after the treatment, with little attention to

Table 52

Social Distance Scale Assessments

Technique	N Actual	N Expected[a]	Mean	SD
Information	10	18	.14	.49
Direct Contact	11	8	.11	.30
Vicarious Experience	7	5	(.22)[b]	(.24)[b]
Persuasive Message	1	2	(.94)[b]	(.00)[b]
Persuasive Message, Contrast	0	1	—	—
Information Plus Contact	5	9	(.31)[b]	(.77)[b]
Information Plus Vicarious	16	5	.07	.15
Systematic Desensitization	1	2	(.11)[b]	(.00)[b]
Positive Reinforcement	0	0	—	—
Other	6	6	(.26)[b]	(.68)[b]
Total	57	56	.16	.41

[a]Based on marginal frequencies rounded to whole numbers, so total does not equal 57.
[b]Based on too few Ds to be interpretable.

assessing behavioral change. This is contrary to Rokeach's (1973) own research in which, for example, he analyzed responses to an NAACP membership solicitation from three to five months after an experimental treatment, and then a year later, to determine behavioral attitude effects (also see Ball-Rokeach, Rokeach, & Grabe, 1984).

Length of Treatment

Among the prior reviewers of the research on the modification of attitudes toward persons with disabilities, Anthony (1972) commented that length of treatment might be an important study characteristic. Information on the total number of hours of treatment was available for 545 (84%) of the effect sizes in our data set.

Length of treatment varied considerably, from .10 hour to over 1,000 hours. The mean number of hours of treatment was 37.14, with a standard deviation of 127.95. But particularly revealing are the median number of treatment hours, 4.00, and the mode of .7 hours--about the length of a typical class period.

For the 545 effect sizes for which the number of hours of experimental treatment was available, there was essentially no relationship between length of treatment and outcomes (r = .02). But as can be seen in Table 53, that overall correlation obscures an apparent interaction between type of technique and length of treatment. For Information and Persuasive Message, there were moderate negative associations (information becomes boring?). The coefficient of .60 for Systematic Desensitization is particularly intriguing. It makes sense that the effects of desensitization would increase with length of treatment--to a certain point. There were no length of treatment outliers for that technique, and the number of hours of

Table 53

Correlations Between Length of Experimental Treatment (Total Hours) and Magnitude of D

Technique	All Available Data			Without Outliers		
	N	r	r^2	N	r	r^2
Information	179	-.21	.05	171	-.04	—[a]
Direct Contact	64	-.04	—[a]	57	.09	.01
Vicarious Experience	58	.06	—[a]	No extreme outliers		
Persuasive Message	23	-.28	.08	19	-.08	.01
Persuasive Message, Contrast	11[b]	.40	.16	No outliers		
Information Plus Contact	71	-.03	—[a]	70	-.20	.04
Information Plus Vicarious	52	-.04	—[a]	51	.05	—[a]
Other	64	.11	.01	63	.29	.09
Systematic Desensitization	21	.60	.36	No outliers		
Total	545	.02	—[a]	431	.01	—[a]

[a] r^2 less than .005.
[b] 10 Ds came from the same study.

treatment were clustered from 5 hours and less. For the other techniques, no relationship was evident.

Visual inspection of scatter-diagrams did not reveal any curvilinear relationships, but it did indicate that a few outliers might be exerting an inordinate effect on some coefficients. Consequently, correlations were rerun with outliers excluded. These coefficients are also in Table 53. The basic effect was to dampen the relationships. For Information, the coefficient dropped from -.21 to -.04; for Persuasive Message, the coefficient dropped from -.28 to -.08. However, in two cases, the association increased: from -.03 to -.20 for Information Plus Contact, and from .11 to .29 for Other. Although the relationship of length of treatment to outcomes is generally small, the effect did not appear to be uniform across attitude modification techniques, and it is not a variable to be totally ignored.

Context

The coding instrument contained a category for Treatment Context, the general milieu or environment within which a study was conducted. As can be seen in Table 54, the effect sizes came largely from two contexts, College-University and Elementary-Secondary Schooling. Those two contexts account for 85 percent (N = 549) of the effect sizes.

The highest percentage of effect sizes (49%; N = 314) came from studies carried out in a college or university environment; the second most frequent environment (36%; N = 235) was elementary and secondary schooling. The large number of studies in both categories, but particularly Elementary-Secondary Schooling, may well be primarily a function of PL 94-142, passed in 1975. There was a dramatic 810 percent increase in effect sizes from Elementary-

Table 54

Treatment Contexts

Context	Effect Sizes (Ds) N	Mean	SD
Elementary-Secondary Schooling	235	.38	.60
College-University	314	.40	.65
Inservice	54	.24	.42
Adult Education	3	(.05)	(.40)
Work	9	(.12)	(.67)
Community	8	(.50)	(.38)
Recreation	7	(.33)	(.30)
Other	14	.49	.65
Total	644	.37	.61

Note. For mean *D*s and standard deviations in parentheses, the number of effect sizes is less than 10 and too few to interpret.

$Eta^2 = .01$.

257

Secondary Education context studies from 1971-75 (N = 10) to 1976-80 (N = 91), with another 40 percent increase to the 1981-85 period (N = 127). The increase in the effect sizes coming from College-University context studies was not as dramatic from 1971-75 (N = 63) to 1976-80 (N = 143), a 127 percent change. And, there was a surprising 63 percent drop from 1976-80 (N = 143) to 1981-85 (N = 53). Our data yield no insights into the reason for that drop. Speculatively, it may be in part due to a shift in researchers' interests, with greater availability of funds for studies at the elementary-secondary level where the thrust for "education for all handicapped children" has been greatest.

There were 10 or more effect sizes for only four treatment contexts—the two noted above and "Inservice" education or training, with N = 54 (8%) and "Other" (N = 14; 2%). Only the means for the first two contexts are presented in Table 55, because of the few effect sizes for the second two. The overall means for the Elementary-Secondary Schooling and College-University contexts (.38 and .40, respectively) are remarkably similar; however, the Inservice mean of .24 (not in Table 55) is considerably lower than both, although the differences are not much above the criterion for triviality (.12). It is of some interest that the Inservice _D_s come basically from three sources: 17 are from studies with elementary school teachers (mean _D_ = .13), 13 are from studies with institutional employees (mean _D_ = .16), and 21 are from inservice with other groups (mean _D_ = .40). (The other three _D_s are accounted for by inservice with special education teachers [N = 1] and police [N = 2].)

Although the Elementary-Secondary Schooling and the College-University mean _D_s are almost identical, there are some interesting within-treatment

Table 55

Treatment Technique Outcomes in Elementary-Secondary Schooling and College-University Contexts

Technique	Elementary-Secondary			College-University		
	N	Mean	SD	N	Mean	SD
Persuasive Message	9	(.80)[a]	(.78)[a]	12	.56	.39
Information Plus Contact	32	.57	.60	47	.55	.73
Direct Contact	22	.69	1.13	46	.36	.53
Vicarious Experience	25	.30	.48	32	.51	.91
Other	27	.49	.56	28	.33	.83
Systematic Desensitization	—	—	--	21	.32	.44
Information	90	.24	.42	94	.35	.62
Information Plus Vicarious	30	.20	.32	21	.28	.36
Persuasive Message, Contrast	—	—	—	11	.13	.33
Positive Reinforcement	—	—	—	2	1.74	.01
Total	235	.38	.60	314	.40	.65

[a]Too few effect sizes (less than 10) to be interpretable.

differences in mean Ds for the two contexts (see Table 55). Of particular interest are the mean Ds for Direct Contact, for which the Elementary-Secondary mean is higher (.69 versus .36, a .33 difference) and for Vicarious Experience, for which the situation is reversed with the College-University mean the higher one (.51 vs. .30, a .21 difference). Only nine Persuasive Message and Systematic Desensitization effect sizes came from studies conducted in an Elementary-Secondary Schooling context.

In short, most studies from which the 644 effect sizes came were conducted in public school or higher education contexts, with very few carried out in other environments, such as at places of employment or in recreation contexts. For those two major context categories, the mean Ds were similar, although both differed somewhat from the mean D for the Inservice context, the only other context with 10 or more Ds, other than the catch-all "Other" with 14. There is some indication of different effects for Direct Contact and Vicarious Experience within the two major context categories, suggesting that context is another source of study variability that should not be ignored.

Setting

Related to, but somewhat different from context is the setting of the research. While "context" refers to general environment, "setting" was defined as the specific type of place where the research was conducted. It is another potential source of variability in study characteristics.

Given the findings to this point, it will come as no surprise that 49 percent of the effect sizes (N = 314) came from studies carried out in regular classrooms (see Table 56), with 57 percent (N = 180) in public school classrooms and 35 percent (N = 109) in higher education classrooms. As would

Table 56

Setting of Treatments

Setting	Effect Sizes (Ds)		
	N	Mean	SD
Regular Classroom	314	.38	.61
Special Education Classroom	35	.21	.68
Institution	22	.34	.77
Hospital	28	.39	.51
Laboratory	23	.46	.66
Individual or Small Group	31	.06	.53
Normal Life	14	.65	.65
Home	2	(.27)[a]	(.03)[a]
Dormitory	4	(.14)[a]	(.54)[a]
Camp	6	(.33)[a]	(.33)[a]
Recreation Facility	1	(.33)[a]	(.00)[a]
Other	32	.32	.38
Combination	71	.51	.54
Can't Tell	61	.39	.71
Total	644	.37	.61

Note. $Eta^2 = .03$.

[a]Too few effect sizes (less than 10) to be interpretable.

be expected. All of the 23 Ds from research in Laboratory settings (15 of which were for Systematic Desensitization as a technique) came from the College-University Context, as did all but five of the Individual or Small Group setting Ds. The low mean D (.06) for Individual or Small Group settings is largely due to a mean D of -.42 for 10 Ds that came from the Other category of treatment techniques.

With the predominance of Regular Classrooms as the research setting from which the 644 effect sizes came, the numbers of other settings within the various treatments are in general quite small. The few worth mention are primarily those in which the presence of persons with disabilities makes the setting particularly appropriate for either Contact or Contact plus Information studies--Special Education Classrooms (13 Contact Ds), Institutions (11 Contact plus Information Ds), Hospitals (19 Contact Ds and 9 Information plus Contact Ds). So, Regular Classrooms dominate the settings, but other settings contribute to variability in study characteristics both within and across types of attitude modification techniques.

Sample Sizes

Data on the size of the experimental treatment group were available for 642 effect sizes (see Table 57). While the range in the size of the treatment groups from which those effect sizes came is large, the modal group was relatively small, N = 20, and the median not much larger, N . Sample sizes for different techniques were fairly similar, except for the small Persuasive Messages mode (6)--which was, however, accompanied by a substantial median (30)--and the high mode (58) for Direct Contact studies. The differences in ranges of sample sizes are, however, striking. Although the minimum Ns are similar, the Persuasive Message (6-44) and Systematic

Table 57

Sizes of Experimental Treatment Groups

	Sample Size Statistics						
Technique	N	Mean	Median	Mode	SD	Range	r[a]
Persuasive Message	23	25	30	6	13	6-44	-.20
Information Plus Contact	79	50	26	26	96	7-862	-.11
Direct Contact	93	69	38	58	104	6-663	-.005
Vicarious Experience	58	35	26	20/52[b]	28	7-198	-.19
Other	71	32	20	20	29	6-148	.31
Systematic Desensitization	21	26	21	20/45[b]	13	13-45	.02
Information	202	43	26	20	58	8-544	.03
Information Plus Vicarious	62	73	37	31	98	13-424	.04
Persuasive Message, Contrast	11	20	19	20	8	13-45	--[c]
Positive Reinforcement	2	10	10	10	.00	—	—
Total	642	47	28	20	73	6-862	-.02

Note. Statistics rounded to whole numbers to reflect that samples are constituted of whole persons. Because of rounding, totals may not agree exactly.

[a]Correlation coefficient for sample size and D.
[b]Bi-modal distributions.
[c]Coefficient not reported because 10 of the 11 data points are from one study.

Message (13-45) studies had much smaller maximum treatment group Ns than the others, with the largest maximum Ns for Direct Contact and Information Plus Contact (663 and 862, respectively). Those differences undoubtedly reflect the predominance of laboratory studies for the first two techniques and course and program evaluation studies for the latter two.

The correlation between Ds and experimental group Ns was practically nil for the total data set (r = -.02). For individual treatment techniques with sufficient numbers of Ds to compute a coefficient, the r's ranged from essentially zero (Information [.03], Direct Contact [-.005], Information Plus Contact [-.11], Information Plus Vicarious Experience [.04], and Systematic Desensitization [.02]) to low and negative (Vicarious Experience [-.19], Persuasive Messages [-.20]), or low and positive (Other [.31]). In the case of the Other category, two outliers (high Ds and Ns) boosted the r.

Limited ranges of Ns do not appear to account for the low coefficients, nor do differences in ranges of Ns for different techniques (Table 54) appear to be a factor in the differing coefficients. The low negative coefficients for Persuasive Message and Vicarious Experience suggest that the size of the treatment groups might be a factor worth further consideration with those treatments. Overall, however, it does not appear that differences in treatment group Ns were systematically related to the outcomes with different attitude techniques.

Summary

Variability in study characteristics other than treatment features are another source of variability among studies grouped for analysis in meta-analytic integrative reviews. All such differences ought to be kept in mind as posing potential "apples and oranges" difficulties in interpretation, even though some do not appear to interact systematically with treatments to affect outcomes. For the study characteristics discussed above (type of comparison, time of posttest, type of dependent measure, context, setting, and sample size), there appear to have been few systematically different effects for different treatments, and those that are discernible are often confounded with other sample attributes. Nevertheless, variability in study characteristics should be an ever-present consideration as our results are reviewed. Further analyses are warranted in an attempt to ferret out the influences which variations in such study characteristics had on study outcomes.

Sample Characteristics

In prior sections, some variations in treatment features within attitude modification technique categories and in other study characteristics have been presented. Next, information will be presented on some attributes of sample Ss that were coded because they might influence outcomes or be important to generalization.

Sample Selection

Information on sample selection does not provide direct evidence of Ss' characteristics. However, the methods by which subjects are obtained may well influence the nature of the sample and, thereby, both the research outcomes and their generalizability. The use of convenient samples in psychological research--frequently sophomores in psychology classes--has often been commented on. The equivalent in educational research is the "intact" group--the conveniently available classroom of students. The effects of using convenient, or intact, groups has not been systematically investigated (except for the impact on statistical inference, e.g., Hopkins, 1982). Another selection technique, the solicitation of volunteers, has been addressed extensively by Rosenthal and Rosnow (1975; Rosnow & Rosenthal, 1976). Their summary of the research evidence indicated that volunteers for research projects do tend to differ from nonvolunteers in ways that could have significant implications for research on modifying attitudes toward persons with disabilities. For example, volunteers are likely to be more intelligent, higher in need for social approval, and less authoritarian than nonvolunteers.

As might be expected (see Table 58), the samples for the studies which yielded our 644 effect sizes came from two major sources: volunteers (N = 220; 34%) and intact groups (N = 294; 46%), together accounting for 80 percent (N = 514) of the effect sizes. For the other 20%, random selection of individuals (N = 31) or of groups then used as the unit of analysis (N = 31) accounted for 62 effect sizes (10%); for 33 effect sizes (5%), selection was categorized as "Other"; and, for 35 effect sizes (5%), the method of sample selection could not be identified.

The mean <u>D</u>s for the different sample selection methods are somewhat perplexing, especially the mean <u>D</u> of .25 for Volunteers, as contrasted to a mean <u>D</u> of .42 for Intact Groups and .53 for Random Samples. Wouldn't volunteers be expected to respond more favorably to efforts to modify attitudes toward a minority group? Alternatively, however, volunteers may have come into modification programs with attitudes already so positive that experimental effects were dampened. In that regard, as did Rosenthal and Rosnow (1975, e.g., p. 49), we included as volunteer subjects not only those who responded to a solicitation to participate in a research project but

Table 58

Methods of Sample Selection by Attitude Modification Technique

Selection

Technique	Can't Tell	Random Samples	Volunteers	Intact Groups	Other	Total
Persuasive Message	—	(.35)	(.60)	(.59)	(1.71)	.67
	—	(.11)	(.46)	(.17)	(.75)	.56
	--	6	9	5	3	23
Information Plus Contact	(1.38)	.67	.40	.50	(.36)	.51
	(.00)	.60	.98	.61	(.01)	.66
	1	11	12*	74**	2	100
Direct Contact	(.34)	--	.28	.49	—	.43
	(.17)	—	.47	.82	—	.73
	5	—	21*	67**	—	93
Vicarious Experience	(-.09)	(.83)	.23	.52	—	.40
	(.00)	(.20)	.75	.80	—	.76
	1	7	30**	20	—	58
Other	(.36)	(.31)	.15	.95	(.75)	.39
	(.28)	(.36)	.53	.79	(.53)	.64
	8	4	39**	15*	5	71
Systematic Desensitization	—	—	.33	—	.30	.32
	—	—	.42	—	.55	.44
	—	—	16	—	5	21
Information	.43	.44	.20	.23	.47	.29
	.36	.38	.50	.56	.45	.51
	10	31	47*	97	18	203
Information Plus Vicarious Experience	.07	(.80)	.19	.20	—	.20
	.14	(.36)	.40	.26	—	.36
	10	3	33	16*	—	62
Persuasive Message, Contrast	—	—	.13	—	—	.13
	—	—	.33	—	—	.33
	--	--	11	—	—	11
Positive Reinforcement	--	—	(1.74)	—	—	1.74
	—	—	(.01)	—	—	.01
	—	--	2	—	—	2
Total	.31	.53	.25	.42	.59	.37
	.34	.42	.60	.67	.60	.61
	35	62	220	294	33	644

Note. Eta^2 for sample selection method and $\underline{D}$ = .04.

*At least 10 fewer cases than expected, based on marginal frequencies.
**At least 10 more cases than expected, based on marginal frequencies.

those who volunteered to participate in activities with disabled persons without knowing they were to be part of a research project. Such persons might have had even higher initial attitudes than would be expected on the general basis of estimated volunteers' traits.

The effects of prior attitudes on outcomes could not be investigated with our data set because there were practically no reports of analyses in which pretreatment attitudes were included as an independent variable, with results reported by levels of antecedent attitudes. The hypotheses that volunteers have higher antecedent attitudes, with higher pretreatment attitudes associated with less change, suggested by our Selection Method finding, could be given a rough check by coding pretreatment attitude levels for volunteers and nonvolunteers in different studies and then comparing Ds for the two groups. That is among the many interesting questions that have arisen from the analyses reported here that call for further use of our extensive data set and collection of research reports.

Of more concern in this section is whether sample selection was related to attitude modification technique outcomes. An answer is, of course, made difficult by the confounding of variables. For example, as indicated in Table 58, intact groups were more heavily represented than volunteers in Information Plus Contact, Direct Contact, and Information effect sizes, with the Volunteer mean D lower in each case. But 47 percent of the Information Plus Contact effect sizes, 49 percent of the Direct Contact effect sizes, and 46 percent of the Information effect sizes came from college and university samples, which are more likely to be obtained through solicitation of volunteers than are, for example, elementary and secondary school samples. (While 40% of the college and university effect sizes came from volunteer samples, only 22% of the elementary and secondary school effect sizes did; or, conversely, while 57% [N = 125] of the effect sizes that came from volunteer samples [N = 22?] were for college or university students, only 24% [N = 53) were for elementary or secondary school students.) The possible effects of methods of sample selection are a relevant consideration in any interpretations of our descriptions of our data set, even though sample selection is confounded with other variables. The nature of those effects is also a viable topic for further research in this attitude modification field.

Grade-Age Levels

The age of the Ss was a sample characteristic that we presumed might be related to treatment outcomes and would be of interest to our readers. However, in coding studies to try out the coding instrument during its development, we found that few authors reported the age of their Ss. Consequently, information of the schooling grade level of Ss was coded because it is a fairly close proxy for age, at least through the undergraduate years of college. Information from that coding is presented in Table 59.

As has been noted already in the discussion of the contexts within which the attitude modification studies were conducted, 85 percent of the effect sizes (N = 549) came from studies carried out in elementary-secondary school and college-university environments. Those figures are reflected in Table 59, with 39 percent (N = 254) of the Ds coming from studies conducted with Ss from preschool through high school and another 43 percent coming from studies with undergraduates (N = 253) and graduate students (N = 29). It is worth noting, too, that 43 percent (N = 110) of the elementary-secondary level Ds are in the Intermediate category (grades 4-6), with another 10 percent (N =

Table 59

Educational-Age Levels of Experimental Treatment Ss

	Effect Sizes (Ds)		
Level	N	Mean	SD
Preschool	3	(.62)[a]	(1.06)[a]
Primary	30	.48	.50
Intermediate	110	.34	.69
Middle School	18	.33	.42
Junior High	13	.41	.26
Senior High	18	.42	.56
Combination	62	.40	.46
Subtotal	254		
Undergraduate	253	.43	.69
Graduate	29	.33	.45
Postprofessional	52	.23	.53
Adult Not in School	47	.19	.43
Can't Tell/Other	9	(.57)[a]	(.73)[a]
Total	644	.37	.61

Note. $Eta^2 = .02$.

[a]Too few effect sizes (less than 10) to be interpretable.

62) in the Combination category, with many combinations that included Intermediate Ss.

In terms of mean Ds, the most apparently interesting aspect of Table 59 is the difference between the mean D for the youngest Ss represented by at least 10 effect sizes, the Primary Ss (grades K-3; mean D = .48) and the older Ss, with mean Ds of .33 (Graduate), .23 (Postprofessional), and .19 (Adult Not In School). Of the 30 Ds for Primary Ss, 14 are from Information study effect sizes—the only treatment technique with 10 or more Ds—with a mean D of .32. The mean D of .48 is the result of 6 Primary Ds in the Information Plus Contact category with a mean of .78. In short, the apparently large mean differences are based on the influence of a small subset of Ds. With that in mind, the mean Ds for various grade-age levels are, overall, surprisingly similar, with the exception of the two adult groups with the lowest means—Postprofessional (mean D = .23) and Adult Not in School (mean D = .19). This similarity is captured by the Eta^2 of .02 for Educational-Age Levels and Ds.

What about differences in grade-age level by treatments? Table 60 contains mean Ds for treatment techniques at each grade-age level. Purposely, only means have been included for which there were at least 10 effect sizes to make more graphic the pattern, including absences, of mean Ds. As noted in earlier sections, Information is clearly the most investigated technique, followed by Information Plus Contact. Just as clearly, the findings come primarily from Ss in the Intermediate grades (with many grade Combinations including intermediate Ss) and from Undergraduates.

A few treatment differences by grade-age level are worth noting. For example, "adult" (Graduate, Postprofessional, and Adult Not In School) Ss

Table 60

Mean $\underline{D}$s for Treatment Techniques and Grade-Age Levels with At Least 10 Effect Sizes

	Grade-Age Level											
Technique	Preschool	Primary	Intermediate	Middle School	Junior High	Senior High	Combination	Undergraduate	Graduate	Postprofessional	Adult	Other
Persuasive Message												
Information Plus Contact			.59 15					.69 36	.09 12		.22 13	
Direct Contact			.81 16				.20 12	.40 44				
Vicarious Experience			.25 17					.46 25				
Other			.20 13					.32 26			.29 11	
Systematic Desensitization								.16 15				
Information		.32 14	.20 36	.20 10			.44 12	.37 79		.19 27		
Information Plus Vicarious Experience			.12 13				.21 22					
Persuasive Message, Contrast												
Positive Reinforcement												

Note. The first number in each cell is the mean $\underline{D}$, the second number is N. Only means based on at least 10 effect sizes are included.

received lower mean Ds than other Ss for Information and Information Plus Contact, but not for Other (technique combinations other than those labeled in the table). Also, the Direct Contact mean D for Intermediate Ss is strikingly higher than those for Combination and Undergraduate Ss. The mean differences of .41 and .61, respectively, are just below and above the criterion of .50 (see Chapter 3) for high practical significance. And, the opposite result is seen for Vicarious Experience, with the Intermediate mean D (.25) .21 lower than that (.46) for Undergraduates.

All in all, the information in Table 60 suggests the caution that is necessary to avoid overgeneralizing the findings in the literature to differing grade-age levels. Further analysis of our data will be undertaken to determine if other sample or study characteristics explain the differences. No analysis, however, can make up for the number of treatment technique-grade-age level combinations for which adequate numbers of Ds are lacking.

Gender

Prior reviewers (e.g., Horne, 1985, pp. 132, 143) have indicated that females tend to have more positive attitudes toward persons with disabilities, and that they may be more likely to change attitudes in a positive direction. In order to determine if gender was an important factor in study outcomes, we recorded the percentage of males in the experimental group whenever information of gender composition was available in the research report.

Percentages of males in the experimental groups are presented in Table 61 for the 339 effect sizes for which that information was available in the reports. Although the mean and median percentages for the total 339 effect

Table 61

Percentages of Males in
Experimental Groups

	Statistics					
Technique	N	Mean	Median	Mode	SD	Range
Persuasive Message	14	21	16	00	23	0–55
Information Plus Contact	57	28	23	00	25	0–84
Contact	38	28	29	00	19	0–61
Vicarious Experience	21	34	46	00	28	0–100
Other	29	44	42	40	11	18–73
Systematic Desensitization	16	11	7	00/7*	14	0–33
Information	115	40	46	00/50*	27	0–100
Information Plus Vicarious	37	41	52	00	29	0–100
Persuasive Message, Contrast	10	40	45	—**	16	21–60
Positive Reinforcement	2	30	30	30	00	--
Total	339	35	36	00	25	0–100

Note. Statistics rounded to whole percentages, so columns and totals may not agree exactly.

*Bi-modal distribution.
**Multi-modal, with only 2 Ds for each %.

275

sizes are nearly identical (35% and 36%, respectively), the mode is zero. That is, the most frequent occurrence was to have no males in the treatment group. That mode is consistent across treatment techniques as well, with the exception of the Information and Persuasive Messages, Contrast modes. Within treatments, Systematic Desensitization stands out as having the most restricted range, with 33 the maximum percentage of males in a treatment group. Also, the Persuasive Message range is from 0% to 55%, also a restricted range.

Correlations between percentages of males in the experimental group and Ds for the 339 effect sizes are reported in Table 62. Overall, there was no relationship between percentage of males and outcomes ($r = .00$). The coefficients for the various treatment techniques range from moderately negative* ($r = -.47$, $r^2 = .22$ for Systematic Desensitization)** to low and positive ($r = .31$, $r^2 = .09$ and $r = .27$, $r^2 = .07$ for Contact and Persuasive Messages, respectively), with most of the coefficients so low as to indicate negligible relationships.

The picture is anything but clear, with the overall lack of relationship, some slight, positive relationships for individual techniques, and the negative relationship for Systematic Desensitization, with its restricted range. Surely, gender is a factor to be considered by those doing further research, or attempting to apply, that technique. On the other hand,

---

*Recall that % of males was recorded, so a negative relationship indicates that with more males in the experimental group, Ds tended to be lower; conversely, a positive relationship indicates that with more males in the experimental group, Ds tended to be higher.

**The $r = -.73$ for Persuasive Messages, Contrast is being ignored in this discussion because it is based on only 10 effect sizes, all from the same study.

Table 62

Correlations Between Percentage of Males in Experimental Groups and Ds

Technique	Statistics N	r	r^2
Persuasive Message	14	.27	.07
Information Plus Contact	57	-.03	.00
Contact	38	.31	.10
Vicarious Experience	21	.17	.03
Other	29	.10	.01
Systematic Desensitization	16	-.47	.22
Information	115	-.03	.00
Information Plus Vicarious	37	.03	.00
Persuasive Message, Contrast	10	-.74[a]	.55
Positive Reinforcement	—	—	—
Total	339	.00	.00

[a]Because all of the data came from one study with 10 effect sizes, this coefficient is ignored in the discussion of this table.

the positive, even if low, coefficients for Persuasive Message and Contact suggest that there are situations in which males may be more responsive to attitude change efforts, even though as a general matter gender seemed to bear little relationship to outcome.

In order to investigate gender effects, we coded two other types of information. First, when a complex analysis of variance was reported with gender and treatment as factors, we recorded, as an effect size, the Eta^2 for the interaction if it, or information to compute it, was available. Second, where Ds could be computed separately for males and females within a treatment by control, treatment by placebo, single-group, pre-posttest, or treatment A versus B comparison, we did so in order to analyze those Ds for differential treatment effects.

The mean Eta^2 for the 36 available treatment by gender interactions was .02, with a standard deviation of .05. Clearly, not much variance was explained by that interaction, which is consistent with the low overall r reported above and with the lack of striking differences in coefficients by treatments. There were too few treatment by gender interactions within any one treatment for interpretation (the largest Ns were for Information, N = 8, and Information Plus Vicarious Experience, N = 7).

It was possible to compute separate Ds for males and females for 24 comparisons. The mean D for females was .41 and for males, .33 (the standard deviation was .49 for each). The mean difference of .08 is too small to be considered anything but trivial, again consistent with the overall r of .00 for percentages of males and Ds. Ten of the 24 Ds for males and females were for the Information treatment technique. The means for females and males were .38 and .31, respectively--again a trivial difference (.07), and consistent with the Information r for percentages of males and Ds of -.03. For no other treatment technique was the N for a gender with treatment analysis greater than 6.

In short, our analyses indicated little consistent evidence that gender is related to attitude change in our data set. Contradictory correlations were obtained for Persuasive Messages and Contact, on the one hand (higher Ds as % of males increased) and Systematic Desensitization (lower Ds as % of males increased), on the other. Analyses of interactions between gender and type of attitude modification technique have not been reported frequently in the literature and deserve more attention in research and application efforts.

Prior Contact

The extent and type of prior contact that Ss have had with persons with disabilities is another variable with potential power for mediating the effects of treatments. Although that factor was ignored by prior reviewers (see Chapter 2), we coded prior contact information for our integrative review. Unfortunately, however, the data did not yield much information about the relations of Ss' prior contacts to treatment outcomes.

The basic reason for the lack of information is that assessment of prior contact was reported for only 260 out of 644 effect sizes (40%). For another 29 effect sizes (4%), prior contact was implicit—e.g., inservice education with experienced psychiatric nurses or special education teachers. For 46 (18%) of the 289 effect sizes for which prior contact was assessed or implicit, no mention of use of the information was made in the report. In the reports for 125 effect sizes (45%), the prior contact information was used only to describe the sample. In studies from which 22 Ds came, prior contact was used to exclude Ss (i.e., to ensure that no Ss had prior contact); for 45 Ds, prior contact was used as a criterion for inclusion in the sample; and, for 14 Ds, Ss were selected from strata of prior contact.

Prior contact was used as a covariate in the analyses for 27 effect sizes; however, the relationship between prior attitudes and outcomes was hardly addressed. Of interest, for example, would be the relationship between prior contact and attitude change in the treatment group or, better yet, a comparison of that relationship for treatment and control groups. For only 4 effect sizes was there also a report of the correlation between prior contact and posttest attitude scores (N = 2) or prior contact and attitude change scores (N = 2). And, for only 4 effect sizes was there a report of such correlations for a treatment and control group.

Of course, comparing treatment-control group correlations is one way of determining if there was a prior contact-by-treatment interaction effect. Alternatively, prior contact could be used with treatment as factors in a complex analysis of variance to determine prior contact-treatment interactions. Such interactions were coded as secondary effect sizes on our coding instrument. Again, the information was too sparse to be of use, with only one effect size ($Eta^2 = .02$).

In short, although prior contact is a potentially important variable, it received so little attention in our population of research reports that nothing can be said about the extent to which it might have mediated treatment effects.

Personality

The characteristics of nondisabled Ss were mentioned as an important theoretical factor in the discussion, earlier in this chapter, of Contact as an attitude modification technique. Personality attributes, such as authoritarianism, might well be important factors in the effects of attitude change efforts. Consequently, we coded analyses of interactions between levels of personality and treatment as secondary effect sizes. As with the Contact studies, in which relationships between personality variables and attitude change were not analyzed, only 3 effect sizes for personality by treatment interactions were identified for the total data set. Although the mean Eta^2 was .01, there were too few effect sizes to provide any information worth interpreting.

Summary

In this chapter, we have presented the major results from the analyses of the data from our population of research reports. The intent has not been to arrive at a statistical estimate of an overall or individual treatment effect. The purpose has been to describe the major dimensions of the body of literature on modifying attitudes toward persons with disabilities to determine if there are indications of effectiveness and with what variables treatment effectiveness might vary.

A summary of the conclusions of research report authors in regard to the effectiveness of the treatments they had investigated indicated that it would be unlikely that clear-cut treatment effects would emerge from analyses of the data set. That expectation was borne out when the mean *D*s (pooling treatment by control, treatment by placebo, and single-group, pre-posttest comparisons) for each of several attitude modification techniques were compared. The mean *D*s could be arranged in rank order, with considerable spread from the highest to the lowest. However, the differences between mean *D*s were, for the most part moderate to small, and the heterogeneity of outcomes, as compared to the differences between mean Ds, was so great that no treatment could be said to have produced overall results clearly superior to any other treatment. Moreover, research with each technique produced both negative changes for treatment groups and negative effect sizes, raising further questions about the uniformity of effects.

To illuminate that heterogeneity and some of the variables that might be associated with it, we reported analyses of concomitant variables: study quality, treatment variations, other study characteristics, and sample characteristics. None provided a clear interpretive path for the heterogeneous results.

Study quality, as indicated by treatment validity, internal validity, and dependent measure validity, was not associated with study outcomes to any appreciable degree. In part, the lack of association was due to lack of variability, with few of the reports receiving high ratings on any of the three quality indicators. It may well be that the ambiguous and conflicting results noted above and throughout the chapter are due to the inadequacies in the primary research studies. However, lacking a large number of high quality studies, our data cannot be used to address that issue. It can only be said that an analysis of outcomes from medium and low quality research has not produced clearcut answers to the questions about attitude modification that provided the impetus for this integrative review.

The analyses indicated considerable variability in treatment features, other study characteristics, and sample characteristics. But the associations of the variables with outcomes were in general very low. There were indications that some study and sample characteristics were differentially represented in the 10 groups of treatment techniques, with some differential effects. However, the confounding of variables, with some nested within each other as well as within treatments, made it difficult to disentangle the nature of covariations. Bangert-Drowns' (1986) portrayal of the general situation in summarizing psychological research provides an apt summary of what this chapter indicates about the research literature on modifying attitudes toward persons with disabilities:

> Research outcomes vary in ways that make generalizable interpretations difficult. Such variation comes from a number of sources. It may reflect real population variation, the effects of different treatment features or study settings, sampling error, selection biases of the reviewer, publication biases, the effects of erroneous or insufficient reporting (unreported spurious influences, computational errors, typographical errors), differing degrees of validity and reliability in the outcome measures, and differences in the range or intensity of the independent variable. The task is enormous, but the power of social scientific inquiry would greatly increase if patterns could be found amid this outcome variation. (p. 396)

The patterns are not yet clear for the body of research we have reviewed. But the review in this chapter has produced suggestions for further, and better, research (discussed further in Chapter 7). The data set will itself be submitted to further analyses in efforts to find regularities. The results presented in the next chapter suggest alternative approaches to the meta-analysis of sets that may also be worth pursuing further.

CHAPTER 6

ALTERNATIVE AND SUPPLEMENTARY DATA EXPLORATIONS

One mildly frustrating challenge in carrying out an integrative review from a quantitative perspective, with a rich data set gathered from a heterogeneous and fairly large population of studies using a comprehensive instrument, is the selection of analyses to be conducted. The matter of selection becomes particularly pressing when there are time constraints, as is inevitable with a funded project. The main analyses, the results of which were presented in Chapter 5, were conducted from a meta-analytic stance consistent with that advocated by Glass (see Chapter 3). The major focus was on individual outcome effect sizes, rather than on studies, and the effect sizes were based on mean differences as indicators of treatment effects. Alternative analytic stances and indicators of treatment effects are reflected in the preliminary results from alternative and supplementary analyses that are reported briefly in this chapter.

Median Effect Sizes

As noted in discussing the Results section of our coding instrument (Chapter 3), there has been criticism in the literature of the meta-analytic strategy of using individual outcome effect sizes as the unit of analysis when studies yield multiple effect sizes. Recognizing that concern, we computed median Ds for studies in order to explore the effects of using study effect sizes in our data analyses. Whether preliminary analyses would indicate differences in results with median Ds versus individual outcome Ds was of particular interest.

Procedures

The synopsis by Bangert-Drowns (1976) of computing a "study effect unit of analysis" sounds procedurally straightforward:

> If a study uses more than one dependent measure, the corresponding effect sizes are either combined if they represent the same construct (e.g., academic achievement) or they are sent to separate analyses if they represent different constructs (e.g., one analysis for academic achievement data, another for attitude toward school data). (p. 393)

However, the procedure becomes complicated when, within each study, the construct--in this case, attitudes toward persons with disabilities--is assessed with quite different types of measures, there are intra-study replications, and there is repeated posttesting. Encompassing all of these variations in one median effect size per study is as likely to lead to misinformation and incorrect conclusions as is the lack of independence of multiple effect sizes from individual studies.

As noted above, in this review the analytic focus was on individual outcome effect sizes. During the funded portion of the project being reported here, adequate time was not available to isolate and treat the various median Ds that had been computed and recorded (by type of assessment and across type of assessment, by replication and time of posttest) in a complete alternative analysis. Instead, one median effect size was selected for each of the studies from which our 644 treatment by control, treatment by placebo, and single-group, pre-posttest Ds came. Used was the overall median D recorded for the first effect size entered for each study. In other words, the medians to be analyzed were for the first-recorded of any intra-study replications and for immediate, rather than follow-up, posttest data. That choice was justified by the lack of differences in mean Ds by replication or time of posttesting (as reported in Chapters 4 and 5). A criterion of 5 was used as the minimum number of medians that had to be available for a mean and standard deviation to be considered adequately stable to be interpretable (as contrasted with 10 for the individual effect size data).

Results

The 644 effect sizes in our main analyses came from 200 studies. Information on the median Ds selected for those 200 studies is presented in Table 63. As can be seen in that table, with median Ds (study effect sizes) as the unit of analysis, the Ns for two more treatment techniques*--Systematic Desensitization and Persuasive Messages, Contrast--were not adequate for interpretation purposes.

The most interesting aspect of Table 63 is that the use of study effect sizes (median Ds) rather than individual effect sizes had little impact on the magnitude of the mean Ds for the treatment techniques or the rank order of the mean Ds. The two exceptions are Systematic Desensitization, which was dropped from the ordering because of a low N, and Other, for which the mean D increased from .39 to .59 (with practically no change in the standard deviation). There is little evidence in Table 63 that using effect sizes from individual findings, rather than study effect sizes, influenced our results appreciably. Nevertheless, further analyses of our data set using median Ds are warranted.

Outliers

Extreme effect sizes, or outliers, are a matter of concern in conducting a meta-analytic type of integrative review. Outliers may have an undue

*There was not an adequate number of Ds for Positive Reinforcement to be considered in the earlier, individual outcomes, analyses.

Table 63

Individual Finding and Median Study
Effect Sizes (*Ds*) for the Treatment Techniques

Technique	Individual Effect Sizes (*Ds*) N	Mean	SD	Median Effect Sizes N	Mean	SD
Persuasive Message	23	.67	.56	7	.73	.49
Information Plus Contact	100	.51	.66	39	.55	.49
Direct Contact	93	.43	.73	35	.50	.76
Vicarious Experience	58	.40	.76	19	.39	.64
Other	71	.39	.64	21	.59	.67
Systematic Desensitization	21	.32	.44	2	(.46)[a]	(.24)[a]
Information	203	.29	.51	56	.22	.42
Information Plus Vicarious	62	.20	.36	18	.23	.41
Persuasive Message, Contrast	11	.13	.33	2	(.01)[a]	(.13)[a]
Positive Reinforcement	2	(1.74)[b]	(.01)[b]	1	(1.75)[a]	(.00)[a]
Total	644	.37	.61	200	.42	.58

Note. Eta^2 for individual effect sizes = .05; for median effect sizes = .11.

[a]Fewer than 5 median effect sizes, so not considered interpretable.
[b]Fewer than 10 effect sizes, so not considered interpretable.

effect on results, especially if clustered within one or a few analysis categories, such as treatment technique. Such effects are a special concern because of the suspicion that unusually large effect sizes may be due to study characteristics--i.e., threats to internal validity--that have invalidly magnified outcomes.

Definition

There are no clear guidelines for defining outliers--that is, for deciding what is an extreme score. In arriving at a definition, we first inspected a frequency distribution for our 644 Ds to determine if there were clear breaks at both ends of the distribution, particularly at 1, 2, and 3 standard deviations from the mean. None was evident at one standard deviation. Moreover, that distance from the mean did not seem adequately extreme to be considered the dividing point for outliers. At two standard deviations, there was a clear break at the negative end of the distribution and a perceptible break at the positive end. Moreover, in a normal distribution, about 95 percent of the distribution would fall within those points. The Ds that fell beyond two standard deviations below the mean constituted .7 percent of the distribution; 4.3 percent were more than 2 standard deviations above the mean. So, despite the obvious lack of normality, by defining "outlier" as a D more than 2 standard deviations from the mean, 5 percent of the effect sizes (N = 33) were excluded from the analyses. We accepted that definition.

Results

As can be seen in Table 64, the effects of dropping the 33 outliers were minor. As would be expected, because they were more positive than negative

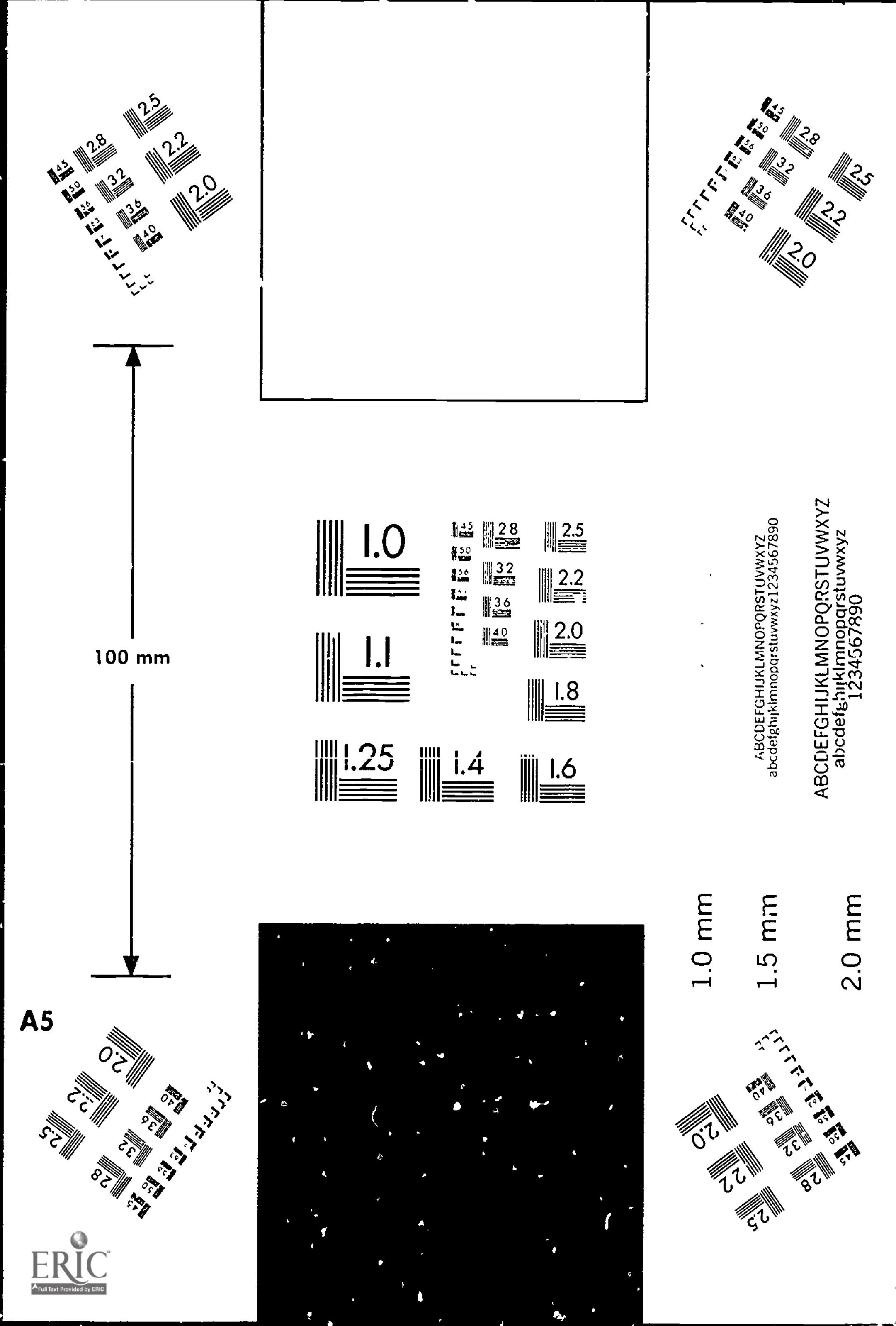

100 mm
A5
1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz1234567890
ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz
1234567890
1.0 mm
1.5 mm
2.0 mm
ERIC
Full Text Provided by ERIC

Table 64

Results for Treatment Techniques
with "Outlier" Ds Excluded

	Effect Sizes (Ds)					
	All			Outliers Excluded		
Technique	N	Mean	SD	N	Mean	SD
Persuasive Message	23	.67	.67	21	.54	.34
Information Plus Contact	100	.51	.66	94	.46	.48
Direct Contact	93	.43	.73	87	.28	.36
Vicarious Experience	58	.40	.76	53	.36	.54
Other	71	.39	.64	65	.30	.47
Systematic Desensitization	21	.32	.44	21	.32	.44
Information	203	.29	.51	197	.28	.43
Information Plus Vicarious	62	.20	.36	62	.20	.36
Persuasive Message, Contrast	11	.13	.33	11	.13	.33
Positive Reinforcement	2	(1.74)[a]	(.01)[a]	—	—	—
Total	644	.37	.61	611	.32	.44

[a]Too few effect sizes (less than 10) to be interpretable.

outliers, the overall mean D dropped, but only from .37 to .32. As should be the case when the range is restricted, the impact on the standard deviation was more noticeable, dropping from .61 to .44. Because the mean D for the highest ranked technique, Persuasive Messages, decreased considerably--from .67 to .54--the overall spread among means for the treatment groups was reduced; but the rankings remained basically the same, except for Direct Contact. That mean D dropped from .43 to .28, moving it from the third rank to a tie for the sixth rank. There also was a minor shift in the order of the Other and Systematic Desensitization means.

It is also worth noting that eliminating outlier Ds did reduce the magnitude of differences between the means for types of comparisons (see Table 65). Although the mean D for each comparison type was reduced, deleting outliers had the greatest effect on the Pre-post mean. Even that change in means (.12), however, was right at the criterion for a trivial difference. The range for the three comparison types dropped from .13 to .05. This negligible result, like those for the treatment technique mean Ds, suggests that considering the more extreme Ds to be a legitimate part of the data set had only a marginal impact on the findings.

Investigation of the reasons for outliers could be revealing in terms of methodological and treatment considerations. In particular, the reduction in the mean Ds for two treatment techniques, Persuasive Messages and Direct Contact, bears further attention. The studies from which the outlier Ds came will be examined in a later phase of this study to determine if there were especially serious threats to internal validity or other aspects of treatment implementation that seem to explain the extreme outcomes.

Table 65

Comparison Type Effect Sizes
With Outlier Ds Eliminated

	Effect Sizes (Ds)					
	All			Outliers Excluded		
	N	Mean	SD	N	Mean	SD
Treatment versus Control	497	.36	.58	477	.32	.42
Treatment versus Placebo	49	.29	.70	45	.20	.49
Pre-post	98	.49	.72	89	.37	.50
Total	644	.37	.61	611	.32	.44

Treatment A vs. B Effect Sizes

As noted in Chapter 4, a basic analytic strategy adopted for this review was to use effect sizes based on the comparison of a treatment against the absence of treatment, in the form of a control or placebo condition or pretest data. Consequently, effect sizes from Treatment A versus Treatment B (A vs. B) comparisons were excluded from our main analyses. It was our hope, however, that a large number of Treatment A versus Treatment B effect sizes would come from studies in which different attitude modification techniques, rather than variations of the same technique, were compared. Those comparisons would be directly relevant to our main analyses. For example, with the mean D for Information Plus Contact (.52) only slightly higher than that for Direct Contact (.43) and considerably higher than that for Information (.29), it would be of interest whether direct comparisons of these techniques yielded a similar ordering of outcomes?

The hope for such analyses turned out to be unrealizable. Of 61 A vs. B Ds, only 14 came from direct comparisons of treatment techniques--12 (from 3 studies) were for Information Plus Contact versus Information and 2 (from one study) were for Contact versus Information.

Results

Clearly, given the small number of Ds and, even more important, the small number of studies, not much credence can be put in summaries of those few A vs. B results. However, rather than to simply ignore what information is there, and to satisfy the curious reader, it is worth noting that of the 12 Information Plus Contact versus Information comparisons, 11 favored Information Plus Contact and 1 did not.

The mean D for these 12 comparisons is .46, with a standard deviation of .33. One study with a sample of university students in an introductory special education course yielded 3 Ds, all positive* (median D = .44); another with a sample of undergraduate students in elementary and secondary education, yielded 6 Ds, all positive (median D = .56). The third study involved sixth, ninth, and twelfth graders; and, an effect size was computed for each grade level. The Ds for the sixth and ninth graders were positive (.29 and .46, respectively), but the twelfth-grade D was negative (-.52). The author of that study (Mulkey, 1980) suggested that the film used for information (A Different Approach) may have required too much sophistication of the sixth and ninth-graders, as contrasted with gaining information through contact and discussions with a disabled person in a wheelchair.

What might one conclude from these few studies and Ds? Not much. There is some confirmation of the rank ordering of Information Plus Contact and Information as attitude modification techniques. But there also is an indication that variations in treatments can be important, as illustrated by the use of a film apparently too subtle for younger students, but more effective than Contact Plus Information for twelfth graders. But with the small Ns, extreme caution must be exercised in even considering these results as suggestive.

Mainstreaming Studies

At the beginning of this project, it was thought that we might locate a number of studies of the effects of mainstreamed classrooms on the attitudes

---

*D's were computed so that a positive D indicates a greater mean gain for the Information Plus Contact group.

of nondisabled students toward persons who have disabilities. For several reasons--the likely long-term duration of mainstreaming contact, the likelihood that research in ongoing school settings would be less well-controlled, and because specific attributes of mainstreaming might be of special interest (such as the pre-mainstreaming preparation of the parents of nondisabled as well as disabled children)--it was decided to code studies of attitude change in mainstreamed classrooms separately. For this purpose, mainstreaming was defined (Conventions, see Appendix C) as:

> a systematic, sustained effort to integrate disabled students in regular classrooms for part or all of their instruction [for one or more periods a day], as contrasted with bringing disabled students into a regular classroom temporarily to provide contact as part of a research project. (pp. C-2/CI 4, C-21/CI 9)

Surprisingly, we found few studies that fit this definition and in which attitudes toward persons with disabilities were assessed (as contrasted, for example, with attitudes toward mainstreaming). Effect sizes from studies involving other types of in-school contact--such as presence of a special classroom in a school, with incidental contact, or with planned contact at lunch and on the playground, through visits to the special classroom, or through providing special assistance to the children with disabilities--were coded under the appropriate attitude modification techniques (Contact, sometimes in combination with Information or Vicarious Experience). Consequently, our analysis of mainstreaming effects turned out to be a supplementary, rather than a mainstream, effort.

Results

Of 20 mainstreaming effect sizes (from 9 studies), 6 were comparisons of different versions of mainstreaming (what we have termed A vs. B studies), leaving only 14 effect sizes from 7 studies. All of the 14 effect sizes came from Ss who were junior high age or below--6 from the intermediate grades (grades 4-6), 1 from junior high school, 5 from grades K-6, and 2 with samples from intermediate through junior high school grades--with too few Ds in any of the grade-age level categories for separate analysis. The mean D was .19, with a standard deviation of .43 and a range from -.81 to 1.05. (The statistics for median Ds were similar, although indicating somewhat greater homogeneity: mean = .23; standard deviation = .32; range, -.33 to .72.)

With the small numbers of effect sizes and studies, it was not feasible to break down the data to analyze the relationships of coded variations in mainstreaming to outcomes. In any event, such analyses would not have been fruitful because of the lack of information in the reports for coding (Can't Tell was the most frequent rating for Type of Instruction; Minutes Per Day, Days Per Week, and Minutes Per Week in Mainstreamed Classroom) and the lack of variation when information was available for coding (None was the most frequent rating for Special Personnel Support [N = 12; Can't Tell = 2], Special Skills Training for Disabled Students [N = 12; Can't Tell = 2], and Special Instruction for Nondisabled Peers [N = 8; Can't Tell = 2; Information = 4]). About the only variability that could be gleaned was that five of the 14 effect sizes came from planned studies, while 9 came from *post hoc* studies; and, the number of months the Ss had been in mainstreamed programs ranged from 6 to 18.

In short, the studies coded in the Mainstreaming category added little information, and there are too few data to warrant further analyses, with one exception. The mean D (.19) for this particular contact situation (Mainstreaming) is lower than the overall means (see Table 60) for Direct

Contact (.43) and Information Plus Contact (.52). It is also lower than the Direct Contact and Information Plus Contact mean D for Intermediate Ss (.81 and .59, respectively), but about the same as the Direct Contact mean D for Combination grade-age level Ss (.20). Whether any other variables in our data set are associated with this variability does merit exploration as a part of further analyses of our Contact effect sizes.

Studies with Effect Size Information Missing

As noted in Chapter 3, commonly in meta-analytic types of integrative reviews, studies are discarded if effect sizes cannot be computed directly or estimated from the inferential statistics that are reported. We coded some information for such studies (see Appendix B for the Coding Instrument) if, as a minimum, the statistical significance of the results could be determined. Our interest was to determine whether such studies differed in characteristics or appeared to produce results different from those studies for which effect size data could be obtained.

The number of results from reports which lacked effect size information, but provided information on statistical significance (called after this, "Information Missing studies"), was 182. Of these, 15 were Treatment A versus B comparisons, leaving 167 outcomes from treatment versus control, treatment versus placebo, and single-group, pre-posttest comparisons. To explore whether the Information Missing studies differed from those for which effect size information was available, our total population of studies, including those with A vs. B comparisons, was deemed pertinent. Of course, only frequencies are available for Information Missing studies. Some pertinent information is reported in Tables 66, 67, and 68.

Study Quality

One might suspect that "Information Missing" studies would be of lower quality. The General Internal Validity* information in Table 66 indicates that to be the case for our data set, although not dramatically so. A higher percentage of effect sizes came from studies with Medium validity (34% vs. 24%) and a lower percentage came from studies with Low validity (63% vs. 75%). There was a barely perceptible difference (3% vs. 1%) at the High validity level, with neither effect sizes or Missing Information results coming from many studies of that quality.

The data in Table 66 also indicate some differences in Types of Comparisons. Information Missing results were less likely than were effect sizes to come from treatment versus control (T vs. C) comparisons (58% vs. 71%) and somewhat more likely to come from single-group, pre-posttest comparisons (21% vs. 14%). Interestingly, while the Pre-post mean D was higher than the T vs. C or T vs. P means (.49, .36, and .29, respectively; see Table 48), for the No Information results (see Table 67), the Pre-post comparisons yielded a lower percentage of positive results then the T vs. C or T vs. P comparisons (36%, 49%, and 41%, respectively) and a higher percentage of negative results (36% for Pre-post, against 10% and 14% for T vs. C and T vs. P [with N only 3], respectively).

There were also two notable differences in Sample Selection. Sample selection procedures for Information Missing results were less likely to be identified (19% Can't Tell vs. 6%). And, Missing Information outcome samples were less likely to be volunteers (24% vs. 34%).

---

*Treatment Validity and Test Validity were not coded for Information Missing studies.

Table 66

Characteristics of "Information Missing" Outcome and Effect Size Studies

General Internal Validity

	Information Missing Outcomes		Effect Sizes	
Level	N	%	N	%
High	2	1	21	3
Medium	44	24	243	34
Low	136	75	441	63
Total	182	100	705	100

Type of Comparison

	N	%	N	%
Treatment Versus Control	106	58	499	71
Treatment Versus Placebo	22	12	49	7
A vs. B	15	8	60	8
Prepost	39	21	97	14
Total	182	99	705	100

Sample Selection

	N	%	N	%
Can't Tell	34	19	43	6
Random	16	9	62	9
Volunteer	44	24	237	34
Intact Group	74	41	327	46
Other	14	8	36	5
Total	182	99	705	100

Note. Because of rounding, percentages do not always sum to 100.

Table 67

Direction of "Missing Information" Outcomes by Comparison Types

	T vs. C		T vs. P		Pre-post		Total	
Direction of Results	N	%	N	%	N	%	N	%
Positive	52	49	9	41	14	36	75	45
Negative	11	10	3	14	14	36	28	17
Can't Tell	43	41	10	45	11	28	64	38
Total	106	100	22	100	39	100	167	100

Report Types

Not surprisingly, Missing Information results were less likely to come from dissertations (24%) than were effect sizes (61%), as can be seen in Table 68. However, even given the constraints on manuscript length for journal publications, it is surprising ("shocking" is really a more apt word) to see that 63 percent of the Missing Information results came from journal articles, while only 20 percent of the effect sizes did. Even journal publication restrictions do not account for that difference, as information such as means and standard deviations is easily reported in tables that take relatively little space. The occasional article discarded from our population because not even a *t*-ratio or F-ratio was reported along with level of statistical significance is even more reprehensible. Blame might be placed on the authors who submit such reports for publication or the editors who accept them. As "gatekeepers" for the profession, we believe that the editors (and their reviewers) bear special responsibility.

Contexts

Differences in Contexts (Table 69) are evident, too. Lower percentages of Missing Information outcome studies were conducted in an Elementary and Secondary Schooling context (26% vs. 36%) or College-University context (30% vs. 50%), but higher percentages were conducted in Inservice Education (15% vs. 8%) and Work (14% vs. 1%) contexts. This finding is particularly important in light of the lower mean *D*s and low numbers of *D*s for the adult groups in our main data analyses, especially when broken down by grade-age levels (see Table 60 and that part of Table 68).

Treatment Techniques and Results

It is also worth noting that there are differences in the percentages of treatment techniques represented in the two populations, as revealed in the

Table 68

Report Types for "Information Missing Studies and Effect Size Outcome

	Information Missing Outcomes		Effect Sizes	
Type	N	%	N	%
Journals	115	63	141	20
Dissertations	44	24	430	61
Theses	—	—	13	2
Convention Papers	4	2	13	2
Unpublished Reports	15	8	30	4
Combinations	4	2	78	11
Total	182	99[a]	705	100

[a]Because of rounding, percentages did not sum to 100.

Table 69

Study Contexts and Grade-Age Levels for "Information Missing" and Effect Size Outcomes

Context	Information Missing Outcome N	Information Missing Outcome %	Effect Sizes N	Effect Sizes %
Can't Tell	2	1	—	—
Elementary and Secondary Schooling	48	26	255	36
College-University	55	30	352	50
Adult Education	10	5	3	0.4
Inservice Education	28	15	57	8
Work	26	14	9	1
Community	—	—	8	1
Recreation	6	3	7	1
Other	7	4	14	2
Total	182	98	705	99.4

Grade-Age Level	Information Missing Outcome N	Information Missing Outcome %	Effect Sizes N	Effect Sizes %
Can't Tell	3	2	6	1
Preschool	—	—	3	0.4
Primary	9	5	30	4
Intermediate	8	4	116	16
Middle School	20	11	18	3
Junior High	4	2	17	2
Senior High	10	5	19	3
Combination	17	9	74	10
Undergraduate	52	29	283	40
Graduate	—	—	30	4
Postprofessional	21	11	53	7
Adult Not in School	30	21	50	7
Other	—	—	5	1
Total	182	99	705	98.4

Note. Because of rounding, percentages do not always sum to 100.

last two sets of columns of Table 70. Persuasive Messages and Systematic Desensitization are not included in the Missing Information outcomes; there are lower percentages of Other and Information Plus Vicarious Information results, but higher percentages of Information Plus Contact and Information results represented in the Missing Information outcomes.

Also of interest in Table 70 are the percentages of Missing Information outcomes that were positive for each treatment technique. Direct Contact has the highest percentage (71%), with Information Plus Contact second (44%) and Information third (37%). That is a reversal in order for Contact and Information Plus Contact from the rankings based on mean Ds.

Other Results

It is evident that the population of studies for which the reports lacked effect size computation information is in some important respects different from the population of studies for which such information was available. A potentially important area of difference, already touched on immediately above, is in the nature of results. To further address that topic, frequencies from treatment by control, treatment by placebo, and single-group, pre-posttest comparisons from the two populations were compared in terms of three aspects of the results from primary research reports often deemed important: the statistical significance of findings (important in terms of discerning any information-provision bias associated with the presence or lack of statistical significance), the direction of findings, and conclusions as to what the findings indicate about treatment effectiveness.

As can be seen in Table 71, the percentages of statistically significant and nonsignificant (at the .05 level) Missing Information results and effect

Table 70

Positive Results for Missing Information Outcomes, Totals of Missing Information Outcomes, and Numbers of Effect Sizes by Treatment Technique

Rankings on Mean Ds	Treatment	Positive Results		Total		Effect Sizes	
		N	% [a]	N	%	N	%
1	Persuasive Message	—	—	—	—	23	4
2	Information Plus Contact	24	44	54	32	100	15
3	Direct Contact	15	71	21	13	93	14
4	Vicarious Experience	8	57	14	8	58	9
5	Other	1	33	3	2	71	11
6	Systematic Desensitization	—	—	—	—	21	3
7	Information	26	37	71	42	203	31
8	Information Plus Vicarious	1	1	3	2	62	10
9	Persuasive Message, Contrast	--	--	—	—	11	2
	Positive Reinforcement	—	—	1	1	2	0.3
	Total	75	45 [b]	167	100	644	99.3 [c]

[a] Percentages of the Total Missing Information results that were positive for each technique.
[b] Not a summation of the column, but the percentage of Total Missing Information results that were positive.
[c] Because of rounding, the percentages did not add up to 100.

304

Table 71

Statistical Significance and Direction of Difference for "Information Missing" Outcomes and Effect Sizes

Statistical Significance

Statistical Significance at .05 Level	Information Missing Outcomes		Effect Sizes	
	N	%	N	%
Not significant	109	65	258	55
Significant	58	35	213	45
Total Available	167	100	471	100

Note. For 173 effect sizes (27%), information on statistical significance was not available.

Phi coefficient = .09.

Direction of Difference

Direction	Information Missing Outcomes		Effect Sizes	
	N	%	N	%
Negative	28	27	150	23
Positive	75	73	494	77
Total	103	100	644	100

Note. For 64 Missing Information outcomes (38%), all of which were statistically nonsignificant, the direction of difference could not be determined.

Phi coefficient = .03.

sizes* differed slightly. While 65 percent of the Information Missing outcomes were not significant at the .05 level, 55 percent of the effect size results were not; and, while 35 percent of the Information Missing outcomes were statistically significant, 45% of the effect size findings were. The Phi coefficient of .09 confirms the small trend toward missing information when statistical significance was not attained. The relationship is even smaller (Phi = .03) for direction of difference, despite a slight tendency for Information Missing Outcomes to reflect more negative (27% vs. 23%) and fewer positive (73% vs. 77%) results.

There are some interesting, although slight, differences in the conclusions about treatment effectiveness drawn by the authors for Missing Information outcomes and effect sizes (see Table 72). For example, 16 percent of the Missing Information authors drew no conclusions at all, as contrasted with 6 percent of the effect size authors. Moreover, only 33 percent of the Missing Information authors concluded that their treatment had no effect, but 10 percent indicated their results were equivocal, as contrasted with 40 and 6 percent for the effect size authors. Also, only 36 percent of the Information Missing authors were coded as drawing the conclusion that their treatment had a positive effect, while 44 percent of the effect size authors did so. That the differences are not dramatic overall is indicated by a Cramer's V of .16.

---

*Information on statistical significance was not available for 173 effect sizes. Most of these were effect sizes that were not tested for statistical significance. For example, a researcher who reported a statistically nonsignificant F-ratio for three treatment means and a control mean would typically not report the inferential results of pair-wise comparisons; we did, however, compute T vs. C effect sizes.

Table 72

Authors' Conclusions About Treatment Effectiveness for "Information Missing" Outcomes and Effect Sizes

Conclusion	Information Missing Outcomes N	Information Missing Outcomes %	Effect Sizes N	Effect Sizes %
No conclusion	26	16	42	6
No effect	56	33	258	40
Equivocal results	17	10	40	6
Positive effect	61	36	284	44
Negative effect	7	4	20	3
Total	167	99	644	99

Note. Because of rounding, percentages do not sum to 100.

Cramer's V = .16.

Differences in the results of studies from reports with and without information for computing effect sizes appear to be small. But it must be remembered that there are differences in the studies in the two populations, including the treatments that are represented in the data sets.

Variance Ratios

The focus of data analysis in most educational and psychological experimental research has been on the central tendencies (usually as indicated by the means) of the groups under study. Although the variability in scores is a key element in the statistical tests used to examine mean differences, little attention is given to differences in dispersion, per se. Reviewers of research, too, have largely assumed that the important question is whether there were mean differences at treatment end, ignoring questions of whether variability increased or decreased, or which result would be desirable. Moreover, whether as a result of the tendency of researchers not to ask questions about the effects of treatments on variability, or as a cause of that tendency, the issues involved in the analysis of changes in variability have not been addressed adequately.

Analyzing Variances

The variance and standard deviation are conventionally used as measures of variability. Differences between or among independent variances are occasionally tested for statistical significance (frequently to test for homogeneity prior to testing mean differences for statistical significance), and a standard error is available for testing the difference between two correlated standard deviations. But nothing equivalent to the repeated measures analysis of variance or analysis of covariance for means is available for standard deviations or variances in the educational and psychological applied research literature.

The issue of an effect size for standard deviations or variances has also not been treated in the literature, as effect sizes for means have been. As we considered possible effect sizes, it seemed clear that differences between pretest and posttest variances for a treatment group and control group could not be treated as indices of change to be subtracted and standardized as mean gains are to obtain Ds. Also, the comparison of pretest and posttest variance ratios by subtracting one from the other or by forming a new ratio (e.g., the pre-treatment variance ratio divided by the posttest variance ratio) made little sense.

It was decided, for exploratory purposes, to disregard pretreatment differences in variability and obtain an estimate of difference in posttreatment variability. That is, the posttest variance for the treatment group was divided by the posttest variance for the control or placebo group, or the pretest variance in a single-group comparison, to obtain a variance ratio. A ratio greater than one indicates greater posttreatment variability for the treatment group; a ratio less than one indicates the reverse--less posttest variability for the treatment group. Of course, a ratio of 1 indicates equal variability.

Overall Results

For the 644 effect sizes used in the main analyses reported in Chapter 5, the posttest standard deviations or variances necessary to compute variance ratios were available for 453, or 70 percent. The mean variance ratio was 1.13. That is, overall, experimental groups were, at posttest, slightly more variable than their untreated comparisons, although the difference from 1 (the ratio for equal variances) is less than the criterion of .18 for triviality of a difference between variance ratios, as defined in Chapter 3. The standard deviation of .91 represents considerable dispersion in variance ratios.

Evaluating Variance Ratios

The evaluation of changes in mean scores on measures of attitudes toward persons with disabilities is usually straightforward: The question is whether the treatment group has a higher posttest mean (or lower, if that indicates more positive attitudes) than is present in the nontreatment comparison data, especially if any change in the treatment group's mean from pretreatment to posttreatment assessment has been in a positive direction.

With variances, evaluation of changes is not so simple. Do we want to increase or decrease variability in attitudes? That depends, and in many situations an answer is difficult to come by. For example, if the treatment group's mean attitude score is initially low and there is little movement in a positive direction, increased variability (i.e., a higher variance ratio) might indicate more positive scores by those who already had high scores--which is good, unless accompanied by less positive scores by those already low. On the other hand, a decrease in variability (i.e., a smaller variance ratio) might reflect upward movement by those at the less favorable end of the attitude scale--which would be good, unless the reduced variance also indicated that those with more positive pretreatment attitudes developed less favorable attitudes during treatment.

Probably the most unambiguous situation would be one in which the mean attitude score for the treatment group became more positive and variability was reduced (the variance ratio became less). That is, in such a valued situation, the central tendency of the treatment group's attitude scores, as compared to the comparison data, would become more positive, but the dispersion of treatment group scores would be reduced. The reduced dispersion would, hopefully, indicate that Ss who had lower pretreatment scores had moved closer to the mean, which had shifted "upward", while those Ss with initially high scores had not regressed more than expected, although a ceiling effect might have precluded higher attitude scores. In terms of our data analyses, one would look for positive Ds accompanied by variance ratios of less than 1.

D-Variance Ratio Relationships

What were the relations between Ds and variance ratios for our 644 effect sizes from treatment versus control, treatment versus placebo, and single-group, pre-posttest comparisons? As reported earlier, the mean D was .37, an indicator of positive change (remembering that underlying the mean is considerable hetereogeneity in individual Ds [SD = .61], as well as negative treatment group changes and negative Ds). However, the mean variance ratio, as noted above, was 1.13, reflecting a slightly greater variance for treatment groups, on the average (again, with considerable heterogeneity in individual variance ratios [SD = .91]).

How about the results for different treatment techniques? As Table 73 shows, the mean variance ratios for three techniques* are basically 1, indicating equal variances. Two of these are slightly below 1 (Information Plus Contact, .91; Information Plus Vicarious Experience (.96), and one is

*Positive Reinforcement is ignored, because the N is only 2, along with Persuasive Message, Contrast, because 10 of 11 effect sizes are from the same study.

Table 73

Ds and Variance Ratios for Attitude Modification Techniques

		Effect Sizes (Ds)				Variance Ratios		
Rank	Technique	N	Mean	SD	Rank	N	Mean	SD
1	Persuasive Message	23	.67	.56	2	18	1.36	.84
2	Information Plus Contact	100	.51	.66	7	70	.91	.57
3	Direct Contact	93	.43	.73	8	64	.86	.40
4	Vicarious Experience	58	.40	.76	3	44	1.37	1.13
5	Other	71	.39	.64	1	65	1.43	1.30
6	Systematic Desensitization	21	.32	.44	5	15	1.04	.41
7	Information	203	.29	.51	4	133	1.13	.93
8	Information Plus Vicarious	62	.20	.36	6	31	.96	.52
9	Persuasive Message, Contrast	11	.13[a]	.33		11	.94[a]	.37[a]
	Positive Reinforcement	2	(1.74)[b]	(.01)[b]		2	(4.13)[b]	(.11)[b]
	Total	644	.37	.61		453	1.13	.91

Note. The Eta^2 for Ds = .05; for variance ratios, Eta^2 = .10.

[a]Ten of 11 Ds came from one study.

[b]Too few effect sizes (less than 10) to be interpretable.

slightly above (Systematic Desensitization, 1.04). Two mean variance ratios are slightly greater, but still small, departures from the "no-difference" value of 1. One of these is less than 1 (Direct Contact, .86) and one is greater than 1 (Information, 1.13). Three variance ratios are sufficiently larger than 1 to indicate a difference of small to moderate magnitude: Persuasive Messages (1.36), Vicarious Experiences (1.34), and Other (1.43).

If anything, then, a slight tendency toward greater posttest variability in treatment groups accompanies the positive mean Ds. In particular, a pattern of higher mean Ds associated with lower variance ratios is not evident in the table. In fact, instead of a negative relationship, there is a low positive correlation ($r = .29$; $r^2 = .08$) between the first eight mean Ds and variance ratios.

To further investigate D-variance ratio relationships, correlation coefficients were computed for individual effect sizes. The coefficient for the 453 individual Ds and the associated variance ratios, as presented in Table 74, is -.03, considerably different from the coefficient of .29 for the mean Ds and variance ratios. Moreover, the correlation coefficients for individual treatments reflect a striking amount of diversity in magnitude and direction of relationship. Three of the coefficients are zero or near zero (Persuasive Messages, $r = -.08$; Direct Contact, $r = -.03$; Information, $r = .00$), indicating relationships between Ds and variance ratios that are basically random*. Two coefficients are positive, with one small (Vicarious Experience, $r = .24$) and the other moderate (Systematic Desensitization, r

*Scatter-diagrams were inspected for curvilinearity and to determine if outliers had undue effects on the coefficients. In no case was there obvious curvilinearity, and outliers fit the relationship expressed by the r.

Table 74

Correlations Between Ds and Variance Ratios

	Individual		Median		Median, Outliers Excluded	
Technique	N	r	N	r	N	R
Persuasive Message	18	-.08	5	(-.88)[a]	—	—
Information Plus Contact	70	-.40	28	-.40	27	-.39
Direct Contact	64	-.03	22	-.40	—	-.40[b]
Vicarious Experience	44	.24	13	.38	12	.11
Other	65	-.38	19	.00	18	-.34
Systematic Desensitization	15	.50	1	—	—	—
Information	133	.00	35	.30	34	-.23
Information Plus Vicarious	31	.33	10	.37	9	(-.04)[a]
Persuasive Message, Contrast	11	-.05	2	—	—	—
Positive Reinforcement	2	—	1	—	—	—
Total	453	-.03	136	.17	135[c]	.08

[a]Too few data points to be interpretable.
[b]No outliers.
[c]This is the N for a coefficient with one outlier deleted, not the total for the separate outlier-excluded analyses.

314

= .50), indicating that as treatment outcomes increased, so did variability. And, two coefficients are negative and moderate (Information Plus Contact, r = -.40; Other, r = -.38), indicating that as outcomes became more positive relative to comparison groups, variability decreased relative to the same comparison data--our postulated valued relationship.

Following up on previously reported exploratory analyses, it was decided to correlate median Ds with median variance ratios to determine what, if any, changes would occur when analyzing by study rather than individual effect size (see Table 74). There was little or no change for three treatments (Information Plus Contact, -.40 vs. -.40; Vicarious Experience, .24 vs. .38; and, Information Plus Vicarious Experience, .33 vs. .37). For two treatments, the coefficients shifted in a positive direction (Other, -.38 vs. .00; and Information, .00 vs. .30). And, for one treatment, the shift was toward greater magnitude for a negative coefficient (Direct Contact, -.03 vs. -.40).

Inspection of the scatter-diagrams indicated, however, that, contrary to what was the case with the analyses for individual effect sizes, one or two outliers ran counter to the trend in some of the bivariate distributions (none of which appeared to be curvilinear). To check on the possible influences, correlations were rerun with obviously contradictory outliers excluded. In no case was more than one outlier deleted; with Direct Contact, none was eliminated. The coefficient for Information Plus Contact remained amazingly constant across all three analyses (-.40, -.40, -.39), while those for Vicarious Experience (.38 vs. .11), Other (.00 vs. -.34), and Information (.30 vs. -.23) changed substantially, with the Other coefficient nearly identical to the coefficient for the individual effect size analysis (-.34 vs. -.38).

All in all, the picture is not particularly clear, except for Information Plus Contact. All analyses indicated the postulated desirable tendency for higher Ds for that treatment to be associated with lower variances. Direct Contact and Other (combinations of techniques) showed strong tendencies in that direction, with Information less strongly so. At the same time, the coefficients for Vicarious Experience were consistently positive, even if small (r = .11), indicating a tendency for larger variances to be associated with higher median Ds.

The differing directions and the range of magnitude of the correlations between Ds and variance ratios present provocative questions about the relative effects of different treatments or, as has been noted in Chapter 5, about the differences in samples (e.g., grade-age levels) with which different treatments have been mainly investigated. Without data in which pretreatment differences are held constant, further analysis did not seem worthwhile. But, at the very least, the results suggest that the effects of treatments on variability in attitudes toward disabled persons are deserving of greater attention by primary researchers and reviewers.

Summary

In this chapter, we have presented a potpourri of exploratory and supplementary analyses. We reported finding little suggestion that the results of analyses with the 644 Ds from treatment versus control, treatment versus placebo, and single-group, pre-posttest comparisons would have been different had we first eliminated outliers (Ds more than 2 standard deviations above or below the mean), or if we had used a median D for each study, rather than analyzing Ds for individual findings. We did report some evidence that studies for which the reports contained insufficient information for computing Ds seemed to be in other ways as well a different population than the studies for which that information was available. We found little information to share from the Mainstreaming studies and the Treatment A versus Treatment B comparisons that were coded. Finally, our brief treatment of the ratios of posttest treatment and comparison variances suggested that variability is a topic that is more deserving of attention by statisticians, primary researchers, and reviewers. In any event, the main foundation for any conclusions to be drawn from this quantitative integrative review remains the analyses of Ds reported in Chapter 5.

CHAPTER 7

WHAT HAS BEEN LEARNED?

As noted in Chapters 1 and 2, prior reviews have not been based on comprehensive collections of research reports or on the extensive, systematic collection and analysis of quantitative data on study outcomes and study characteristics. An assumption underlying this review was that the inability of prior reviewers to draw firm conclusions about the effectiveness of attitude modification techniques was likely due, at least in part, to the small samples of prior studies that were reviewed and the lack of systematic data collection and analysis. A meta-analytic type of integrative review of the research on modifying attitudes toward persons with disabilities was proposed and initiated with the hope of bringing order to the literature where other reviews had not done so. As the reader of Chapters 5 and 6 knows, that hope turned out to be in vain.

Treatment Efficacy

Even with a population of studies based on an exhaustive search of the literature and with quantitative integrative review techniques, clear-cut indications were not found of the overall efficacy of techniques for modifying attitudes toward disabled persons or of reliable differences in efficacy between techniques. The mean $\underline{D}$ for our 644 treatment versus control, treatment versus placebo, and single-group, pre-posttest comparisons was .37 (for a positively skewed distribution). This is a moderate, not particularly large, effect size. It, as well as the standard deviation of .61, reflects the instances in our data set of negative changes by treatment groups (12%) and negative effect sizes (23%), which indicate that at treatment end, the comparison group had the higher mean.

We were able to rank order the mean Ds for treatment techniques from .67 for Persuasive Messages to .20 for Information Plus Vicarious Experiences. There were, however, negative effect sizes for each treatment. Moreover, the variances indicated considerable overlap between the distributions of Ds for the various treatment techniques.

Concomitant Variables

A search for concomitant variables which might explain or help to make sense out of those results was not particularly fruitful. Surprisingly, "quality of study" indicators were not related to outcomes, a matter to which we will return later in this chapter.

Treatment variation. There was a great deal of variation in treatments categorized under similar labels, such as Information and Direct Contact. For the most part, the proportion of variance in outcomes associated with these variations was low (.07 or less), although type of experience was associated with 20% of the variance in Ds for Vicarious Experience studies, and type of message presentation was associated with 28% of the variance in Persuasive Message Ds. There were some apparent differential effects. However, nesting of treatments within the types of disabilities toward which attitude change efforts were directed and cells that were empty, or nearly so, precluded conclusions about interactions.

Other study characteristics. Studies also differed in a variety of other ways, including the length of treatment and time of posttest, the type of dependent measures, the contexts and settings within which they were carried out, and sample size. These variations explained very little of the variance in Ds, with no r^2 or Eta^2 greater than .05. In most cases, the majority of effect sizes fell into one or two characteristic categories. For most of the variations, any relations with Ds were consistent across treatments. One exception was length of treatment. The overall r for length of treatment and Ds was .02, but there were some differences in coefficients within treatment categories--with low negative coefficients for Information, Information Plus Contact, and Systematic Desensitization, and a moderate positive r for Systematic Desensitization. Another exception was context. The predominant contexts were Elementary-Secondary Schooling and College-University, with some nesting of treatments within contexts (e.g., no Persuasive Messages or Systematic Desensitization effect sizes from the Elementary-Secondary context) and some different results (e.g., a higher Direct Contact D in the Elementary-Secondary context, with a reversal for Vicarious Experiences). Again, nesting and empty or low N cells make difficult any conclusions about the association of treatment outcomes with other study characteristics.

Sample characteristics. Variations in sample characteristics also accounted for little of the variance in Ds, with no Eta^2 or r^2 larger than .04 for method of sample selection, grade-age level, or gender. (The relations of prior contact and personality variables to outcomes could not be analyzed because they were basically ignored in the primary research reports.) There were some differential effects for samples selected by different methods, especially volunteers; but they were confounded with context (volunteers were more likely to come from college-university studies). There also appeared to be treatment effect size differences by grade-age levels, but with nesting and small Ns or empty cells, that could not be discerned with certainty. Gender, too, seemed to be related to different results across treatments, but the relationships were moderate and inconsistent.

Summary. As a consequence of the unevenly distributed variations, with many cells empty or with low Ns, and the nesting of treatments, the analysis of potential concomitant variables was not particularly productive, except for indicating areas to be addressed in future research. Rather than drawing conclusions about the conditions under which different attitude modification techniques had been more or less successful, the major conclusion had to be that there had been a great deal of variety in the conditions under which the effectiveness of the various attitude modification techniques was investigated, that the variations have not been systematically controlled, and that, for that reason, they confounded efforts to draw conclusions about treatment effectiveness.

Summation

All possible data analyses could not be conducted within the time span of the funded project for which this report has been prepared, and further analyses will be conducted for other reports to groups of professionals.* However, at this time, the status of the research field might best be summarized with the flavor of the quote from Towner (1984) which we used in Chapter 1 to indicate that another review of the literature was warranted:

*Papers are now scheduled for presentation at annual meetings of the Council on Exceptional Children in Chicago on April 22, 1987, and the American Educational Research Association in Washington, D. C. on April 24, 1987, and at the Third Annual Canadian Congress of Rehabilitation sponsored by the Canadian Rehabilitation Council for the Disabled in Quebec on June 2, 1987. Papers dealing with specific factors in attitude change, such as the target attitudes, the grade-age level of subjects, and individual modification techniques will be developed for presentation at the meetings of other professional organizations. For example, a presentation on implications for social studies curricula has been proposed for the annual meeting of the National Council for the Social Studies in Dallas in November, 1987. Several manuscripts for journal articles, including one in which the prior reviews are analyzed, are under preparation, as well.

> The applications [of similar techniques] yielded discouraging and contradictory findings. Both positive and negative attitudinal changes, in addition to numerous reports of [statistically] nonsignificant changes, resulted from interactions [of nondisabled persons] with disabled persons as well as from the provision of educational and general information. (p. 249)

The results of this review are likely to be disappointing not only for persons seeking guidelines for attitude modification programs, but for those interested in the implications of this body of research for the validity of attitude modification theory. Very few of the investigators based their treatments explicitly upon any theory. Consequently, little worthwhile information was obtained.

Quality of Research

Earlier references to the lack of systematic variations in treatment and other study characteristics have more than incidental importance. If the comprehensive quantitative review of literature reported in prior chapters did not lead to firm conclusions about the efficacy of various attitude modification techniques for changing attitudes toward persons with disabilities, it did yield some more definite conclusions in regard to the quality of the research in this area.

As noted in Chapter 5 and mentioned above, research quality indicators were not related to study outcomes in our data set. That result may have been due in large part to the few studies which were coded as being of high quality. One coding difficulty lay with reporting--that is, the frequent lack of sufficient information in reports to make methodological categorizations. Whether due to a predilection on the part of authors not to give adequate details, or to the pressures, real or imagined, to make research reports brief, particularly for journals, important information in regard to procedures and threats to internal validity was omitted.

Beyond that coding constraint, however, many methodological weaknesses were obvious in the studies that made up our population. Included were, for example, failures to verify implementation of the independent variable, neglecting to report reliability coefficients, much less to compute them for the scores used in the investigation, the tendency to use only Likert-type questionnaires to assess attitudes, failures to incorporate blinded test administration in the procedures, and reliance upon statistical significance as an indicator of the importance of outcomes.

Perhaps most important of all, replications of studies were, for all intents and purposes, missing from the population of reports. Moreover, the welter of study characteristics and outcomes belie the hope of some persons that meta-analytic reviews of the literature will be an adequate substitute for the careful planning of replications as an element of programmatic research.

This "finding"--that is, that the bulk of the research in the field has not been methodologically strong--may be the most important one to come out of this integrative review. Many implications for sounder research and research reporting are embedded in our Results chapters. Probably none is more important than the obvious need for programmatic research, for planned series of replications rather than helter-skelter studies depending upon the availability of samples and situations and the researchers' (particularly graduate students and their advisors) particular interests at that time.

A major lesson from this integrative review, then, is that meta-analysis is not likely to be an adequate substitute for programmatic, replication-oriented research in this field. With the multitude of potentially relevant variables in the area of attitude modification and the nesting of treatments within levels of those variables, it was not possible to isolate through post hoc analysis the possible concomitant and contaminating factors that might account for the equivocal, even contradictory, results from the body of primary studies.

At the same time, another dilemma has been posed. It is that even with careful attention to design, it is difficult to conduct valid studies in an applied area such as modifying attitudes toward disabled persons. Unfortunately, it only takes one serious threat to internal validity to invalidate the results from an otherwise strong study. This frailty of research designs in applied human fields makes the call for replication all the more urgent. It will, however, be especially important that those who carry out replications attend carefully to methodology so that the same design weaknesses will not be pervasive, thus producing a systematic bias in findings across studies.

Lessons About Integrative Reviews

Jackson (1978, 1980) has enunciated well the position that the procedural steps involved in conducting an integrative review are analogous to those for a piece of primary research. That position served as the conceptual frame for the review of prior reviews in Chapter 2 and for the approach to our integrative review.

As researchers with considerable prior experience in conducting primary research studies, however, we have discerned one important difference, at least from primary research studies that involve the typical collection of pre- and posttest data using a group-administered test, without the use of observational or other intensive data collection techniques to assess either treatment implementation or outcomes. That is, doing a proper meta-analysis is very labor intensive and, with a large number of studies, very time consuming. The development of a coding instrument (categories, definitions for categories, and conventions for using the categories) is as demanding as the development of any complex instrument which multiple observers must use reliably. (Fortunately, much time and effort were saved for us by the availability of the expertise and prior work of Karl White and Glendon Casto of Utah State University's Early Intervention Research Institute.) Coding itself, as we have detailed in Chapter 3, is a time-consuming process—and one, our experience confirmed, in which senior staff should be involved.

As we noted at the beginning of Chapter 5, however, a major difference between a properly conducted integrative review and a typical piece of primary research is the complexity of data analysis, if one has gathered data with a coding instrument which validly reflects the complexity of research contexts and situations. Despite the advice of experienced meta-analysts, we underestimated the time that would be needed to fully analyze our data.

Of course, an integrative review with a population of studies as extensive and complicated as ours should not be thought of as a task to be completed within a limited period of time, defined, for example, by a funding period. As journal articles and further reports to professional groups are prepared, analytic suggestions made in Chapters 5 and 6 will be followed up and other analyses conducted. Of particular interest will be more fine-grained analyses of attitude modification techniques by such variables as grade-age level and disability attitude targets, in an attempt to discern any regularities which we did not detect through the analyses which underlay this report.

What is the Reality?

We have emphasized the need for not only better designed research studies but for a more productive research strategy, i.e., replication, in the investigation of modifying attitudes toward persons with disabilities. We have also alluded to the possibility that the internal validity of attitude modification studies in this area may be intrinsically frail, because of the difficulties involved in studying such phenomena in applied settings. If that is the case, even with careful replication the accumulation of findings that indicate clearly what attitude modification techniques are most effective, or which are most effective with which types of persons for changing attitudes toward what types of disabilities, may turn out to be a difficult, if not impossible, goal to attain. That state of affairs may explain the results of this review and of the more limited reviews that preceded it.

Another possibility has to be considered as well, however. That is, that the state of research identified in earlier limited reviews of the literature and confirmed by our comprehensive, quantitative review is not a function of either poor design or inherent methodological deficiencies, but a reflection of reality. For example, Cronbach (1975) has argued that complex interactions among variables is the natural state of affairs with psychological phenomena. He has suggested that the complications are so great that when researchers begin to attend to all of the potential interactions, they enter a metaphorical "hall of mirrors". Also arguing for a nonsimplistic view of human behavior, Perrow (1981) has contended that social-psychological phenomena may be much less amenable to systemization, and much more unpredictable, than most of those engaged in "social science" research are willing to admit.

Physicists David Crutchfield and his associates (1986) have cast the lack of predictability of some phenomena in scientific terms. Based on their analyses of physical phenomena, they argue that "chaos"--that is, randomness generated according to orderly principles--may be a fundamental restriction on our ability to develop cause-and-effect conclusions in some areas of study. That is, contrary to the frequent assumption that the prediction of such phenomena as the weather, the flow of water in a mountain stream, and human behavior is possible, if only sufficient amounts of information can be gathered and processed, they contend that randomness, or chaos, may be fundamental.

Chaos comes about according to understandable rules that are not in themselves based on chance, and the chaotic behavior is, itself, lawful. However, with chaotic events--such as weather changes, turbulence of a mountain stream, and perhaps some areas of human behavior--"small uncertainties are amplified" and behavior becomes unpredictable. A speck of dust observed under a microscope moves continuously and erratically because it is bombarded by surrounding water molecules caught in thermal motion. As Crutchfield et al. put it, "the web of causal influences among the subunits can become so tangled that the resulting pattern of behavior becomes quite random" (p. 46).

Chaotic behavior may stem from initially simple interactions between a few components. The individual elements are simple and rule-consistent; but complex interactions make prediction impossible, with exponential growth in the inability to predict reliably as the chaotic behavior continues. Human creativity, Crutchfield and his co-authors note, may reflect and contribute to chaos in human behavior, as small fluctuations in thinking are amplified and molded into "macroscopic coherent mental states that are experienced as thoughts" (p. 57).

Are attitudes and attitude change chaotic phenomena? That is, are attitudes pertubations that result from interactions among relatively few and simple elements whose effects are magnified by the cognitive power of the mind, so that the attitude-related behavior has, for the purposes of scientific systemization, large random elements to it? Chaos would explain the consistent conclusions in reviews of the literature of equivocal findings. Improvements in research are necessary, as sketched out above, before it will be possible to know whether the observed state of affairs in research on modifying attitudes toward disabled persons is an artifact of methodology or a natural state of affairs.

In any event, the prognostication need not be pessimistic. On the one hand, with improved research methods, researchers may be able to determine how to have the desired influence on attitudes toward persons with disabilities. On the other hand, should it emerge that we are dealing with chaotic phenomena, research efforts could be directed toward understanding how to channel behavior based on negative attitudes toward disabled persons so that such attitudes will not have destructive effects, much as—to use a turbulent water metaphor—researchers continue to study not only the factors that affect water movement but how to channel water to avoid its destructive forces, even though they are unable to predict the motions of individual molecules or droplets of turbulent water.

Channelling of water is done to reduce the effects of turbulence. Similarly, we have seen that the channelling of attitude-related behavior—for example, by Public Law 94-142, judicial decisions, and the policies and actions of public and private agencies--can have (as it has in race relations) positive long-term impacts on attitudes while delimiting the potentially restrictive and dehumanizing effects of discriminatory behavior. Research into the modification of attitudes should be more methodologically and strategically sound in order to improve the likelihood that it will lead to better understanding of the phenomenon; yet, how to limit the debilitating effects of negative attitudes may be an equally, if not more, important policy concern and productive line of inquiry.

References

Abroms, Kippy, & Kodera, Thomas. (1979). Acceptance hierarchy of handicaps: Validation of Kirk's statement, "Special education often begins where medicine stops." Journal of Learning Disabilities, 12, 15-20.

Alexander, C., & Strain, P. (1978). A review of educators' attitudes toward handicapped children and the concept of mainstreaming. Psychology in the Schools, 15, 390-396.

Allport, Gordon W. (1954). The nature of prejudice. Garden City, NY: Doubleday.

Aloia, G., Beaver, R., & Pettus, W. (1978). Increasing initial interactions among integrated EMR students and their nonretarded peers in a game-playing situation. American Journal of Mental Deficiency, 82, 573-579.

Altrocchi, J., & Eisdorfer, C. (1961). Changes in attitudes toward mental illness. Mental Hygiene, 45, 563-570.

Amir, Yehuda. (1969). Contact hypothesis in ethnic relations. Psychological Bulletin, 71(5), 319-342.

Anthony, W. (1969). The effects of contact on an individual's attitude toward disabled persons. Rehabilitation Counseling Bulletin, 12, 168-170.

Anthony, W. (1972). Societal rehabilitation: Changing society's attitudes toward the physically and mentally disabled. Rehabilitation Psychology, 19, 117-126.

Antonak, Richard F. (1980). A hierarchy of attitudes toward exceptionality. The Journal of Special Education, 14, 231-241.

Antonak, Richard F. (1986, June). Methods to measure attitudes toward people who are disabled. Paper presented at the Conference on Attitudes Toward Persons with Disabilities, Hofstra University, Hofstra, New York.

Ball-Rokeach, Sandra J., Rokeach, Milton, & Grube, Joel W. (1984). The great American values test: Influencing behavior and belief through television. New York: The Free Press.

Bangert-Drowns, Robert L. (1986). Review of developments in meta-analytic method. Psychological Bulletin, 99(3), 388-399.

Begab, M. (1969). The effect of differences in curricula and experiences on social work students' attitudes and knowledge about mental retardation. Dissertation Abstracts, 29, 4111-4112.

Bernotavicz, Freda. (1979). Changing attitudes towards the disabled through visual presentations: What the research says. (Report to the National Institute of Handicapped Research, Grant No. 12-P-59142). The Disability, Attitudes and Media Project, Human Services Development Institute, Center for Research and Advanced Study, University of Southern Maine.

Bogdan, Robert, & Biklen, Douglas. (1977). Handicapism. Social Policy, 7(5), 14-19.

Borg, W., & Gall, M. (1983). Educational Research (4th ed.). New York: Longman.

Bowe, Frank. (1978). Handicapping America: Barriers to disabled persons. New York: Harper & Row.

Bowe, Frank. (1980). Rehabilitating America: Toward independence for disabled and elderly people. New York: Harper & Row.

Bracey, Gerald W. (1986). Research: Tips for readers of research. Phi Delta Kappan, 67, pp. 395-6.

Bracht, Glenn H., & Glass, Gene V. (1968). The external validity of experiments. American Educational Research Journal, 5(4), 437-474.

Bradfield, R. H., Brown, J., Kaplan, P., Rickert, E., & Stannard, R. (1973). The special child in the regular classroom. Exceptional Children, 39, 384-390.

Brennan, J., & Margolin, R. (1954). Utilizing community resources for rehabilitation of psychiatric patients. Personnel and Guidance Journal, 32, 330-335.

Brooks, B., & Bransford, L. (1971). Modification of teachers' attitudes toward exceptional children. Exceptional Children, 38, 259-260.

Bryant, Fred B., & Wortman, Paul M. (1985). Methodological issues in the meta-analysis of quasi-experiments. Evaluation Studies Review Annual, 10, 629-648.

Bullock, R. J., & Svyantek, Daniel J. (1985). Analyzing meta-analysis: Potential problems, an unsuccessful replication, and evaluation criteria. Journal of Applied Psychology, 70(1), 108-115.

Campbell, Donald T., & Stanley, Julian C. (1963). Experimental and quasi-experimental designs for research. Chicago: Rand McNally.

Carlberg, Conrad G., Johnson, David W., Johnson, Roger, Maruyama, Geoffrey, Kavale, Kenneth, Kulik, Chen-Lin C., Kulik, James A., Lysakowski, Richard S., Pflaum, Susanna W., & Walberg, Herbert J. (1984). Meta-analysis in education: A reply to Slavin. Educational Researcher, 13(8), 16-23.

Carver, Ronald P. (1978). The case against statistical significance testing. Harvard Educational Review, 48(3), 378-399.

Chubon, R. (1982). An analysis of research dealing with the attitudes of professionals toward disability. Journal of Rehabilitation, 48(1), 25-29.

Cleland, C., & Chambers, W. (1959). Experimental modification of attitudes as a function of an instructional tour. American Journal of Mental Deficiency, 64, 124-130.

Clore, G., & Jeffrey, K. (1972). Emotional role-playing, attitude change, and attraction toward a disabled person. Journal of Personality and Social Psychology, 23, 105-111.

Cohen, Jacob. (1977). Statistical power analysis for the behavioral sciences. New York: Academic Press.

Cole, F. (1971). Contact as a determinant of sighted persons' attitudes toward the blind. Dissertation Abstracts, 31, 6892-6893.

Cook, Thomas D., & Campbell, Donald T. (1979). Quasi-experimentation: Design & analysis issues for field settings. Chicago: Rand McNally.

Cook, Thomas D., & Leviton, Laura C. (1980). Reviewing the literature: A comparison of traditional methods with meta-analysis. Journal of Personality, 48, 449-472.

Cooper, Harris M. (1982). Scientific guidelines for conducting integrative research reviews. Review of Educational Research, 52, 291-302.

Cooper, Harris M. (1984). The integrative research review: A systematic approach. Beverly Hills: Sage.

Cooper, Harris M., & Rosenthal, Robert. (1980). Statistical versus traditional procedures for summarizing research findings. Psychological Bulletin, 87(3), 442-449.

Cowen, E., Underberg, R., & Verrillo, R. (1958). The development and testing of an attitude to blindness scale. The Journal of Social Psychology, 48, 297-30[illegible]

Cronbach, Lee J. (1975). Beyond the two disciplines of scientific psychology. American Psychologist, 30, 116-127.

Crutchfield, James P., Farmer, J. Doyne, Packard, Norman H., & Shaw, Robert S. (1986). Chaos. Scientific American, 255(6), 46-57.

Dahl, H., Horsman, K., & Arkell, R. (1978). Simulation of exceptionalities for elementary school students. Psychological Reports, 42, 573-574.

Donaldson, Joy. (1974). Effects of live, video, and audio presentations by a panel of physically disabled individuals on attitudes toward disabled persons. Unpublished doctoral dissertation, University of Kentucky.

Donaldson, Joy. (1976). Channel variations and effects on attitudes toward physically disabled individuals. Audio-Visual Communication Review, 24, 135-144.

Donaldson, Joy. (1980). Changing attitudes toward handicapped persons: A review and analysis of research. Exceptional Children, 46, 504-513.

Donaldson, J., & Martinson, M. (1977). Modifying attitudes toward physically disabled persons. Exceptional Children, 44, 337-341.

Dye, Celeste A. (1978). Effects of persuasion and autotelic inquiry methods on attitude change. Perceptual and Motor Skills, 47, 943-949.

Dye, Celeste A. (1980). Autotelic inquiry: A learning approach to attitude change. Educational Gerontology, 5, 239-248.

English R. William. (1966). Assessment, modification, and stability of attitudes toward blind persons. Unpublished master's thesis, Southern Illinois University.

English, R. William. (1971). Correlates of stigma towards physically disabled persons. Rehabilitation Research and Practice Review, 2(4), 1-17.

Esposito, Beverly G., & Peach, Walter J. (1983). Changing attitudes of preschool children toward handicapped persons. Exceptional Children, 49(4), 361-363.

Esposito, Beverly G., & Reed, Thomas M., II. (1986). The effects of contact with handicapped persons on young children's attitudes. Exceptional Children, 53(3), 224-229.

Evans, J. (1976). Changing attitudes toward disabled persons: An experimental study. Rehabilitation Counseling Bulletin. 19, 572-579.

Eysenck, H. J. (1978). An exercise in mega-silliness. American Psychologist, 33, 517.

Feldman, K. A. (1971). Using the work of others: Some observations on reviewing, integrating, and consolidating findings. Sociology of Education, 44, 86-102.

Fenton, T. (1975). The effects of inservice training on elementary classroom teachers' attitudes toward and knowledge about handicapped children. Dissertation Abstracts International, 35, 5966A.

Forader, A. (1970). Modifying social attitudes toward the physically disabled using three different modes of instruction. Dissertation Abstracts, 30, 4360B.

Frith, Greg H., & Mitchell, Janet W. (1981). The attitudes of nonhandicapped students toward the mildly retarded: A consideration in placement decisions. Education and Training of the Mentally Retarded, 16, 79-83.

Gage, N. L. (1982). The future of educational research. Educational Researcher, 11(8), 11-12.

Gallo, P. S. (1978). Meta-analysis--A mixed meta-phor. American Psychologist, 33, 515-517.

Gay, L. (1976). Educational research: Competencies for analysis and application. Columbus, OH: Charles E. Merrill.

Glass, Gene V. (1976). Primary, secondary, and meta-analysis of research. Educational Researcher, 5(10), 3-8.

Glass, Gene V. (1977). Integrating findings: The meta-analysis of research. Review of Research in Education, 5, 351-379.

Glass, Gene V. (1978). Reply to Mansfield and Bussey. Educational Researcher, 7, 3.

Glass, Gene V. (1980, December). On criticism of our class size/student achievement research: No points conceded. Phi Delta Kappan, 242-244.

Glass, Gene V, McGaw, B., & Smith, M. (1981). Meta-analysis in social research. Beverly Hills, CA: Sage.

Glass, Gene V, & Smith, Mary Lee. (1978). Reply to Eysenck. American Psychologist, 33, 517-518.

Glass, R., & Meckler, R. (1972). Preparing elementary teachers to instruct mildly handicapped children in regular classrooms: A summer workshop. Exceptional Children, 39, 152-156.

Goodman, H., Gottlieb, J., & Harrison, R. (1972). Social acceptance of EMR's integrated into a non-graded elementary school. American Journal of Mental Deficiency, 76, 412-417.

Gottlieb, Jay. (1972). Bi-cultural study of attitude change and behavior toward retardates. Unpublished doctoral dissertation, Yeshiva University.

Granofsky, J. (1953). Modification of attitudes toward the visibly disabled: An experimental study of the effectiveness of social contact in producing a modification of the attitudes of non-disabled females toward visibly disabled males. Dissertation Abstracts, 16, 1182-1183.

Green, Bert F., & Hall, Judith A. (1984). Quantitative methods for literature reviews. Annual Review of Psychology, 35, 37-53.

Greenstein, Theodore. (1976). Behavior change through value self-confrontation: A field experiment. Journal of Personality and Social Psychology, 34(2), 254-262.

Haddle, H. (1974). The modification of attitudes toward disabled persons: The case for using systematic desensitization as an attitude-change strategy. American Foundation for the Blind Research Bulletin, No. 28.

Hafer, M., & Narcus, M. (1979). Information and attitudes toward disability. Rehabilitation Counseling Bulletin, 23(2), 95-102.

Harasymiw, S., & Horne, M. (1975). Integration of handicapped children: Its effect on teacher attitudes. Education, 96, 152-155.

Harth, R. (1973). Attitudes and mental retardation: Review of the literature. Training School Bulletin, 69, 150-164.

Hays, William L. (1973). Statistics for the social sciences (2nd ed.). New York: Holt, Rinehart and Winston.

Hedges, Larry V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. Journal of Educational Statistics, 6(2), 107-128.

Hedges, Larry V., & Olkin, Ingram. (1980). Vote-counting methods in research synthesis. Psychological Bulletin, 88, 359-369.

Hedges, Larry V., & Olkin, Ingram. (1985). Statistical methods for meta-analysis. New York: Academic Press.

Hedges, Larry V., & Olkin, Ingram. (1986). Meta analysis: A review and a new view. Educational Researcher, 15(8), 14-21.

Hersh, A., Carlson, R., & Lossino, D. (1975). Normalized interaction with families of the mentally retarded. Mental Retardation, 15, 32-33.

Hicks, J., & Spaner, F. (1962). Attitude change and mental hospital experience. Journal of Abnormal and Social Psychology, 65, 112-120.

Holzberg, J., & Gewirtz, H. (1963). A method of altering attitudes toward mental illness. Psychiatric Quarterly Supplement, 37, 56-61.

Hopkins, Kenneth D. (1982). The unit of analysis: Group means versus individual observations. American Educational Research Journal, 19(1), 5-18.

Horne, M. (1979). Attitudes and mainstreaming: A literature review for school psychologists. Psychology in the Schools, 16, 61-67.

Horne, Marcia D. (1985). Attitudes toward handicapped students: Professional, peer, and parent reactions. Hillsdale, NJ: Lawrence Erlbaum.

Horne, Marcia D. (1986, June). Modifying peer attitudes toward the handicapped: Procedures and research issues. Paper presented at the Conference on Attitudes Toward Persons With Disabilities, Hofstra University, Hempstead, NY.

Hovland, C., Janis, I., & Kelley, H. (1953). Communication and persuasion. New Haven: Yale University Press.

Hunter, John E., Schmidt, Frank L., & Jackson, Gregg B. (1982). Meta-analysis: Cumulating research findings across studies. Beverly Hills: Sage.

Hyde, J. S. (1981). How large are cognitive gender differences? A meta-analysis using ω^2 and d. American Psychologist, 36, 892-901.

Insko, Chester A. (1967). Theories of attitude change. New York: Meredith Publishing.

Jackson, Gregg B. (1978). Methods for reviewing and integrating research in the social sciences. Final Technical Report to National Science Foundation for Grant #DIS 76-20398. Washington, DC: The George Washington University, Social Research Group.

Jackson, G. (1980). Methods for integrative reviews. Review of Educational Research, 50, 438-460.

Johannsen, W. (1969). Attitudes toward mental patients: A review of empirical research. Mental Hygiene, 53, 218-228.

Jones, Reginald L. (Ed.). (1984). Attitudes and attitude change in special education: Theory and practice. Reston, VA: The Council for Exceptional Children.

Kearney, N., & Roccio, P. (1956). The effect of teacher education on the teacher's attitudes. Journal of Educational Research, 49, 703-708.

Kiesler, Charles A., Collins, Barry E., & Miller, Norman. (1969). Attitude change: A critical analysis of theoretical approaches. New York: John Wiley & Sons.

Kirby, F., & Toler, H. (1970). Modification of pre-school isolate behavior: A case study. Journal of Applied Behavior Analysis, 3, 309-314.

Kleinfield, Sonny. (1979). The hidden minority: A profile of handicapped Americans. Boston: Little, Brown, and Company.

Kuhn, J. (1971). A comparison of teachers' attitudes toward blindness and exposure to blind children. The New Outlook for the Blind, 65, 337-340.

Kulik, James A., & Kulik, Chen-Lin C. (1986, April). Operative and interpretable effect sizes in meta-analysis. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

La Bue, A. (1959). Teacher classroom attitudes. Journal of Teacher Education, 10, 433.

Ladas, H. (1980). Summarizing research: A case study. Peview of Educational Research, 50(4), 597-624.

Lane, P. (1976). Evaluative statements by prospective students as a function of ethnic and retardation labels. Dissertation Abstracts International, 37, 1491-A.

Lapp, Bernard. (1974). The effects of behavior modification inservice training on teacher behavior, teacher attitudes, and knowledge of behavior modification. Unpublished doctoral dissertation, University of Connecticut.

Lazar, A., Gensley, J., & Orpet, R. (1971). Changing attitudes of young mentally gifted children toward handicapped persons. Exceptional Children, 37, 600-602.

Levitt, Edith, & Cohen, Shirley. (1976). Attitudes of children toward their handicapped peers. Childhood Education, 52, 171-173.

Lewis, I., & Cleveland, D. (1966). Nursing students' attitudinal changes following a psychiatric affiliation. Journal of Psychiatric Nursing, 4, 223-231.

Leyser, Y., & Gottlieb, J. (1980). Improving the social status of rejected pupils. Exceptional Children, 48, 459-461.

Light, Richard J. (1979). Capitalizing on variation: How conflicting research findings can be helpful for policy. Educational Researcher, 8(8), 3-8.

Light, Richard J., & Pillemer, David B. (1982). Numbers and narrative: Combining their strengths in research reviews. Harvard Educational Review, 52(1), 1-26.

Light, Richard J., & Pillemer, David B. (1984). Summing up: The science of reviewing research. Cambridge: Harvard University Press.

Light, Richard J., & Smith, P. (1971). Accumulating evidence: Procedures for resolving contradictions among different research studies. Harvard Educational Review, 41(4), 429-471.

Makas, Elaine. (1985, August). The measurement of attitudes toward disabled people: A new approach. Paper presented at the Annual Convention of the American Psychological Association, Los Angeles.

Makas, Elaine. (1986, April). The relationship between contact with and attitudes toward people with disabilities: A question of theory or of method. Paper presented at the annual meeting of The Society for the Study of Chronic Illness, Impairment, & Disability, Reno, Nevada.

Mansfield, R. S., & Bussey, T. V. (1977). Meta-analysis of research: A rejoinder to Glass. Educational Researcher, 6, 3.

Marlowe, M. (1979). The games analysis intervention: A procedure to increase the peer acceptance and social adjustment of a retarded child. Education and Training of the Mentally Retarded, 14, 262-268.

Matkin, Ralph E., Hafer, Marilyn, Wright, W. Russell, & Lutzker, John R. (1983). Pretesting artifacts: A study of attitudes toward disabilities. Rehabilitation Counseling Bulletin, 26(5), 342-348.

McGaw, Barry, & Glass, Gene V. (1980). Choice of the metric for effect size in meta-analysis. American Educational Research Journal, 17(3), 325-337.

Middleton, J. (1953). The prejudices and opinions of mental hospital employees regarding mental illness. American Journal of Psychiatry, 110, 133-138.

Mitchell, Marlys M. (1976). Teacher attitudes. The High School Journal, 59, 302-312.

Mouly, G. (1978). Educational research: The art and science of investigation. Boston: Allyn and Bacon.

Mulkey, Sidney W. (1980). Standard stimulus effects on attitudes toward disabled persons: An experimental study. Unpublished doctoral dissertation, The Florida State University.

Murray, J. (1958). An experiment in changing the attitudes of employers toward mental illness. Mental Hygiene, 42, 402.

Nunnally, N. (1961). Popular conceptions of mental health. New York: Holt, Rinehart and Winston.

Oberle, Judson B. (1975). The effect of personalization and quality of contact on changing expressed attitudes and hiring preferences toward disabled persons. Unpublished doctoral dissertation, Syracuse University.

Orwin, Robert G., & Cordray, David S. (1985). Effects of reporting on meta-analysis: A conceptual framework and reanalysis. Psychological Bulletin, 97(1), 134-147.

Osgood, G., Succi, G., & Tannebaum, P. (1957). The measurement of meaning. Urbana: University of Illinois Press.

Ozyurek, Mehmet. (1977). Effects of live, audio, and print presentations of a discussion about physical disabilities on attitude modification toward disabled persons in Turkey. Unpublished doctoral dissertation, University of Northern Colorado.

Parish, T., Eads, G., [illegible]e, N., & Piscitello, M. (1977). Assessment and attempted modification of future teachers' attitudes toward handicapped children. Perceptual and Motor Skills, 44, 540-542.

Perkins-Karniski, M. (1978). The effect of increased knowledge of body systems and functions on attitudes toward the disabled. Rehabilitation Counseling Bulletin, 22, 16-20.

Perrow, Charles. (1981). Disintegrating social sciences. New York University Education Quarterly, 12(2), 2-9.

Pettit, L. (1956). Attitudes of relatives of long-hospitalized mental patients regarding convalescent leave. Mental Hygiene, 40, 251.

Preece, Peter F. W. (1983). A measure of experimental effect size based on success rates. Educational and Psychological Measurement, 43, 763-766.

Pulton, T. (1976). Attitudes toward the physically disabled: A review and a suggestion for producing positive attitude change. Physiotherapy Canada, 28, 83-88.

Quay, L., Bartlett, C., Wrightsman, L., & Catron, D. (1961). Attitude change in attendant employees. The Journal of Social Psychology, 55, 27-31.

Rabkin, J. (1972). Opinions about mental illness: A review of the literature. Psychological Bulletin, 77(3), 153-171.

Rapier, J., Adelson, R., Carey, R., & Croke, K. (1972). Changes in children's attitudes toward the physically handicapped. Exceptional Children, 39, 219-223.

Richardson, S. A., & Ronald, L. (1977). The effect of a physically handicapped interviewer on children's expression of values toward handicap. Rehabilitation Psychology, 24(4), 211-218.

Richardson, Stephen, Goodman, Norman, Hastorf, Albert, & Dornbusch, Stanford. (1961). Cultural uniformity in reaction to physical disabilities. American Sociological Review, 26, 241-247.

Rokeach, Milton. (1968). Beliefs, attitudes, and values. San Francisco: Jossey-Bass, Inc.

Rokeach, Milton. (1971). Long-range experimental modification of values, attitudes, and behavior. American Psychologist, 26(5), 453-459.

Rokeach, Milton. (1973). The nature of human values. New York: The Free Press.

Rokeach, Milton, & McLellan, D. Daniel. (1972). Feedback of information about the values and attitudes of self and others as determinants of long-term cognitive and behavioral change. Journal of Applied Social Psychology, 2(3), 236-251.

Rosenthal, Robert. (1978). Combining results of independent studies. Psychological Bulletin, 85(1), 185-193.

Rosenthal, Robert. (1984). Meta-analytic procedures for social research. Beverly Hills, CA: Sage.

Rosenthal, Robert, & Rosnow, Ralph L. (1975). The volunteer subject. New York: John Wiley and Sons.

Rosenthal, Robert, & Rubin, Donald B. (1982). Comparing effect sizes of independent studies. Psychological Bulletin, 92(2), 500-504.

Rosenthal, Robert, & Rubin, Donald D. (1982). A simple, general purpose display of magnitude of experimental effect. Journal of Educational Psychology, 74(2), 166-169.

Rosnow, Ralph L., & Rosenthal, Robert. (1976). The volunteer subject revisited. Australian Journal of Psychology, 28, 97-108.

Rosswurm, M. (1980). Changing nursing students' attitudes toward persons with physical disabilities. ARN, 5, 12-14.

Rowe, Joanne, & Stutts, Rose Marie. (In press). Impact of experience with disabled individuals on teachers' attitudes. Adapted Physical Activity Quarterly.

Rusalem, H. (1967). Engineering changes in public attitudes toward a severely disabled group. Journal of Rehabilitation, 33(3), 26-27.

Sadlick, Marie, & Penta, Frank B. (1975). Changing nurse attitudes toward quadriplegics through use of television. Rehabilitation Literature, 36, 274-278, 288.

Sandler, A., & Robinson, R. (1981). Public attitudes and community acceptance of mentally retarded persons: A review. Education and Training of the Mentally Retarded, 16, 97-103.

Segal, S. (1978). Attitudes toward the mentally ill: A review. Social Work, 23, 211-217.

Sellin, D., & Mulchahay, R. (1965). The relationship of an institutional tour upon opinions about mental retardation. American Journal of Mental Deficiency, 70, 408-412.

Shaver, James P. (1979). The productivity of educational research and the applied-basic research distinction. Educational Researcher, 8(1), 3-9. (a)

Shaver, James P. (1979). The usefulness of educational research in curricular/instructional decision-making in social studies. Theory and Research in Social Education, 7(3), 21-46. (b)

Shaver, James P. (1980, April). Readdressing the role of statistical tests of significance. Paper presented at the annual meeting of the American Educational Research Association, Boston.

Shaver, James P. (1983). The verification of independent variables in teaching methods research. Educational Researcher, 12(8), 2-9.

Shaver, James P. (1985). Chance and nonsense: A conversation about interpreting tests of statistical significance, Part 1. Phi Delta Kappan, 67(Sept.), 138-141. (a)

Shaver, James P. (1985). Chance and nonsense: A conversation about interpreting tests of statistical significance, Part 2. Phi Delta Kappan, 67(Oct.), 138-141. (b)

Shaver, James P., & Curtis, Charles K. (1981). Handicapism and equal opportunity: Teaching about the disabled in social studies. Reston, VA: Foundation for Exceptional Children. (a)

Shaver, James P., & Curtis, Charles K. (1981). Handicapism: Another challenge for social studies Social Education, 45(3), 208-211. (b)

Shaver, James P., & Norton, Richard S. (1980). Populations, samples, randomness, and replication in two social studies journals. Theory and Research in Social Education, 8(2), 1-10. (a)

Shaver, James P., & Norton, Richard S. (1980). Randomness and replication in ten years of the American Educational Research Journal. Educational Researcher, 9(1), 9-15. (b)

Shotel, J., Iano, R., & McGettigan, J. (1974). Teacher attitudes associated with integration of handicapped children. In G. J. Warfield (Ed.), Mainstream currents. Reston, VA: Council for Exceptional Children, 91-97.

Simpson, R., Parrish, N., & Cook, J. (1976). Modification of attitudes of regular class children towards the handicapped for the purpose of achieving integration. Contemporary Educational Psychology, 1, 46-51.

Simpson, S. N. (1980). Comment on "meta-analysis of research on class-size and achievement". Educational Evaluation and Policy Analysis, 2, 81-83.

Siperstein, Gary N., Bak, John J., & Gottlieb, Jay. (1977). Effects of group discussion on children's attitudes toward handicapped peers. The Journal of Educational Research, 70, 131-134.

Slavin, Robert E. (1984). Meta-analysis in education: How has it been used? Educational Researcher, 13(8), 6-15.

Slavin, Robert E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. Educational Researcher, 15(9), 5-11.

Smith, Mary Lee. (1980). Publication bias and meta-analysis. Evaluation in Education, 4, 22-24.

Smith, Mary Lee, & Glass, Gene V. (1980). Meta-analysis of research on class size and its relationship to attitudes and instruction. American Educational Research Journal, 17(4), 419-434.

Stevens, T., & Braun, B. (1980). Measures of regular classroom teachers' attitudes toward handicapped children. Exceptional Children, 46, 292-294.

Stock, William A., Okun, Morris A., Haring, Marilyn J., Miller, Wendy, & Ceurvorst, Robert W. (1982). Rigor in data synthesis: A case study of reliability in meta-analyis. Educational Researcher, 11(6), 10-14, 20.

Strain, P., & Timm, M. (1974). An experimental analysis of social interaction between a behaviorally disordered preschool child and her classroom peers. Journal of Applied Behavior Analysis, 7, 583-590.

Strauch, J. (1970). Social contact as a variable in the expressed attitudes of normal adolescents toward EMR pupils. Exceptional Children, 36, 495-500.

Thompson, Wayne N. (1975). The process of persuasion: Principles and readings. New York: Harper & Row.

Towner, A. (1984). Modifying attitudes toward the handicapped: A review of the literature and methodology. In R. Jones (Ed.), Attitudes and attitude change in special education: Theory and practice. Reston: VA: The Council for Exceptional Children.

Triandis, Harry C. (1971). Attitude and attitude change. New York: John Wiley & Sons.

Triandis, Harry C., Adamopoulos, John, & Brinberg, David. (1984). Perspectives and issues in the study of attitudes. In Reginald L. Jones (Ed.), Attitudes and attitude change in special education: Theory and practice. Reston, VA: The Council for Exceptional Children.

Tringo, John. (1970). The hierarchy or preference toward disability groups. The Journal of Special Education, 4, 295-306.

Voeltz, Luanna M. (1980). Children's attitudes toward handicapped peers. American Journal of Mental Deficiency, 84, 455-464.

Voeltz, L. (1982). Effects of structured interactions with severely handicapped peers on children's attitudes. American Journal of Mental Deficiency, 86, 380-390.

Wagner, Richard V., & Sherwood, John J. (Eds.). (1969). The study of attitude change. Belmont, CA: Brooks/Cole.

Wallston, B., Blanton, R., Robinson, J., & Pollchink, L. (1972). Community resources development in rehabilitation of the handicapped. Nashville, TN. (ERIC Document Reproduction Service, Document No. ED 078 239)

Walberg, Herbert J. (1986). Syntheses of research on teaching. In Merlin C. Wittrock (Ed.), Handbook of research on teaching (3rd ed.) (pp. 214-229). New York: Macmillan Publishing Company.

Warren, S., Turner, D., & Brody, D. (1964). Can education students' attitudes toward the retarded be changed? Mental Retardation, 2, 235-242.

Watts, William A. (1984). Attitude change: Theories and methods. In Reginald L. Jones (Ed.), Attitudes and attitude change in special education: Theory and practice. Reston, VA: Council for Exceptional Children.

Westwood, M., Vargo, J., & Vargo, F. (1981). Methods for promoting attitude change toward and among physically disabled persons. Journal of Applied Rehabilitation Counseling, 12(4), 220-225.

White, K., Bush, D., & Casto, G. (1985-86). Learning from reviews of early intervention. The Journal of Special Education, 19(4), 417-428.

White, Karl R., & Casto, Glendon. (1985). An integrative review of early intervention efficacy studies with at-risk children: Implications for the handicapped. Analysis and Intervention in Developmental Disabilities, 5, 7-31.

White, Karl R., Myette, Beverly, Baer, Ric rd, & Taylor, Cie. (1982). A meta-analysis of previous research on the treatment of hyperactivity (Final Report for Grant # HEW/OE/NIE-G-80-0008). Logan, UT: Exceptional Child Center, Utah State University. (ERIC Document Reproduction Service No. ED 224 218)

Willson, Victor L., & Putnam, Richard R. (1982). A meta-analysis of pretest sensitization effects in experimental design. American Educational Research Journal, 19(2), 249-258.

Wilson, E., & Alcorn, D. (1969). Disability simulation and development of attitudes toward the exceptional. The Journal of Special Education, 3, 303-307.

Wilson, G. Terence, & Rachman, S. J. (1983). Meta-analysis and the evaluation of psychotherapy outcome: Limitations and liabilities. Journal of Consulting and Clinical Psychology, 51(1), 54-64.

Wolf, Fredric M. (1986). Meta-analysis: Quantitative methods for research synthesis. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-001. Beverly Hills: Sage Publications.

Wolins, L. (1962). Responsibility for raw data. American Psychologist, 17, 657-658.

Wyrick, J. (1968). The effect of the lecture method on attitudes toward physical disability. Unpublished master's thesis, University of Kansas.

Yerxa, E. (1971). The effects of a dyadic, self-administered instructional program in changing the attitudes of female college students toward physically disabled persons. Dissertation Abstracts International, 32, 1931-1932A.

Yuker, Harold E. (1983). The lack of a stable order of preference for disabilities: Response to Richardson and Ronald. Rehabilitation Psychology, 28(2), 93-103.

Yuker, Harold E. (1986, June). The effects of contact on attitudes toward disabled persons: Some empirical generalizations. Abstracts of papers presented at the Attitudes Toward Persons With Disabilities Conference, Hofstra University, Hempstead, New York.

Yuker, Harold E., & Block, J. R. (1986). Research with the attitudes towards disabled persons scales (ATDP): 1960-1985. Hempstead, NY: Center for the Study of Attitudes Toward Persons With Disabilities, Hofstra University.

Yuker, Harold E., Block, J. R., & Younng, Janet H. (1970). The measurement of attitudes toward disabled persons. Albertson, NY: Human Resources Center. (ED 044 853)

Zimbardo, Philip G., Ebbesen, Ebbe B., & Maslach, Christina. (1977). Influencing attitudes and changing behavior (2nd ed.). Reading, MA: Addison-Wesley.

Zukerman, R. (1975). Changes in knowledge and attitudes as a result of participation in a teacher education game on the labeling of handicapped children. Dissertation Abstracts International, 36, 6031-6032A.

APPENDIX A

STATEMENT OF GENERAL PURPOSE AND POPULATIONS

META-ANALYSIS: MODIFYING ATTITUDES TOWARD DISABLED PERSONS

STATEMENT OF GENERAL PURPOSE AND POPULATIONS

The purpose of the meta-analysis project is to conduct a comprehensive, systematic review of the literature to determine what is known about modifying attitudes toward disabled persons. One objective is to identify research questions on which further investigation is especially needed, as well as those that do not merit further research. A secondary purpose is to analyze the quality of the research, both to determine if effect sizes covary with quality and because the quality of research is of interest in itself.

The population of studies is all English-language research reports of research to modify attitudes toward disabled persons. In initial screening, titles or references to studies will be examined for words that indicate that intent for the research. In screening actual reports for inclusion, only those reports that contain a clear statement of intent to investigate (1) the "modification of attitudes" (2) toward "persons with disabilities" (whether specifically disabilities or in general) will be included. Included will be journal articles, master's theses, doctoral dissertations, and other unpublished papers which are identifiable through conventional computer and hand search techniques and the bibliographies of reports that are read. The primary emphasis, therefore, will be on research conducted in North America, particularly the USA; but research from other countries will not be excluded.

The research of interest is that carried out to investigate empirically the effects of interventions, or treatments, on the attitudes of nondisabled persons toward disabled persons. Correlational research will not be included. For example, contact with disabled persons is one common type of intervention. However if a researcher went into a school district and obtained information to categorize students according to the amount of their contact with disabled persons during the prior year and then compared attitude means for contact-hour groups or correlated amount of contact with scores on an attitude toward disabled measure, that study would not qualify. Experimental and quasi-experimental studies are of particular interest. Pre-experimental single-group studies that involved a planned intervention and pretest-posttest data will be also included, as will static group designs used to investigate program effects. Mainstreaming studies will be included if the effects on attitudes toward disabled persons have been investigated, although much of this research has involved the use of pre-experimental designs.

Research reports of the effects of interventions on the attitudes of any age or occupational group are of interest as long as the research was directed toward changing attitudes toward disabled or handicapped persons.

"Disabled or handicapped persons" is defined in terms of conventional special education categories: mentally retarded, hard of hearing, deaf, speech impaired, visually handicapped, seriously emotionally disturbed (or mentally ill), orthopedically impaired, deaf-blind, multi-handicapped, and learning disabilities, as well as general categories such as "the disabled" or "the handicapped". Studies of subjects from populations such as "disruptive" students or "slow learners" are not to be included.

Attitudes toward disabled persons is the dependent variable of interest. Attitudes are considered to have cognitive, affective, and behavioral components and may be assessed in a variety of ways, including paper-and-pencil tests which have items that are cognitive-affective mixtures, assessments of changes in voluntary interactions with disabled persons, or reactions on projective-type tests. Measures which assess only knowledge about the disabled do not qualify for this meta-analysis unless the research report authors consider them to be attitude assessments, nor do measures which assess attitudes toward mainstreaming. General measures of attitudes toward children or other people will not be included unless specifically aimed at disabled persons or a particular type of disability, through instructions to the Ss or because of the content of the study--e.g., an attempt to change parents' attitudes toward their disabled children.

Measures such as sociometric scales, friendship choices, or positive interactions are relevant only if the researcher(s) consider them to be assessments of attitudes. Observational or other data gathered and/or reported so that the behaviors or responses of nondisabled Ss toward disabled persons or the direction of behavioral or response change cannot be identified will not be included, even if considered in the report to be attitude assessments.

The intent is to obtain effect sizes to be analyzed. However, studies which are experimental or quasi-experimental in nature for which effect sizes cannot be obtained will be coded for other information if time is available. Those studies will be analyzed separately.

APPENDIX B

CODING INSTRUMENTS

(1) Coding Instrument

(2) Prior Contact Coding Sheet

(3) Contact Coding Sheet

(4) ES Information Missing Coding Instrument

(5) Supplementary Sheets
- (a) Effect Sizes
- (b) Information Request
- (c) Comments on Study
- (d) Effect Size Computations
- (e) Comments on Conventions
- (f) Report Disposition

Report ID# ________
Time start ________
Time end ________
Instrument ____ of ____

________ Coder   ________ Date

META-ANALYSIS: MODIFYING ATTITUDES TOWARD DISABLED PERSONS

CODING INSTRUMENT

________ (Author(s)/Year)   ________ (Abbreviated Title & Source)

Checklist

___ 1. Citation checked
___ 2. References checked
___ 3. Every space marked
___ 4. ES data available
___ 5. ES's computed
___ 6. ES's checked
___ 7. Comments on Conventions Sheet
___ 8. Comments on Study Sheet
___ 9. Scoring log completed
___ 10. Report disposition sheet completed

A. General Information

	ES__	ES__	ES__	ES__	
1-2	M A	M A	M A	M A	*1. Project code
3-6	_ _ _ _	_ _ _ _	_ _ _ _	_ _ _ _	*2. Report ID #
7-8	0 1	0 1	0 1	0 1	*3. Card #
9-10	_ _	_ _	_ _	_ _	*4. Year of publication
11	_	_	_	_	*5. Type of report (1=journal; 2=book chapter; 3=book; 4=dissertation; 5=thesis; 6=convention paper; 7=unpublished report; 8=other: Explain ________; 9=combination, specify ________)

6. Effect Size(s)

Description of ES Comparison(s)

ES# ___ ________________
ES# ___ ________________
ES# ___ ________________
ES# ___ ________________

350

	ES__	ES__	ES__	ES__	
12-13	_ _	_ _	_ _	_ _	a. N of ES's
14-15	_ _	_ _	_ _	_ _	b. ID of ES
16	_	_	_	_	c. Level (1=primary; 2=secondary)
17-18	_ _	_ _	_ _	_ _	d. Type of comparison (1=treatment vs. control; 2=treatment vs. placebo; 3=treatment A vs. B; 4=pre-post; Interaction, treatment by: 5=gender, 6=age/grade, 7=testing, 8=personality, specify ______, 9=other, specify ______; Treatment within: 10=gender, female; 11=gender, male; 12=other, specify ______; Prior contact: 13=within treatment; 14=across groups; 15=interaction)
19	_	_	_	_	*7. Target Population (0=not mentioned; 1=term used; 2=defined; 3=population described; 4=1&2; 5=1&3)
20	_	_	_	_	*8. Accessible Population (0=not mentioned; 1=term used; 2=defined; 3=described; 4=1&2; 5=1 & 3; 6=1, 2, & 3)
					9. Replication (0=No; 1=direct; 2=systematic; 3=pseudo, within)
21	_	_	_	_	a. of other research
22	_	_	_	_	b. within study
					B. Description of Sample(s)
					1. N:
23-27	_ _ _ _ _	_ _ _ _ _	_ _ _ _ _	_ _ _ _ _	a. of total sample
28-31	_ _ _ _	_ _ _ _	_ _ _ _	_ _ _ _	b. of experimental Ss
32-35	_ _ _ _	_ _ _ _	_ _ _ _	_ _ _ _	c. of control Ss

	ES__	ES__	ES__	ES__	
					2. Sample Selection (0=can't tell; 1=random; 2=solicited/volunteer; 3=captive/intact group; 4=random, group; 5=other, specify: ________)
36	_	_	_	_	Experimental
37	_	_	_	_	Control
					3. % Male
38-40	___	___	___	___	Total Sample
41-43	___	___	___	___	Experimental
44-46	___	___	___	___	Control
47	_	_	_	_	4. Treatment Context (0=can't tell; 1=elementary or secondary schooling; 2=college/university education; 3=adult education; 4=inservice; 5=work; 6=community; 7=recreation; 8=other, specify ________________)
					5. Educational Level of Ss (0=can't tell; 1=preschool; 2=primary; 3=intermediate; 4=middle school; 5=junior high school; 6=senior high school; 7=combination, specify ________________; 8=undergraduates; 9=graduate students; 10=post-professional; 11=adults not in school; 12=other, ________________)
48-49	__	__	__	__	Experimental
50-51	__	__	__	__	Control
					6. University Students According to Major (0=not applicable; 1=can't tell; 2=elementary education; 3=secondary education; 4=elementary & secondary education; 5=education, unspecified; 6=nursing; 7=occupational & physical therapy; 8=philosophy; 9=psychology; 10=rehabilitation counseling; 11=social work; 12=sociology; 13=special education; 14=communicative disorders; 15=medicine; 16=other, specify ____________)
52-53	__	__	__	__	Experimental
54-55	__	__	__	__	Control

	ES__	ES__	ES__	ES__	
					7. Occupation of Nonstudent Ss (0=not applicable; 1=can't tell; 2=child care workers; 3=community recreation workers; 4=employees in institutions; 5=regular class teachers, elementary; 6=middle or juniorhigh; 7=high school; 8=school administrators; 9=special class teachers; 10=vocational rehabilitation counselors; 11=parents; 12=police; 13=medical; 14=general public; 15=other, specify ____________)
56-57	__ __	__ __	__ __	__ __	Experimental
58-59	__ __	__ __	__ __	__ __	Control
					8. Prior Experiences with Disabled Persons (0=can't tell; 1=none; 2=as parents; 3=as siblings; 4=classmates; 5=as teachers; 6=in school; 7=as co-workers; 8=as supervisors; 9=in work setting; 10=as clients/patients; 11=general, not specific; 12=combination, specify ________; 13=other, specify ____________)
60-61	__ __	__ __	__ __	__ __	Experimental
62-63	__ __	__ __	__ __	__ __	Control
64	__	__	__	__	9. Country of subjects (1=USA; 2=Canada; 3=Australia/New Zealand; 4=Europe, specify ____________; 5=other, specify ________________)
					C. Treatment/Intervention
65	__	__	__	__	1. Basis (0=can't tell; 1=theory, explicit; 2=prior research, no citations; 3=prior research, few citations; 4=prior research, case developed; 5=practical experience/insight; 6=other, specify ________)
					2. Attitude Change Theory
66	__	__	__	__	a. Theory (1=S-R/behavioral; 2=conditioning; 3=congruity/equilibrium; 4=social judgment; 5=functional; 6=combination, specify ____________)
67	__	__	__	__	b. Relationship to treatment (1=mentioned but not used; 2=brief allusion; 3=explicit, well developed basis; 4=post hoc interpretation; 5=implicit). If mainstreaming study with no prior intervention to change attitudes, skip to C.10., and X-out Sections C.3-9.

	ES__	ES__	ES__	ES__	
					3. Setting (0=can't tell; 1=regular classroom; 2=special classroom; 3=home; 4=institution; 5=group home; 6=hospital; 7=dormitory; 8=playground; 9=camp; 10=recreation facility; 11=laboratory; 12=individual/small group; 13=normal life; 14=other, specify ________; 15=combination, specify ________)
68-69	__ __	__ __	__ __	__ __	Experimental
70-71	__ __	__ __	__ __	__ __	Control
					4. Treatment/Intervention Technique(s) (0=none/control; 1=placebo; 2=information; 3=direct contact; 4=vicarious experience; 5=positive reinforcement; 6=persuasive message; 7=persuasive messages, contrast; 8=2&3; 9=2&4; 10=other, specify ____; 11=systematic desensitization)
72-73	__ __	__ __	__ __	__ __	Experimental
74-75	__ __	__ __	__ __	__ __	Control
					a. Information (1) Type (0=none; 1=etiology; 2=characteristics; 3=problems; 4=similarities with nondisabled; 5=prostheses and special equipment; 6=famous disabled people; 7=legal rights; 8=parenting/management; 9=self; 10=social relations; 11=other, specify ________; 12=combination, specify ________)
76-77	__ __	__ __	__ __	__ __	Experimental
78-79	__ __	__ __	__ __	__ __	Control

(End of Card #1)

* *

1-2	M A	M A	M A	M A	Project code
3-6	__ __ __ __	__ __ __ __	__ __ __ __	__ __ __ __	Report ID#
7-8	0 2	0 2	0 2	0 2	Card #

	ES__	ES__	ES__	ES__	
					(2) Delivery mode (0=none; 1=lecture; 2=discussion; 3=lecture/discussion; 4=print; 5=panel discussion-disabled; 6=panel discussion-nondisabled; 7=speaker-disabled; 8=speaker-nondisabled; 9=video or films; 10=pictures/photos/film-strips; 11=case study; 12=audio; 13=simulations; 14=regular course, specify ____________; 15=regular program, specify ____________; 16=other, specify __________; 17=combination, specify ________________)
9-10	__	__	__	__	Experimental
11-12	__	__	__	__	Control
					b. Direct contact (0=none; 1=as companions; 2=as peer tutors; 3=in cooperative learning groups; 4=as classmates; 5=as classmates, behavior modified; 6=as students, behavior modified; 7=practice teaching-special classes; 8=class-room observation; 9=supervised playground activities; 10=in recreation programs; 11=panel discussions; 12=guest speaker(s); 13=visit to institution/residential facilities; 14=integrated community programs; 15=as teacher/counselor; 16=as co-workers; 17=other, specify __________; 18=combination, specify ________________)
13-14	__	__	__	__	Experimental
15-16	__	__	__	__	Control
					c. Vicarious experience (0=none; 1=role play-contact; 2=role play-disabled; 3=simulation; 4=observation of role play or simulation; 5=videotapes or films; 6=case studies; 7=pictures/photos; 8=print fiction/biography; 9=dolls/puppets 10=other, specify ______________; 11=combination, specify ______________)
17-18	__	__	__	__	Experimental
19-20	__	__	__	__	Control
					d. Positive reinforcement (0=none; 1=covert; 2=overt)
21	_	_	_	_	Experimental
22	_	_	_	_	Control
					e. Persuasive message (0=none; 1=video/film; 2=audio; 3=print; 4=expert; 5=expert, disabled; 6=self-presentation;7=other,specify ________________; 8=combination, specify ________________)
23	_	_	_	_	Experimental
24	_	_	_	_	Control

	ES__	ES__	ES__	ES__	
25-26	_ _	_ _	_ _	_ _	5. Treatment/Intervention to Change Attitudes Toward: (1=disabled in general; 2=physically disabled in general; 3=retarded, general; 4=moderately retarded; 5=severely retarded; 6=mentally ill; 7=emotionally disturbed; 8=visually impaired/blind; 9=hearing impaired/deaf; 10=learning disabled; 11=speech/language impaired; 12=cerebral palsied; 13=epileptic; 14=para/quadriplegic; 15=other physically impaired, specify ____________; 16=autistic; 17=health impaired; 18=multiply handicapped; 19=other, specify ______________; 20=combination, specify ______________)
					6. Treatment/Intervention Conducted By: (0=can't tell; 1=experimenter; 2=project assistants; 3=regular staff; 4=combination, specify ____________; 5=other, specify ____________; 6=not applicable)
27	_	_	_	_	Experimental
28	_	_	_	_	Control
					7. Length of Treatment/Intervention
29	_	_	_	_	a. Information available (0=no; 1=yes)
					b. Days/week
30	_	_	_	_	Experimental
31	_	_	_	_	Control
					c. Minutes/day
32-34	_ _ _	_ _ _	_ _ _	_ _ _	Experimental
35-37	_ _ _	_ _ _	_ _ _	_ _ _	Control
					d. Number of weeks (If less than 1, insert x's)
38-40	_ _ _	_ _ _	_ _ _	_ _ _	Experimental
41-43	_ _ _	_ _ _	_ _ _	_ _ _	Control
					e. Total # of hours
44-48	_ _ _ _ . _	_ _ _ _ . _	_ _ _ _ . _	_ _ _ _ . _	Experimental
49-53	_ _ _ _ . _	_ _ _ _ . _	_ _ _ _ . _	_ _ _ _ . _	Control

	ES__	ES__	ES__	ES__	
					8. Verification of Treatment Implementation (0=none; 1=not necessary; 2=systematic observation; 3=nonstructured observation; 4=interviews with Ss; 5=questionnaires to Ss; 6=intervenor follow-up; 7=other, specify ______________; 8=combination, specify ________________)
54	_	_	_	_	Experimental
55	_	_	_	_	Control
					a. Reporting (0=not applicable; 1=data; 2=data and analysis presented; 3=assertion by author(s))
56	_	_	_	_	Experimental
57	_	_	_	_	Control
					b. Degree of implementation claimed (0=not applicable; 1=none; 2=some; 3=mostly; 4=complete)
58	_	_	_	_	Experimental
59	_	_	_	_	Control
					c. Basis for author's conclusion re implementation (0=not applicable; 1=can't tell; 2=author's judgment or inference; 3=statistical significance; 4=inspection of data; 5=combination, specify ________________; 6=other, specify ______________________________)
60	_	_	_	_	Experimental
61	_	_	_	_	Control
62	_	_	_	_	d. Actual implementation (1=complete; 2=mostly; 3=only in part)
63	_	_	_	_	e. Description of treatment adequate for replication (0=no; 1=somewhat; 2=yes)

	ES__	ES__	ES__	ES__	
					9. Treatment validity
64	_	_	_	_	a. Implementation (from 8d)
65	_	_	_	_	b. Hawthorne
66	_	_	_	_	c. John Henry
67	_	_	_	_	d. Treatment diffusion
68	_	_	_	_	e. Dissatisfaction/resentment
69	_	_	_	_	f. Novelty/disruption
70	_	_	_	_	g. Experimenter effect/expectations
71	_	_	_	_	h. Treatment/experimenter confounded
72	_	_	_	_	i. Test x treatment interaction
73	_	_	_	_	j. Multiple treatment interference
74	_	_	_	_	k. General Treatment Validity (1=excellent; 2=fair; 3=poor)
					10. Mainstreaming
75	_	_	_	_	a. Presence (0=no mainstreaming; 1=mainstreaming only; 2=pretreatment plus mainstreaming)
76	_	_	_	_	b. Type of study (0=can't tell; 1=planned; 2=post hoc)
77	_	_	_	_	c. Instruction in mainstreamed classes (0=can't tell; 1=standard group/class instruction; 2=cooperative learning; [illegible]ndividualized instruction; 4=peer tut[illegible] 5=combination, specify ____________; 6=other, specify ____________)
78	_	_	_	_	d. Special personnel support (0=can't tell; 1=none; 2=teacher preparation; 3=consultant help; 4=other, specify ____________)
79	_	_	_	_	e. Special skills training for disabled students (0=can't tell; 1=none; 2=social; 3=academic; 4=both)

(0=can't tell; 1=not a plausible threat; 2=minor problem; 3=substantial problem; 4=major problem; 5=not applicable. For 3, 4, & 5, write reason next to item.)

(End of Card #2)

* *

	ES__	ES__	ES__	ES__	
1-2	M A	M A	M A	M A	Project Code
3-6	_ _ _ _	_ _ _ _	_ _ _ _	_ _ _ _	Report ID#
7-8	0 3	0 3	0 3	0 3	Card #
9	_	_	_	_	f. Type of special skills training. (0=not applicable; 1=coaching; 2=modeling; 3=counseling; 4=direct reinforcement; 5=group contingencies; 6=diagnostic/prescriptive; 7=cognitive control; 8=combination, specify ____________; 9=other, specify ________)
10	_	_	_	_	g. Special instruction for nondisabled peers (0=can't tell; 1=none; 2=information; 3=vicarious experience; 4=reinforcement; 5=persuasive messages)
11	_	_	_	_	h. Parent education (0=can't tell; 1=none; 2=disabled children; 3=nondisabled children; 4=2&3)
12	_	_	_	_	i. Type of parent education (0=can't tell; 1=not applicable; 2=information; 3=vicarious experience; 4=reinforcement; 5=persuasive messages)
13-14	_ _	_ _	_ _	_ _	j. Disabled children in mainstreamed classes (0=can't tell; 1=mildly and moderately retarded; 2=emotionally disturbed; 3=visually impaired; 4=blind; 5=hearing impaired; 6=deaf; 7=communication disordered; 8=physically and health impaired; 9=learning disabled; 10=combination, specify __________; 11=others, specify ______________)
15-17	_ _ _	_ _ _	_ _ _	_ _ _	k. Number of minutes handicapped children spend daily in mainstreamed class
18	_	_	_	_	l. Number of days per week in mainstreamed class
19-22	_ _ _ _	_ _ _ _	_ _ _ _	_ _ _ _	m. Total minutes per week in mainstreamed class
23-25	_ _ _	_ _ _	_ _ _	_ _ _	n. Months in mainstreamed program when outcomes assessed
26	_	_	_	_	o. Outcomes measured for (1=nondisabled peers; 2=teachers; 3=parents, disabled children; 4=parents, nondisabled children; 5=administrators; 6=combination, specify ____________)

	ES__	ES__	ES__	ES__	D. Dependent Measures
27	_	_	_	_	*1. "Attitude" defined (0=no; 1=affective; 2=cognitive; 3=behaviors; 4=1&2; 5=1&3; 6=1, 2, & 3)
28-29	_ _	_ _	_ _	_ _	*2. Number of dependent measures. List (by name of instrument and form, if possible, or general category): 1. ________ 4. ________ 2. ________ 5. ________ 3. ________ 6. ________
30	_	_	_	_	3. Use of common instrument (0=no; 1=ATDP; 2=OMI; 3=RGEPS; 4=ATHI; 5=ATBS; 6=MRAI; 7=MTAI, revised; 8=combination, specify ________)
31	_	_	_	_	4. Common instrument modified (0=not applicable; 1=no; 2=for different population of Ss; 3=for different disabilities; 4=other, specify ________)
32	_	_	_	_	5. Source of data (1=self-report; 2=opinion of other, teacher; 3=opinion of other, administrator; 4=opinion of other, specify ________ ________; 5=observation; 6=nonproject request)
33-34	_ _	_ _	_ _	_ _	6. Type of assessment (1=interview, structured; 2=interview, nonstructured; 3=attitude questionnaire; 4=sociometric measure; 5=peer assessment; 6=social distance scale; 7=informal observation; 8=systematic observation; 9=semantic differential; 10=telephone or mail survey; 11=telephone or mail request; 12=Q-sort; 13=projective test, pictures; 14=sentence completion; 15=adjective checklist; 16=rankings; 17=other, specify ________)
35	_	_	_	_	7. Source of instrument (0=not applicable; 1=can't tell; 2=teacher-made; 3=project developed; 4=prior research; 5=other, specify ________)

	ES__	ES__	ES__	ES__	
36	_	_	_	_	8 Development by project (0=not applicable; 1=description not provided; 2=description not adequate; 3=adequate description)
					9. Reliability of scores
37	_	_	_	_	a. Mentioned? (0=no; 1=coefficient reported; 2=yes, but no coefficient)
38	_	_	_	_	b. Source (0=not applicable; 1=can't tell; 2=computed on sample; 3=reported from other research; 4=pilot study; 5=combination, specify ________)
39-40	_ _	_ _	_ _	_ _	c. Method (0=not applicable; 1=can't tell; 2=test-retest; 3=internal consistency; 4=alternate forms; 5=inter-observer--%; 6=inter-observer--r; 7=intra-observer--%; 8=intra-observer--r; 9=categorization reliability; 10=combination, specify ____________________)
41-42	_ _	_ _	_ _	_ _	d. Coefficient
43	_	_	_	_	e. Magnitude (0=not applicable; 1=.80-1.0; 2=.60-.79; 3=.0-.59)
					10. Validity of scores
44	_	_	_	_	a. Discussed (0=no; 1=moderately; 2=comprehensively)
45	_	_	_	_	b. Type (0=not applicable; 1=general; 2=face; 3=construct-discrimination or correlations; 4=construct--expert judgment; 5=concurrent; 6=combination, specify __________________)
46	_	_	_	_	c. Source (0=not applicable; 1=not mentioned; 2=general reference to literature; 3=citation of research; 4=project data; 5=inference without data; 6=combination, specify _______________)
47	_	_	_	_	d. Reactivity of measure (1=low; 2=moderate; 3=high)
48	_	_	_	_	e. Adequacy of validity (1=low; 2=moderate; 3=high)

	ES__	ES__	ES__	ES__	
					11. Data collection
49	_	_	_	_	a. Type (0=can't tell; 1=regular staff, 2=researcher; 3=research assistant; 4=noncontact; 5=other, specify ________)
50	_	_	_	_	b. Blinded collection (0=can't tell; 1=not applicable; 2=no; 3=partial—experimental or control, or pre or post only; 4=yes)
51	_	_	_	_	12. Blinded scoring (0=can't tell; 1=not applicable; 2=no; 3=partial—experimental or control, or pre or posttest only; 4=yes)
52	_	_	_	_	13. Time of posttest (0=can't tell; 1=immediate; 2=delayed; 3=follow-up)
53-56	_ _ _._	_ _ _._	_ _ _._	_ _ _._	14. Weeks after intervention to posttest
					E. Internal Validity
57	_	_	_	_	1. Design (1=pre-post, control; 2=posttest-only; 3=Solomon 4-group; 4=nonequivalent control group; 5=single subject; 6=pre-post, one-group; 7=static group; 8=other, specify ________)
58	_	_	_	_	2. Assignment to groups (0=can't tell; 1=random; 2=match-random; 3=select or match from different group; 4=random assignment of intact groups; 5=convenience; 6=not applicable; 7=other, specify ________)
					3. Threats
59	_	_	_	_	a. Treatment Validity (from C.9.k., p. 9)
60	_	_	_	_	b. Maturation
61	_	_	_	_	c. History
62	_	_	_	_	d. Testing
63	_	_	_	_	e. Instrumentation
64	_	_	_	_	f. Statistical regression
65	_	_	_	_	g. Selection
66	_	_	_	_	h. Experimental mortality
67	_	_	_	_	i. General Internal Validity (1=high; 2=medium; 3=low)

(0=can't tell; 1=not a plausible threat; 2=minor threat; 3=substantial threat; 4=major problem. For 3 & 4, write reason next to item.)

	ES__	ES__	ES__	ES__	F. Results
68	_	_	_	_	1. Statistical significance (0=not available; 1=not significant at .05 level; 2=significant at the .05 level)
69	_	_	_	_	2. Author's conclusions about effectiveness a. Qualified? (0=no; 1=sample; 2=interactions; 3=design; 4=measures; 5=treatment verification; 6=replication; 7=other, specify ______ ________; 8=combination, specify __________)
70	_	_	_	_	b. Treatment effective? (0=none; 1=didn't have effect; 2=data equivocal; 3=produced effect; 4=produced negative effect)
71	_	_	_	_	3. Effect size(s) a. Available (0=no; 1=yes, positive change; 2=yes, negative change; 3=no, negative; 4=no, positive)
72-77	_ _ _ _._ _	_ _ _ _._ _	_ _ _ _._ _	_ _ _ _._ _	b. D. 1) ± D

(End of Card #3)

* *

1-2	M A	M A	M A	M A	Project code
3-6	_ _ _ _	_ _ _ _	_ _ _ _	_ _ _ _	Report ID#
7-8	0 4	0 4	0 4	0 4	Card #
9-10	_ _	_ _	_ _	_ _	2) Source (0=not applicable; 1=reported; 2=calculated; 3=t or ANOVA F; 4=correlated t; 5=correlated t (.50); 6=n-way ANOVA; 7=COVAR; 8=COVAR (.50); 9=proportions, chi-square; 10=other nonparametric; 11=r_{pb}; 12=significance level; 13=other, specify __________)

	ES__	ES__	ES__	ES__	
11	_	_	_	_	3) Scale of $\bar{X}$ difference (0=not applicable; 1=raw post; 2=raw gain; 3=covariance adjusted; 4=residual gain
12-13	__	__	__	__	4) Standard deviation (0=not applicable; 1=post control/placebo; 2=pretest; 3=1&2 pooled; 4=1-way ANOVA SSs; 5=n-way ANOVA SSs; 6=1-way ANCOVA SSs; 7=n-way ANCOVA SSs; 8=adjusted COVAR sd; 9=adjusted gain sd; 10=pooled post)
14-19	____.___	____.___	____.___	____.___	5) Mdn primary D for type of assessment
20-25	____.___	____.___	____.___	____.___	6) Mdn primary D, overall
					c. Correlation
26	_	_	_	_	1) Type (0=none; 1=r_{pb}; 2=E^2; 3=Phi; 4=Cramer's V; 5=other, specify ________)
27-29	___	___	___	___	2) ± coefficient
30	_	_	_	_	3) Source (0=not applicable; 1=reported; 2=calculated; 3=estimated from D)
31-33	___	___	___	___	4) Mdn primary coefficient for type of assessment
34-36	___	___	___	___	5) Mdn primary coefficient, overall
					d. Variance ratio
37-41	___.___	___.___	___.___	___.___	1) Ratio (S_E^2/S_C^2)
42-46	___.___	___.___	___.___	___.___	2) Mdn primary variance ratio, type of assessment
47-51	___.___	___.___	___.___	___.___	3) Mdn primary variance ratio, overall

	ES__	ES__	ES__	ES__	
					G. Supplemental Information
					1. Information gain
52	_	_	_	_	a. Reported? (0=not applicable; 1=yes; 2=no)
53	_	_	_	_	b. Type of report (0=not applicable; 1=verbal; 2=statistical significance; 3=descriptive statistics; 4=2&3)
54	_	_	_	_	c. Conclusions (0=not applicable; 1=clear gain; 2=no gain; 3=mixed results; 4=can't tell/inconclusive)
55-60	____.__	____.__	____.__	____.__	d. Effect size--D
61	_	_	_	_	2. Type of study (0=can't tell; 1=course evaluation; 2=program evaluation; 3=experimental treatment; 4=mainstreaming, variation in time)
					H. Coding Summary
62-64	___	___	___	___	1. Minutes spent coding
65	_	_	_	_	2. Coder (1=Curtis; 2=Jesunathadas; 3=Shaver; 4=Strong)

RETURN TO PAGE 1 AND COMPLETE CHECKLIST

Coder ____ Date ____
Time start ____
Time end ____
Sheet ___ of ___

Report ID# ________

____
Author(s)/Date

META-ANALYSIS: MODIFYING ATTITUDES TOWARD DISABLED PERSONS

I. Prior Contact Coding Sheet

	ES__	ES__	ES__	ES__	
1-2	M A	M A	M A	M A	*1. Project code
3-6	_ _ _ _	_ _ _ _	_ _ _ _	_ _ _ _	2. Report ID
7-8	0 5	0 5	0 5	0 5	3. Card #
9	_	_	_	_	4. Coder (1=Curtis; 2=Jesunathadas; 3=Shaver; 4=Strong)
10	_	_	_	_	5. Prior contact assessed (0=No; 1=yes; 2=implicit)
11-12	_ _	_ _	_ _	_ _	6. Use--selection/assignment (0=not applicable; 1=none; 2=describe sample only; 3=eligibility as Ss--none; 4=eligibility as Ss--yes; 5=selection strata; 6=assignment strata; 7=5&6; 8=covariate; 9=other, ________; 10=combination, ________)
13	_	_	_	_	7. Use--outcome analysis, correlational (0=not applicable; 1=none, 2=within treatment, post; 3=within treatment, change; 4=treatment vs. control, post; 5=treatment vs. control, change)
14-15	_ _	_ _	_ _	_ _	8. ID of ES
16-17	_ _	_ _	_ _	_ _	9. N of added secondary ESs
18-19	_ _	_ _	_ _	_ _	10. Type of assessment (0=Not applicable; 1=can't tell; 2=questionnaire; 3=interview; 4=setting; 5=other, specify, ________; 6=2&3; 7=2&4; 8=3&4; 9=2,3,&4; 10=combination, ________)
20	_	_	_	_	11. Type of prior contact setting (0=not applicable; 1=mainstreamed classroom; 2=mainstreamed, school; 3=mainstreamed, regular teacher; 4=special education; 5=institution; 6=family; 7=other ________; 8=combination, ________)

	ES__	ES__	ES__	ES__	
21-22	_ _	_ _	_ _	_ _	12. Definition of degree of contact (0=not applicable; 1=can't tell; 2=amount; 3=number; 4=frequency; 5=length; 6=intensity; 7=dichotomy; 8=type of relationship; 9=institutional visit; 10=other, specify, __________; 11=combination, __________)
23	_	_	_	_	13. Quality of contact (0=not applicable; 1=can't tell; 2=no; 3=yes)
24	_	_	_	_	14. Direction of quality comparison (0=not applicable; 1=positive vs. neutral; 2=negative vs. neutral; 3=positive vs. negative)
25-26	_ _	_ _	_ _	_ _	15. Type of disability (0=not applicable; 1=disabled in general; 2=physically disabled in general; 3=retarded, general; 4=moderately retarded; 5=severely retarded; 6=mentally ill; 7=emotionally disturbed; 8=visually impaired/blind; 9=hearing impaired/deaf; 10=learning disabled; 11=speech/language impaired; 12=cerebral palsied; 13=epileptic; 14=para/quadriplegic; 15=other physically impaired, specify __________; 16=autistic; 17=health impaired; 18=multiply handicapped; 19=other, specify __________; 20=combination, specify __________)
27-29	_ _ _	_ _ _	_ _ _	_ _ _	16. Minutes spent coding

Coder ____ Date ____

Time: Start ____
End ____
Sheet ____ of ____

Report ID# ____

____
Author(s)/Date

META-ANALYSIS: MODIFYING ATTITUDES TOWARD DISABLED PERSONS

J. Contact Coding Sheet

	ES__	ES__	ES__	ES__	
1-2	M A	M A	M A	M A	1. Project code
3-6	_ _ _ _	_ _ _ _	_ _ _ _	_ _ _ _	2. Report ID
7-8	0 6	0 6	0 6	0 6	3. Card #
9	_	_	_	_	4. Coder (2=Jesunathadas; 3=Shaver)
					5. Status
10	_	_	_	_	a. Age (0=can't tell; 1=disabled younger; 2=same; 3=disabled older; 4=variety)
11	_	_	_	_	b. Educational-vocational prestige (0=can't tell; 1=disabled lower; 2=same; 3=disabled higher)
12	_	_	_	_	c. Helping relationship (0=none; 1=professional 2=preprofessional; 3=nonprofessional; 4=mutual help; 5=disabled the helper)
13	_	_	_	_	d. Overall (0=can't tell; 1=disabled lower; 2=equal; 3=disabled higher)
14-15	_ _	_ _	_ _	_ _	6. ID of ES
					7. Type of contact
16	_	_	_	_	a. Voluntariness (0=can't tell; 1=assigned; 2=role choice; 3=voluntary; 4=varied)
17	_	_	_	_	b. Intimacy (0=can't tell; 1=no interaction; 2=casual personal contact; 3=close personal contact; 4=varied contact; 5=potential contact)
18	_	_	_	_	c. Cooperation-competition (0=can't tell; 1=no opportunity; 2=not necessary; 3=implicit cooperation; 4=explicit cooperation; 5=implicit competition; 6=explicit competition; 7=combination of cooperation and competition)

	ES__	ES__	ES__	ES__	
19	_	_	_	_	d. Reinforcement (0=can't tell; 1=none; 2=shared, intrinsic; 3=shared, external; 4=nondisabled; 5=2 & 4; 6=3 & 4)
20	_	_	_	_	e. Pleasantness (0=can't tell; 1=no; 2=yes)
21	_	_	_	_	f. Modeling (0=can't tell; 1=none; 2=by peers; 3=by significant others)
22	_	_	_	_	8. Institutional/authority/peer support (0=can't tell; 1=no; 2=yes)
					9. Characteristics of disabled persons
23-24	__	__	__	__	a. Disability (0=can't tell; 1=combination (specify ______________; see p. J-7 of Conventions)
25	_	_	_	_	b. Negative stereotype (0=can't tell; 1=yes; 2=no)
26	_	_	_	_	c. Competence (0=can't tell; 1=lacked; 2=acknowledge/acceptance; 3=competent)
					10. Characteristics of nondisabled
27	_	_	_	_	a. Personality related (0=not assessed; 1=not tested; 2=no; 3=mixed results; 4=yes)
28	_	_	_	_	b. Prior attitudes related (0=not assessed; 1=not tested; 2=no; 3=mixed results; 4=yes)
29	_	_	_	_	11. Treatment (1=contact; 2=contact & information; 3=contact & vicarious experience)
30-32	___	___	___	___	12. Minutes spent coding

Report ID# ________
Time start ________
Time end ________
Instrument ____ of ____

________ Coder ________ Date

META-ANALYSIS: MODIFYING ATTITUDES TOWARD DISABLED PERSONS

CODING INSTRUMENT

ES Information Missing

________________ (Author(s)/Year) ________________ (Abbreviated Title & Source)

Checklist

___ 1. Citation checked
___ 2. References checked
___ 3. Every space marked
___ 4. ES data available
___ 5. ES's computed
___ 6. ES's checked
___ 7. Comments on Conventions Sheet
___ 8. Comments on Study Sheet
___ 9. Scoring log completed
___ 10. Report disposition sheet completed

A. General Information

	ES__	ES__	ES__	ES__	
1-2	M A	M A	M A	M A	*1. Project Code
3-6	_ _ _ _	_ _ _ _	_ _ _ _	_ _ _ _	*2. Report ID #
7-8	0 1	0 1	0 1	0 1	*3. Card #
9-10	_ _	_ _	_ _	_ _	*4. Year of publication
11	_	_	_	_	*5. Type of report (1=journal; 2=book chapter; 3=book; 4=dissertation; 5=thesis; 6=convention paper; 7=unpublished report; 8=other: Explain ________ ________)

6. Effect Size(s)

Description of ES Comparison(s)

ES# ___ ________________________________

ES# ___ ________________________________

ES# ___ ________________________________

ES# ___ ________________________________

	ES__	ES__	ES__	ES__	
12-13	_ _	_ _	_ _	_ _	a. N of ES's
14-15	_ _	_ _	_ _	_ _	b. ID of ES
16	_	_	_	_	c. Level (1=primary; 2=secondary)
17-18	_ _	_ _	_ _	_ _	d. Type of comparison (1=treatment vs. control; 2=treatment vs. placebo; 3=treatment A vs. B; 4=pre-post; Interaction treatment by: 5=gender, 6=age/grade, 7=testing, 8=personality, specify ________________, 9=other, specify ________________; Treatment within: 10=gender, female; 11=gender, male; 12=other, specify ________________)
19	_	_	_	_	*7. Target Population (0=not mentioned; 1=term used; 2=defined; 3=population described; 4=1&2; 5=1&3)
20	_	_	_	_	*8. Accessible Population (0=not mentioned; 1=term used; 2=defined; 3=described; 4=1&2; 5=1 & 3; 6=1, 2, & 3)
					9. Replication (0=No; 1=direct; 2=systematic)
21	_	_	_	_	a. of other research
22	_	_	_	_	b. within study
					B. Description of Sample(s)
					1. N:
23-27	_ _ _ _ _	_ _ _ _ _	_ _ _ _ _	_ _ _ _ _	a. of total sample
28-31	_ _ _ _	_ _ _ _	_ _ _ _	_ _ _ _	b. of experimental Ss
32-35	_ _ _ _	_ _ _ _	_ _ _ _	_ _ _ _	c. of control Ss

	ES__	ES__	ES__	ES__	
					2. Sample Selection (0=can't tell; 1=random; 2=solicited/volunteer; 3=captive/intact group; 4=random, group; 5=other, specify: ________)
36	_	_	_	_	Experimental
37	_	_	_	_	Control
					3. % Male
38-40	_ _ _	_ _ _	_ _ _	_ _ _	Total Sample
41-43	_ _ _	_ _ _	_ _ _	_ _ _	Experimental
44-46	_ _ _	_ _ _	_ _ _	_ _ _	Control
47	_	_	_	_	4. Treatment Context (0=can't tell; 1=elementary or secondary schooling; 2=college/university education; 3=adult education; 4=inservice; 5=work; 6=community; 7=recreation; 8=other, specify ____________________)
					5. Educational Level of Ss (0=can't tell; 1=preschool; 2=primary; 3=intermediate; 4=middle school; 5=junior high school; 6=senior high school; 7=combination, specify ____________________; 8=undergraduates; 9=graduate students; 10=post-professional; 11=adults not in school; 12=other, ____________________)
48-49	_ _	_ _	_ _	_ _	Experimental
50-51	_ _	_ _	_ _	_ _	Control
					6. University Students According to Major (0=not applicable; 1=can't tell; 2=elementary education; 3=secondary education; 4=elementary & secondary education; 5=education, unspecified; 6=nursing; 7=occupational & physical therapy; 8=philosophy; 9=psychology; 10=rehabilitation counseling; 11=social work; 12=sociology; 13=special education; 14=communicative disorders; 15=medicine; 16=other, specify ________________)
52-53	_ _	_ _	_ _	_ _	Experimental
54-55	_ _	_ _	_ _	_ _	Control

	ES__	ES__	ES__	ES__	
					7. Occupation of Nonstudent Ss (0=not applicable; 1=can't tell; 2=child care workers; 3=community recreation workers; 4=employees in institutions; 5=regular class teachers, elementary; 6=middle or juniorhigh; 7=high school; 8=school administrators; 9=special class teachers; 10=vocational rehabilitation counselors; 11=parents; 12=police; 13=medical; 14=general public; 15=other, specify ________)
56–57	__ __	__ __	__ __	__ __	Experimental
58–59	__ __	__ __	__ __	__ __	Control
					8. Prior Experiences with Disabled Persons (0=can't tell; 1=none; 2=as parents; 3=as siblings; 4=classmates; 5=as teachers; 6=in school; 7=as co-workers; 8=as supervisors; 9=in work setting; 10=as clients/patients; 11=general, not specific; 12=combination, specify ________; 13=other, specify ________)
60–61	__ __	__ __	__ __	__ __	Experimental
62–63	__ __	__ __	__ __	__ __	Control
64	__	__	__	__	9. Country of subjects (1=USA; 2=Canada; 3=Australia/New Zealand; 4=Europe, specify ________; 5=other, specify ________)
					C. Treatment/Intervention
65	__	__	__	__	1. Basis (0=can't tell; 1=theory, explicit; 2=prior research, no citations; 3=prior research, few citations; 4=prior research, case developed; 5=practical experience/insight; 6=other, specify ________)
					2. Attitude Change Theory
66	__	__	__	__	a. Theory (1=S-R/behaviora[illegible]nditioning; 3=congruity/equilibrium; 4=[illegible]ocial judgment; 5=functional; 6=combination, specify ________)
67	__	__	__	__	b. Relationship to treatment (1=mentioned but not used; 2=brief allusion; 3=explicit, well developed basis; 4=post hoc interpretation; 5=implicit). If mainstreaming study with no prior intervention to change attitudes, skip to C.10., and X-out Sections C.3-9.

	ES__	ES__	ES__	ES__	
					3. Setting (0=can't tell; 1=regular classroom; 2=special classroom; 3=home; 4=institution; 5=group home; 6=hospital; 7=dormitory; 8=playground; 9=camp; 10=recreation facility; 11=laboratory; 12=individual/small group; 13=normal life; 14=other, specify ______; 15=combination, specify ______)
68-69	__ __	__ __	__ __	__ __	Experimental
70-71	__ __	__ __	__ __	__ __	Control
					4. Treatment/Intervention Technique(s) (0=none/control; 1=placebo; 2=information; 3=direct contact; 4=vicarious experience; 5=positive reinforcement; 6=persuasive message; 7=persuasive messages, contrast; 8=2&3; 9=2&4; 10=other, specify ______)
72-73	__ __	__ __	__ __	__ __	Experimental
74-75	__ __	__ __	__ __	__ __	Control
					a. Information
					(1) Type (0=none; 1=etiology; 2=characteristics; 3=problems; 4=similarities with nondisabled; 5=prostheses and special equipment; 6=famous disabled people; 7=legal rights; 8=parenting/management; 9=self; 10=social relations; 11=other, specify ______; 12=combination, specify ______)
76-77	__ __	__ __	__ __	__ __	Experimental
78-79	__ __	__ __	__ __	__ __	Control

(End of Card #1)

* *

1-2	M A	M A	M A	M A	Project code
3-6	__ __ __ __	__ __ __ __	__ __ __ __	__ __ __ __	Report ID#
7-8	0 2	0 2	0 2	0 2	Card #

	ES__	ES__	ES__	ES__

(2) Delivery mode (0=none; 1=lecture; 2=discussion; 3=lecture/discussion; 4=print; 5=panel discussion-disabled; 6=panel discussion-nondisabled; 7=speaker-disabled; 8=speaker-nondisabled; 9=video or films; 10=pictures/photos/filmstrips; 11=case study; 12=audio; 13=simulations; 14=regular course, specify ________; 15=regular program, specify ________; 16=other, specify ________; 17=combination, specify ________)

9-10	__	__	__	__	Experimental
11-12	__	__	__	__	Control

b. Direct contact (0=none; 1=as companions; 2=as peer tutors; 3=in cooperative learning groups; 4=as classmates; 5=as classmates, behavior modified; 6=as students, behavior modified; 7=practice teaching-special classes; 8=classroom observation; 9=supervised playground activities; 10=in recreation programs; 11=panel discussions; 12=guest speaker(s); 13=visit to institution/residential facilities; 14=integrated community programs 15=as teacher/counselor; 16=as co-workers; 17=other, specify ________; 18=combination, specify ________)

13-14	__	__	__	__	Experimental
15-16	__	__	__	__	Control

c. Vicarious experience (0=none; 1=role play-contact; 2=role play-disabled; 3=simulation; 4=observation of role play or simulation; 5=videotapes or films; 6=case studies; 7=pictures/photos; 8=print fiction/biography; 9=dolls/puppets 10=other, specify ________; 11=combination, specify ________)

17-18	__	__	__	__	Experimental
19-20	__	__	__	__	Control

d. Positive reinforcement (0=none; 1=covert; 2=overt)

21	_	_	_	_	Experimental
22	_	_	_	_	Control

e. Persuasive message (0=none; 1=video/film; 2=audio; 3=print; 4=expert; 5=expert, disabled; 6=self-presentation;7=other,specify ________; 8=combination, specify ________)

23	_	_	_	_	Experimental
24	_	_	_	_	Control

	ES__	ES__	ES__	ES__	
25-26	__ __	__ __	__ __	__ __	5. Treatment/Intervention to Change Attitudes Toward: (1=disabled in general; 2=physically disabled in general; 3=retarded, general; 4=moderately retarded; 5=severely retarded; 6=mentally ill; 7=emotionally disturbed; 8=visually impaired/blind; 9=hearing impaired/deaf; 10=learning disabled; 11= speech/language impaired; 12=cerebral palsied; 13=epileptic; 14=para/quadriplegic; 15=other physically impaired, specify ______________; 16=autistic; 17=health impaired; 18=multiply handicapped; 19=other, specify ______________; 20=combination, specify ______________)
					6. Treatment/Intervention Conducted By: (0= can't tell; 1=experimenter; 2=project assistants; 3=regular staff; 4=combination, specify ______________; 5=other, specify ______________; 6=not applicable)
27	__	__	__	__	Experimental
28	__	__	__	__	Control
					7. Length of Treatment/Intervention
29	__	__	__	__	a. Information available (0=no; 1=yes)
					b. Days/week
30	__	__	__	__	Experimental
31	__	__	__	__	Control
					c. Minutes/day
32-34	__ __ __	__ __ __	__ __ __	__ __ __	Experimental
35-37	__ __ __	__ __ __	__ __ __	__ __ __	Control
					d. Number of weeks (If less than 1, insert x's)
38-40	__ __ __	__ __ __	__ __ __	__ __ __	Experimental
41-43	__ __ __	__ __ __	__ __ __	__ __ __	Control
					e. Total # of hours
44-48	__ __ __ . __	__ __ __ . __	__ __ __ . __	__ __ __ . __	Experimental
49-53	__ __ __ . __	__ __ __ . __	__ __ __ . __	__ __ __ . __	Control

	ES__	ES__	ES__	ES__	
					8. Verification of Treatment Implementation (0=none; 1=not necessary; 2=systematic observation; 3=nonstructured observation; 4=interviews with Ss; 5=questionnaires to Ss; 6=intervenor follow-up; 7=other, specify ________; 8=combination, specify ________)
54	_	_	_	_	Experimental
55	_	_	_	_	Control
					a. Reporting (0=not applicable; 1=data; 2=data and analysis presented; 3=assertion by author(s))
56	_	_	_	_	Experimental
57	_	_	_	_	Control
					b. Degree of implementation claimed (0=not applicable; 1=none; 2=some; 3=mostly; 4=complete.
58	_	_	_	_	Experimental
59	_	_	_	_	Control
					c. Basis for author's conclusion re implementation (0=not applicable; 1=can't tell; 2=author's judgment or inference; 3=statistical significance; 4=inspection of data; 5=combination, specify ________; 6=other, specify ________)
60	_	_	_	_	Experimental
61	_	_	_	_	Control
62	_	_	_	_	d. Actual implementation (1=complete; 2=mostly; 3=only in part)
63	_	_	_	_	e. Description of treatment adequate for replication (0=no; 1=somewhat; 2=yes)

No.	ES__	ES__	ES__	ES__	Item
					9. Treatment validity
64	_	_	_	_	a. Implementation (from 8d)
65	_	_	_	_	b. Hawthorne
66	_	_	_	_	c. John Henry
67	_	_	_	_	d. Treatment diffusion
68	_	_	_	_	e. Dissatisfaction/resentment
69	_	_	_	_	f. Novelty/disruption
70	_	_	_	_	g. Experimenter effect/expectations
71	_	_	_	_	h. Treatment/experimenter confounded
72	_	_	_	_	i. Test x treatment interaction
73	_	_	_	_	j. Multiple treatment interference
74	_	_	_	_	k. General Treatment Validity (1=excellent; 2=fair; 3=poor)
					10. Mainstreaming
75	_	_	_	_	a. Presence (0=no mainstreaming; 1=mainstreaming only; 2=pretreatment plus mainstreaming)
76	_	_	_	_	b. Type of study (0=can't tell; 1=planned; 2=post hoc)
77	_	_	_	_	c. Instruction in mainstreamed classes (0=can't tell; 1=standard group/class instruction; 2=cooperative learning; 3=individualized instruction; 4=peer tutoring; 5=combination, specify ____________ ______; 6=other, specify ______________________)
78	_	_	_	_	d. Special personnel support (0=can't tell; 1=none; 2=teacher preparation; 3=consultant help; 4=other, specify ______________________)
79	_	_	_	_	e. Special skills training for disabled students (0=can't tell; 1=none; 2=social; 3=academic; 4=both)

(Items f–j: 0=can't tell; 1=not a plausible threat; 2=minor problem; 3=substantial problem; 4=major problem; 5=not applicable. For 3, 4, & 5, write reason next to item.)

(End of Card #2)

* *

	ES__	ES__	ES__	ES__	
1-2	M A	M A	M A	M A	Project Code
3-6	_ _ _ _	_ _ _ _	_ _ _ _	_ _ _ _	Report ID#
7-8	0 3	0 3	0 3	0 3	Card #
9	_	_	_	_	f. Type of special skills training. (0=not applicable; 1=coaching; 2=modeling; 3=counseling; 4=direct reinforcement; 5=group contingencies; 6=diagnostic/prescriptive; 7=cognitive control; 8=combination, specify ____________; 9=other, specify ________)
10	_	_	_	_	g. Special instruction for nondisabled peers (0=can't tell; 1=none; 2=information: 3=vicarious experience; 4=reinforcement; 5=persuasive messages)
11	_	_	_	_	h. Parent education (0=can't tell; 1=none; 2=disabled children; 3=nondisabled children; 4=2&3)
12	_	_	_	_	i. Type of parent education (0=can't tell; 1=not applicable; 2=information; 3=vicarious experience; 4=reinforcement; 5=persuasive messages)
13-14	_ _	_ _	_ _	_ _	j. Disabled children in mainstreamed classes (0=can't tell; 1=mildly and moderately retarded; 2=emotionally disturbed; 3=visually impaired; 4=blind; 5=hearing impaired; 6=deaf; 7=communication disordered; 8=physically and health impaired; 9=learning disabled; 10=combination, specify ________; 11=others, specify ____________)
15-17	_ _ _	_ _ _	_ _ _	_ _ _	k. Number of minutes handicapped children spend daily in mainstreamed class
18	_	_	_	_	l. Number of days per week in mainstreamed class
19-22	_ _ _ _	_ _ _ _	_ _ _ _	_ _ _ _	m. Total minutes per week in mainstreamed class
23-25	_ _ _	_ _ _	_ _ _	_ _ _	n. Months in mainstreamed program when outcomes assessed
26	_	_	_	_	o. Outcome measured for (1=nondisabled peers; 2=teachers; 3=parents, disabled children; 4=parents, nondisabled children; 5=administrators; 6=combination, specify ____________)

	ES__	ES__	ES__	ES__	D. Dependent Measures
27	_	_	_	_	*1. "Attitude" defined (0=no; 1=affective; 2=cognitive; 3=behaviors; 4=1&2; 5=1&3; 6=1, 2, & 3)
28-29	_ _	_ _	_ _	_ _	*2. Number of dependent measures. List (by name of instrument and form, if possible, or general category): 1. ________ 4. ________ 2. ________ 5. ________ 3. ________ 6. ________
30	_	_	_	_	3. Use of common instrument (0=no; 1=ATDP; 2=OMI; 3=RGEPS; 4=ATHI; 5=ATBS; 6=MRAI; 7=MTAI, revised; 8=combination, specify ________)
31	_	_	_	_	4. Common instrument modified (0=not applicable; 1=no; 2=for different population of Ss; 3=for different disabilities; 4=other, specify ________)
32	_	_	_	_	5. Source of data (1=self-report; 2=opinion of other, teacher; 3=opinion of other, administrator; 4=opinion of other, specify ________; 5=observation; 6=nonproject request)
33-34	_ _	_ _	_ _	_ _	6. Type of assessment (1=interview, structured; 2=interview, nonstructured; 3=attitude questionnaire; 4=sociometric measure; 5=peer assessment; 6=social distance scale; 7=informal observation; 8=systematic observation; 9=semantic differential; 10=telephone or mail survey; 11=telephone or mail request; 12=Q-sort; 13=projective test, pictures; 14=sentence completion; 15=adjective checklist; 16=rankings; 17=other, specify ________)
35	_	_	_	_	7. Source of instrument (0=not applicable; 1=can't tell; 2=teacher-made; 3=project developed; 4=prior research; 5=other, specify ________)

	ES__	ES__	ES__	ES__	
36	_	_	_	_	8. Development by project (0=not applicable; 1=description not provided; 2=description not adequate; 3=adequate description)
					9. Reliability of scores
37	_	_	_	_	a. Mentioned? (0=no; 1=coefficient reported; 2=yes, but no coefficient)
38	_	_	_	_	b. Source (0=not applicable; 1=can't tell; 2=computed on sample; 3=reported from other research; 4=pilot study; 5=combination, specify ________)
39-40	_ _	_ _	_ _	_ _	c. Method (0=not applicable; 1=can't tell; 2=test-retest; 3=internal consistency; 4=alternate forms; 5=inter-observer--%; 6=inter-observer--r; 7=intra-observer--%; 8=intra-observer--r; 9=categorization reliability; 10=combination, specify ____________________)
41-42	_ _	_ _	_ _	_ _	d. Coefficient
43	_	_	_	_	e. Magnitude (0=not applicable; 1=.80-1.0; 2=.60-.79; 3=.0-.59)
					10. Validity of scores
44	_	_	_	_	a. Discussed (0=no; 1=moderately; 2=comprehensive)
45	_	_	_	_	b. Type (0=not applicable; 1=general; 2=face; 3=construct-discrimination or correlations; 4=construct--expert judgment; 5=concurrent; 6=combination, specify ____________________)
46	_	_	_	_	c. Source (0=not applicable; 1=not mentioned; 2=general reference to literature; 3=citation of research; 4=project data; 5=inference without data; 6=combination, specify ________________)
47	_	_	_	_	d. Reactivity of measure (1=low; 2=moderate; 3=high)
48	_	_	_	_	e. Adequacy of validity (1=low; 2=moderate; 3=high)

	ES__	ES__	ES__	ES__	
					11. Data collection
49	_	_	_	_	a. Type (0=can't tell; 1=regular staff, 2=researcher; 3=research assistant; 4=noncontact; 5=other, specify ______________)
50	_	_	_	_	b. Blinded collection (0=can't tell; 1=not applicable; 2=no; 3=partial—experimental or control, or pre or post only; 4=yes)
51	_	_	_	_	12. Blinded scoring (0=can't tell; 1=not applicable; 2=no; 3=partial—experimental or control, or pre or posttest only; 4=yes)
52	_	_	_	_	13. Time of posttest (0=can't tell; 1=immediate; 2=delayed; 3=follow-up)
53-56	_ _ _._	_ _ _._	_ _ _._	_ _ _._	14. Weeks after intervention to posttest
					E. Internal Validity
57	_	_	_	_	1. Design (1=pre-post, control; 2=posttest-only; 3=Solomon 4-group; 4=nonequivalent control group; 5=single subject; 6=pre-post, one-group; 7=static group; 8=other, specify ______________)
58	_	_	_	_	2. Assignment to groups (0=can't tell; 1=random; 2=match-random; 3=select or match from different group; 4=random assignment of intact groups; 5=convenience; 6=not applicable; 7=other, specify ______________)
					3. Threats
59	_	_	_	_	a. Treatment Validity (from C.9.k., p. 9)
60	_	_	_	_	b. Maturation
61	_	_	_	_	c. History
62	_	_	_	_	d. Testing (0=can't tell; 1=not a plausible threat; 2=minor threat; 3=substantial threat; 4=major problem. For 3 & 4, write reason next to item.)
63	_	_	_	_	e. Instrumentation
64	_	_	_	_	f. Statistical regression
65	_	_	_	_	g. Selection
66	_	_	_	_	h. Experimental mortality
67	_	_	_	_	i. General Internal Validity (1=high; 2=medium; 3=low)

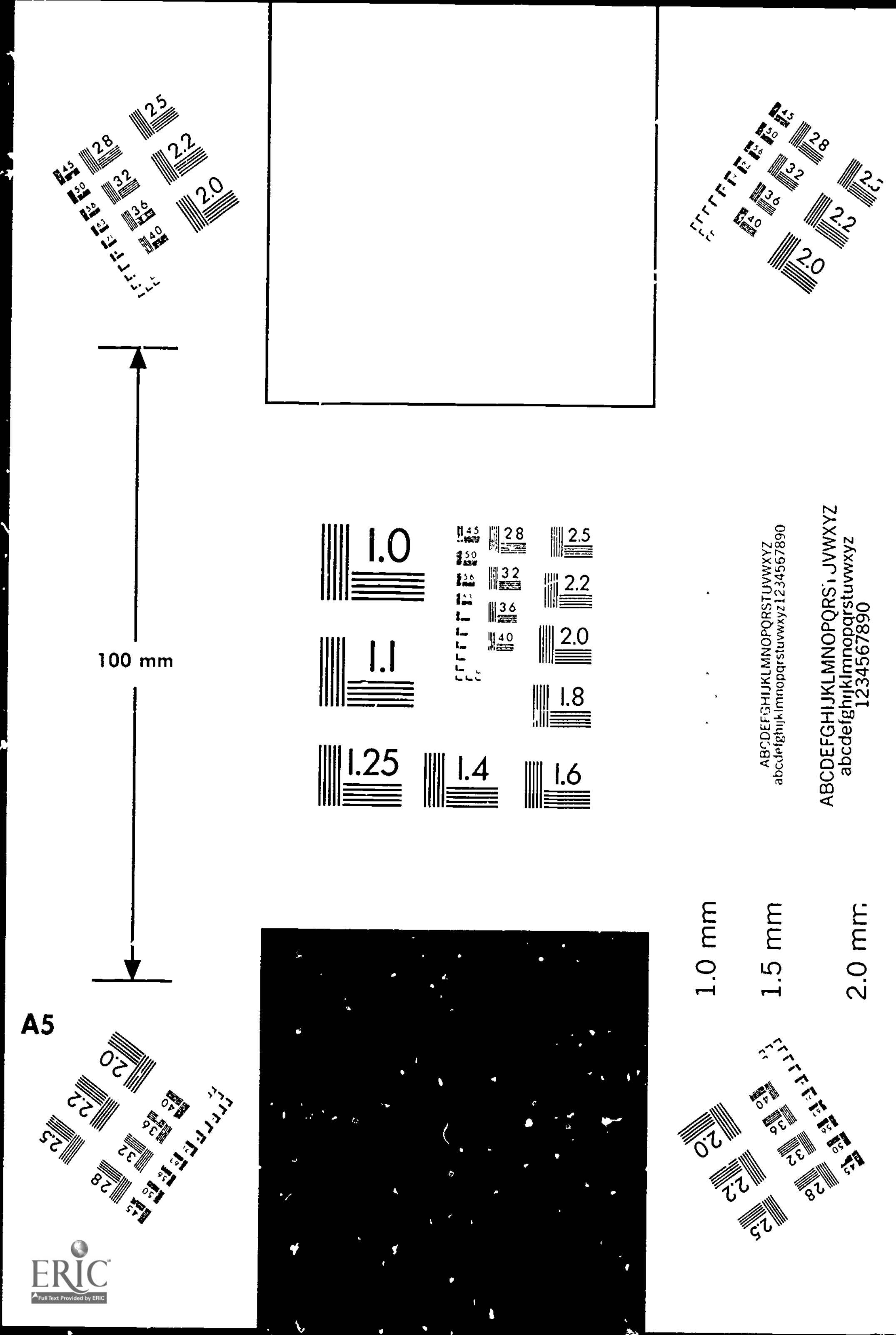
	ES__	ES__	ES__	ES__	F. Results
68	_	_	_	_	1. Statistical significance (0=not available; 1=not significant at .05 level; 2=significant at the .05 level)
					2. Author's conclusions about effectiveness
69	_	_	_	_	a. Qualified? (0=no; 1=sample; 2=interactions; 3=design; 4=measures; 5=treatment verification; 6=replication; 7=other, specify ________; 8=combination, specify ______)
70	_	_	_	_	b. Treatment effective? (0=none; 1=didn't have effect; 2=data equivocal; 3=produced effect; 4=produced negative effect)
					3. Effect size(s)
71	_	_	_	_	a. Available (0=no; 1=yes)
72–77	_ _ _ . _ _	_ _ _ . _ _	_ _ _ . _ _	_ _ _ . _ _	b. D. 1) ± D

(End of Card #3)

* *

1–2	M A	M A	M A	M A	Project code
3–6	_ _ _ _	_ _ _ _	_ _ _ _	_ _ _ _	Report ID#
7–8	0 4	0 4	0 4	0 4	Card #
9–10	_ _	_ _	_ _	_ _	2) Source (0=not applicable; 1=reported; 2=calculated; 3=t or ANOVA F; 4=correlated t; 5=correlated t (.50); 6=n-way ANOVA; 7=COVAR; 8=COVAR (.50); 9=proportions, chi-square; 10=other nonparametric; 11=r_{pb}; 12=significance level; 13=other, specify ______)
11	_	_	_	_	3) Scale of $\overline{X}$ difference (0=not applicable; 1=raw post; 2=raw gain; 3=covariance adjusted; 4=residual gain

	ES__	ES__	ES__	ES__	
12–13	__ __	__ __	__ __	__ __	4) Standard deviation (0=not applicable; 1=post control/placebo; 2=pretest; 3=1&2 pooled; 4=1-way ANOVA SSs; 5=n-way ANOVA SSs; 6=1-way ANCOVA SSs; 7=n-way ANCOVA SSs; 8=adjusted COVAR sd; 9=adjusted gain sd; 10=pooled post)
14–19	__ __ __ . __ __	__ __ __ . __ __	__ __ __ . __ __	__ __ __ . __ __	5) Mdn primary $\underline{D}$ for type of assessment
20–25	__ __ __ . __ __	__ __ __ . __ __	__ __ __ . __ __	__ __ __ . __ __	6) Mdn primary $\underline{D}$, overall
					c. Correlation
26	__	__	__	__	1) Type (0=none; $1=r_{pb}$; $2=E^2$; 3=Phi; 4=Cramer's V; 5=other, specify ________)
27–29	__ __ __	__ __ __	__ __ __	__ __ __	2) $\pm$ coefficient
30	__	__	__	__	3) Source (0=not applicable; 1=reported; 2=calculated; 3=estimated from $\underline{D}$)
31–33	__ __ __	__ __ __	__ __ __	__ __ __	4) Mdn primary coefficient for type of assessment
34–36	__ __ __	__ __ __	__ __ __	__ __ __	5) Mdn primary coefficient, overall
					d. Variance ratio
37–41	__ __ . __ __	__ __ . __ __	__ __ . __ __	__ __ . __ __	1) Ratio (S_E^2/S_C^2)
42–46	__ __ . __ __	__ __ . __ __	__ __ . __ __	__ __ . __ __	2) Mdn primary variance ratio, type of assessment
47–51	__ __ . __ __	__ __ . __ __	__ __ . __ __	__ __ . __ __	3) Mdn primary variance ratio, overall



100 mm
A5
1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz1234567890
1.0 mm
1.5 mm
2.0 mm

	ES__	ES__	ES__	ES__	
					G. Supplemental Information
					1. Information gain
52	_	_	_	_	a. Reported? (0=not applicable; 1=yes; 2=no)
53	_	_	_	_	b. Type of report (0=not applicable; 1=verbal; 2=statistical significance; 3=descriptive statistics; 4=2&3)
54	_	_	_	_	c. Conclusions (0=not applicable; 1=clear gain; 2=no gain; 3=mixed results; 4=can't tell/inconclusive)
55-60	_ _ _ _ . _ _	_ _ _ _ . _ _	_ _ _ _ . _ _	_ _ _ _ . _ _	d. Effect size—D
61	_	_	_	_	2. Type of study (0=can't tell; 1=course evaluation; 2=program evaluation; 3=experimental treatment; 4=mainstreaming, variation in time)
					H. Coding Summary
62-64	_ _ _	_ _ _	_ _ _	_ _ _	1. Minutes spent coding
65	_	_	_	_	2. Coder (1=Curtis; 2=Jesunathadas; 3=Shaver; 4=Strong)

RETURN TO PAGE 1 AND COMPLETE CHECKLIST

Report ID #: ______

______ Coder ______ Date

META-ANALYSIS: MODIFYING ATTITUDES TOWARD DISABLED PERSONS

EFFECT SIZES

(Author(s)/Year) Abbreviated Title and Source

Dependent Measures: ______

Groups: ______ ______

	Comparison (List primary first)	Dependent Measure	Primary or Secondary (circle one)
ES# 1			P S
ES# 2			P S
ES# 3			P S
ES# 4			P S
ES# 5			P S
ES# 6			P S
ES# 7			P S
ES# 8			P S
ES# 9			P S
ES# 10			P S
ES# 11			P S
ES# 12			P S
ES# 13			P S
ES# 14			P S
ES# 15			P S
ES# 16			P S
ES# 17			P S
ES# 18			P S
ES# 19			P S
ES# 20			P S

META-ANALYSIS: MODIFYING ATTITUDES TOWARD DISABLED PERSONS

INFORMATION REQUEST

Coder: ____________

Date: ____________

Report ID#: ____________

Reference: ______________________________ ______________________________

____

Author's address: ____

____

Dependent measure(s):

Information needed:

Coder Date

META-ANALYSIS: MODIFYING ATTITUDES
TOWARD DISABLED PERSONS

COMMENTS ON STUDY

(Brief abstract, if not on report; special strengths, weaknesses, significance; nuances of treatment, design, assessment; conclusions)

ID # Author(s)/Year Abbreviated Title & Source

Report ID# ______________

EFFECT SIZE COMPUTATIONS

ES__	ES__
ES__	ES__

Coder ______________ Date ________

META-ANALYSIS: MODIFYING ATTITUDES
TOWARD DISABLED PERSONS

COMMENTS ON CONVENTIONS

ID #	Author(s)/Year	Abbreviated Title & Source

Coder ____________ Date ________

META-ANALYSIS: MODIFYING ATTITUDES
TOWARD DISABLED PERSONS

REPORT DISPOSITION

____________ ID # ____________ Author(s)/Year ____________ Abbreviated Title & Source

_____ 1. Coding subsumed other reports.
ID # or #'s of other reports: ____________

_____ 2. Coding complete except for ES's.

_____ 3. ES's computed
Date

_____ 4. ES's checked
Date

_____ 5. Abstract available

_____ report

_____ comment sheet

_____ 6. COMMENTS ON CONVENTIONS sheet

_____ 7. COMMENTS ON STUDY sheet

_____ 8. Additional information needed from author(s)

_____ a. Form completed

_____ b. Request sent ________ Date

_____ c. Additional information received ________ Date

_____ d. Follow-up or thank you sent ________ Date

_____ 9. A vs. C, B vs. C Log Sheet

APPENDIX C*

CODING CONVENTIONS

(1) Conventions for Use of Coding Instrument

(2) Computation of Effect Sizes

(3) Conventions Addenda

---

*The pages for Appendix C are numbered in two ways: (1) consecutively with the rest of the report, and (2) based on the coding instrument categories and page numbers, as discussed in Chapter 3. The first number is (1); the number following the colon is (2).

META-ANALYSIS: MODIFYING ATTITUDES TOWARD DISABLED PERSONS

CONVENTIONS FOR USE OF CODING INSTRUMENT

General Instructions

1. Use of Numerical List. When you pick up a report to code, put your initials and the date next to the report ID# on the Numerical List of reports. When a report is used for a reliability check, write "Rel." and the date next to the ID# on the Numerical List. If the information to code one or more primary effect sizes is available, but information for one or more others has to be requested, indicate that on the Numerical List by writing "Info. Req." by your initials. When the information arrives and you are able to code the report, scratch out the "Info. Req." note. If no ES can be coded without additional information, have the report removed from the Numerical List and make a marginal note on the Alphabetical List that information has been requested. Put it back on the lists, with a new number, when the needed information arrives. The purpose of these procedures is to make the Numerical List the source for a quick check on the status of coding—both for general information and for deciding when reliability checks are appropriate.

2. EFFECT SIZES and INFORMATION REQUEST Sheets. The first step in coding a report, after skimming it to get a general sense of purpose and procedures, is to fill out an EFFECT SIZES sheet. If information is not available to compute the primary ESs, do not code the report. Complete an INFORMATION REQUEST sheet and give it to the secretary. Attach the report and the EFFECT SIZES sheet to the REQUEST sheet. Once a letter is written requesting the information, the report and ESs sheet will be put in the "Information

Requested" stack in Room 423. When the information arrives, retrieve the report and code it.

3. COMMENTS ON STUDY Sheets. Complete a COMMENTS ON STUDY sheet for every report you code, even if the only comment written on it is "No comments" or "Abstract in article". If the report does not have an abstract, write a brief one on this sheet. Be certain to write on this sheet any characteristics of the study that may be helpful or interesting in later organization or interpretation of analyses. Such things as gender balance if exact figures are not available to be coded, other design or sample characteristics, treatment characteristics, interactions of interest but not coded, authors' comments indicating special significance or interpretations of results, and good discussions of test validity are among the types of comments to record. Also, note the reasons for important coding decisions that may be questioned later--such as why a dependent measure was not used in an effect size.

4. EFFECT SIZE COMPUTATION Sheets. Be certain to record all major steps in your computations of D's, correlation coefficients, and variance ratios. If additional sheets are needed, enter the Report ID# on each and staple them to the ES COMPUTATION sheet for the study. Submit COMPUTATION sheets along with the Coding Instrument and other sheets to be filed once you complete the coding of the study.

5. COMMENTS ON CONVENTIONS Sheets. Whenever you have difficulty coding a study, note on a COMMENTS ON CONVENTIONS sheet the difficulty and how you resolved it. The difficulty might involve how to define groups for ESs, an instrument that doesn't fit well into any category in D.6. (p. 11), a study

which involved more than the 999.9 total hours provided for in C.7.e. (p. 7), difficulties in fitting delivery modes to the categories in C.4.a.(2) (p. 6), and so on. We will use the COMMENTS ON CONVENTION sheets both in interpreting analyses and in writing the final report.

6. Entering Numerals. Be sure that the numbers you write on the Coding Instrument can be read by the person keypunching. That means that they must be dark enough to be legible, and written carefully so that numerals will not be confused.

7. Coding Time. It is best to code a report in a single sitting if possible. Enter your start and stop time at the top of page 1 so that you can complete item G.1. at the end of the Coding Instrument. If you are not able to complete a report in one sitting, be sure to repeat your previous reading to the point where you have an adequate context for the coding. Also, enter the stop and start times for each of the sittings so that you can add up the times for category G.1.

8. Completing Spaces. Each code space must have something in it. Be sure to fill in all spaces, including leading zero's (eg, ID. #0024), pluses and minuses, and x's. Many of the categories in the Coding Instrument have a code for indicating "not applicable" or "can't tell." In cases where specific data are called for, such as the percentage of males in the experimental and control groups, and the data are not available, insert x's. Also, if an ES is coded for a pre-post, single-group design, or similar design, enter x's in any control group spaces, if a "Not Applicable" code is available.

9. Completing Checklist. Be sure to go over all items on the checklist after coding a study. A REPORT DISPOSITION sheet should be completed on every study and COMMENTS ON CONVENTIONS and COMMENTS ON STUDY sheets should be completed as appropriate. On the COMMENTS ON CONVENTIONS sheet, note any special difficulties in coding, any use of "other" or "combination" categories that raise questions about the adequacy of the other categories, and any decisions not covered by the Conventions that you had to make about category definitions in order to code the study. The COMMENTS ON STUDY (COS) sheets will provide a source of information about each study that will be more easily accessible than items on the completed Coding Instrument. If the report does not include an Abstract, write a brief abstract on the COS sheet. Also, note anything that makes the study particularly noteworthy, such as particular strengths or weakness, unusualness of approach or population. Also note any nuances of treatment, design, or assessment that may not be obvious or noticed on the Coding Instrument. In analyzing the data and writing up the results, it will be important, but difficult, to "keep in touch" with the many studies we will have coded. The COS sheets will be crucial both for alerting us to significant study characteristics and variations and for identifying studies to review again, and even re-code, to assist in our interpretations and discussions.

10. One Study/Multiple Reports. When multiple documents report analyses of the same data, they should be coded as a single study. One document may be adequately comprehensive to include others (e.g., a dissertation which encompasses one or more articles published from it, or a preliminary report encompassed by a final report). Assign an ID# to the main report and list the other documents with the same ID#, but with A, B, etc. added. Or, the

data in multiple reports may all come from one study, even though no one document is totally comprehensive; in that case, score the multiple reports as if they were one. Assign an ID# to one of the reports (the major one, if such can be identified; or, failing that, the one with the earliest date; or, finally, by alphabetizing authors' names, then titles). Use that ID# with an A, B, or C, etc. (ordering by authors' last name and then by date of publication) on the other reports. Use only the ID# for the coded or major document in the data columns of the coding sheet but be certain that the Coding Instrument and the REPORT DISPOSITION sheet have recorded on them the identification number of any subsumed document. On the EFFECT SIZES sheet and at the bottom of page one of the Coding Instrument, where the ESs are described, use the ID number and letters to indicate the articles from which the different effect sizes were taken.

11. Reading. Typically, you should skim each report before attempting to score it. Often you will find information in unexpected places. For example, you may find information describing the sample in the Conclusions rather than in the Sample section. In dissertations, the Acknowledgments can be a rich source of information—for example, in regard to whether persons other than the author carried out the treatment or gathered data.

12. Rounding Computations. In computing effect sizes (ESs), carry-out computations to the fourth decimal place and then round the ES to two decimal places. (You may also need to round in inserting other numerical data.) Round up if the number in the third decimal place is 6 or higher, if it is a 5 followed by a 6 or higher, or if it is a 5 followed by a 5 which is followed by a 6 or higher.

CODING INSTRUMENT

A. GENERAL INFORMATION

The Coding Instrument is set up with four columns for effect sizes. Some studies may have fewer than four effect sizes, some may have mor . When there are more than four effect sizes, use multiple Coding Instruments. The ES numbers should be filled in at the top of each sheet. Complete an EFFECT SIZES sheet before assigning ES numbers.

A.1. Project code. To identify IBM cards for the project, the code M A (meta-analysis) will be punched in each. The project code is typed on the Coding Instrument.

A.2. Report ID#. Write in the ID number of the study for each ES. Be certain to fill in all spaces--using zero's not x's. The ID number will be the same for every effect size for a study. For multiple documents scored as one, use one ID# (see prior item 8).

A.3. Card numbers. Card numbers will be typed on the code sheets.

A.4. Year of publication. It is assumed that all studies will have been published in the twentieth century, so insert the last two numbers of the publication year. Note that for dissertations, the date of the dissertation is the number to enter, not the date on which it was included in Dissertation Abstracts. If a report does not have a date on it, go to its list of references and find the most recent reference, add two years, and then enter that date. Note the asterisk (*4.). Whenever an asterisk is included for a category, the information is the same across all ESs.

A.5. Type of report. "Dissertation" refers to a doctoral study and "thesis" refers to a master's study. Unpublished report ("7") includes all unpublished reports other than convention papers ("6"), including government reports, reports that are part of occasional paper series, and mimeographed publications of a research center. If multiple reports are available for the same study (e.g., a dissertation and a journal article), code "9" and enter the types of reports used.

A.6. Effect Size(s).

Description of ES Comparison(s)

The number of effect sizes is a function of the number of dependent measures and the number of comparisons which are of interest. Complete an EFFECT SIZES sheet, providing a brief description of each effect size, including the dependent measure and the groups compared (see A.6.d. below). List primary effect sizes (see A.6.c., below) first, followed by secondary effect sizes. Then, enter in the left hand column of the Coding Instrument the ES# from the EFFECT SIZES sheet. These will be the numbers which you will place at the top of each page and which you will enter in A.6.b. ID of ES. Also enter the ES descriptions. Be certain to paper clip the EFFECT SIZES sheet to the Coding Instrument when you are done with the report.

A.6.a. N of ESs. The N of ESs is taken from the EFFECT SIZES sheet. Two N's will be entered: The N of primary ESs and of secondary ESs. That is the number of primary effect sizes will be entered in each primary ES column, and the number of secondary effect sizes will be entered in each secondary ES column. Even if there are Prior Contact secondary ESs (see A.6.d., #'s 13, 14, 15), enter the number of secondary ESs coded on the original Coding

Instrument (pp. 1-16) only. The number of Prior Contact secondary ESs will be coded on page 1 (I.9) of the Prior Contact Coding Sheet. For studies for which information is lacking to code ESs, but for which statistical significance results are available (coded on the ES Information Missing CODING INSTRUMENT), the N of ESs is the number you would code if ES data were available. If there are two sets of ESs for a report, one coded on a regular and one on an ES Information Missing CODING INSTRUMENT--i.e., ID# (*)--the N of ESs is entered separately for each set.

A.6.b. ID of ES. Fill in the numbers assigned on the EFFECT SIZES sheet. Note that although the N of ESs is figured separately for regular and Missing Information sets of ESs, ES ID numbers are to be assigned serially for all ESs, with the regular ESs numbered first. The keypunch operator will punch the ES number from item A.6.b. You should write ES numbers at the top of each page for your own information as you code.

A.6.c. Level. Two levels of ESs will be scored: Primary and secondary. A "primary" effect size is one which involves the comparison of a treatment group with a control or placebo group or the comparison of two treatment groups. If a delayed posttest is administered, treatment vs. control/placebo or Treatment A vs. B comparisons for it also yield primary ESs. For pre-experimental single-group designs, pre-posttest comparisons yield primary ESs.

"Secondary" effect sizes are of two types: Interactions, and treatment comparisons within levels of classification variables, i.e., within a subgroup. For example, in addition to comparing an experimental and a control group for the total sample, data may be available for a treatment-control comparison for males and for females. These data may be part of a table for which there is an interaction effect size, or may be presented in a table even though the interaction effect was not analyzed.

A.6.d. Type of comparison. A "treatment group" is one to which some treatment or intervention has been applied. In some studies, interventions to modify attitudes will be compared with each other, with no control or placebo group. This will yield "Treatment A vs. B" comparisons. If, however, Treatment A and Treatment B, or Treatment A_1, A_2, A_3 (i.e., the same treatment is applied to different groups), are each compared to a control or a placebo condition (C), code the A vs. C, B vs. C, or A_1 vs. C, A_2 vs. C, etc. ESs, but not A vs. B or A_1 vs. A_2, etc. Enter the study on the A vs. C, B vs. C Log Sheet for easy identification in case we later want to run A vs. B analyses. A "control group" is one to which no treatment has been applied, while a "placebo group" is one which receives a treatment intended to have no effect. In a study with more than one control or placebo group, unless the groups are clearly different, pool the means to compare with the treatment means in ESs. With a separate-sample, pretest-posttest, control group design (Campbell & Stanley, p. 54), code the design as "1" and code the pretest group as a control group throughout.

As you go through the coding sheet, you will find that data spaces are often labeled "experimental" and "control". In coding, experimental = treatment. Use the "control" space for placebos, too. For Treatment A vs. B comparisons, use the "experimental" space for what you have labeled on page 1 as "Treatment A" and the "control" space for what you have labeled, "Treatment B". If, in a comparison of two treatments, one can be expected to be more powerful than the other, label it "Treatment A". For example, if one group is provided information about disabled persons in an effort to change attitudes and another group in the same study is provided the information plus contact with disabled persons, the latter treatment would be assumed to be more powerful, with its combination of elements. It would be labeled Treatment A, and coded in the "experimental" group spaces. If no potential difference in treatment power can be discerned, the labeling of Treatment A's and B's (and C's and D's, etc., if necessary) is arbitrary. It is essential in any case that the same order be preserved in entering data throughout the coding sheet.

"Pre-post" effect sizes involving pre-posttest data from the same group ("4"), will be coded only for single-group, pre-post designs and will, for that pre-experimental design, be primary effect sizes.

Factorial designs yield secondary ES's to be coded. Four types of interactions are identified specifically on the Coding Instrument: Treatment by gender ("5"), treatment by age/grade ("6"), treatment by testing ("7"), and treatment by personality ("8"). In addition, there is a place to specify an interaction involving some other classification variable ("9"). "Treatment by personality" refers to an interaction between the treatment and levels of a personality measure of some sort. Specify the personality measure in the space provided. For interactions, Eta^2 will be the effect size. Typically, the only within-group secondary ESs to be coded (including for single-group, pre-post designs) are for gender ("10"; "11"). But if results are reported by grade (or age) level and differ across grade (or age) levels, code within-grade or within-age level ESs. In a factorial analysis in which the treatment by gender (sex) interaction (or treatment by grade interaction) is reported, within-gender (or within-grade or within-age level)

ESs are redundant; code only the interaction ES, but note on the COMMENTS ON STUDY sheet the pattern of treatment by gender or treatment by grade means. The "other" category ("12") would be coded for within-grade or within-age ESs, or for another such comparison of special interest.

Categories "13", "14" and "15" are used for Prior Contact ESs. Enter "13" if there is an analysis of prior contact levels within the treatment group; enter "14" if there is analysis of treatment effects within prior contact levels across groups (as treatment effects within levels of gender were coded); enter "15" if the interaction of prior contact and treatment is analyzed. If "15" can be coded, do not code "13" or "14". If "14" can be coded, do not code "13". If more than one definition of prior contact (e.g., number and frequency) are analyzed separately, or if quality of prior contact is analyzed separately, each is the basis for an ES.

A.7. Target Population. The target population is the universe or group to which the researcher would like to generalize. "Zero" is scored if there is no mention of such a group. "1" is used if the term " target population," is used in discussing the purpose of the study. "2" is scored if characteristics of the target population are mentioned explicitly—as, e.g., "The results of this study are meant to be generalizable to the attitudes of middle-class high school students toward persons who are blind." If the target population is actually described using data, such as from national tests or the census, "3" is to be scored. Combinations to be scored are indicated. Codable references to target populations will be found in the sections on Purpose and Methods. Despite what is said on item 10 above (p. 4) in regard to finding information in unexpected places, do not infer the target population from the Discussion or Conclusions sections, as the issue is whether the target population was identified prior to the study.

A.8. Accessible population. The accessible population is the group from which the researcher drew the sample or samples. It may be identical with the target population. Or, when the researcher uses all of an accessible group--e.g., all of the students in the fifth grade in a school district or all students in a college of education--the accessible population and sample are identical. If, however, the researcher uses an "available" subgroup (e.g., 50 students enrolled in two sections of a college course or two classrooms of fifth-grade students in a school or school district), that group is his/her sample, "drawn", however inappropiately, from a larger accessible population (e.g., all students enrolled in the course and similar course, all students in a college, all fifth-grade students in a school or school district). This distinction may be difficult to make. But the question to ask is, were other similar Ss available within the context from which the researcher selected the group or groups to use? Researchers usually consider the convenient (i.e., captive, intact) groups they use to be samples; rarely is it intended that the convenient sample encompasses the entire accessible population. In many cases, the sample will be defined or described, but not the accessible population; and, rarely are they identical unless the researcher says so specifically.

If the term "accessible population" is used to describe the group from which the sample was selected, code "1". If general characteristics of the accessible population are indicated, code "2". For example, the report might say, "Junior and senior students were selected from four urban Chicago high schools." If, however, the characteristics of the accessible population are described based on data available prior to or as part of the study, code "3". For example, the following would be a "3": "The sample was drawn from the senior class of nursing students at the University of Illinois. Made up of 60% females and 40% males, 30% had at least one year of experience working in a hospital." Use codes for combinations of 1, 2, and 3, as appropriate. Note that in many cases the researcher may describe the sample for the research but not mention accessible population. In that case, code "0".

9. Replication. Replication involves repetition of a study to determine if the findings will hold up under either the same or different conditions. "Direct replication" refers to a study conducted with the same population and treatment conditions as the prior study. "Systematic replication" involves planned variation in population and/or conditions. A replication may be a repetition of prior research (A.9.a.), or there may be a replication within a study (A.9.b.), as when a researcher carries out a study at different grade levels or in different schools or with subjects with different characteristics to determine if the results will hold up across the various conditions. A study may involve both replication of other research and replication within itself. If research is reported in which there are applications of the treatment to different populations of Ss, which should not be pooled to get ESs, code "3" in category A.9.b. within study, even if the researcher apparently had not intended to carry out a replication, and note what you have done on the Coding Instrument and on the COMMENTS ON CONVENTIONS sheet. If the chronological order of within study replications is clear, those applications of treatment following the first one are coded as replications; if the chronological order is not clear, code all applications as replications.

B. DESCRIPTION OF SAMPLE(S)

B.1. N: a. of total sample. The number to be inserted here for primary ESs is the total number of subjects included in the effect size data at the time of data analysis--that is, the number in the comparison with which the project ended, not started. If N is presented, and no other information is provided or inferable from analysis tables, assume that it is the N at time of data analysis. Sometimes the total N will not be given but can be inferred rather precisely from references to the number of groups and whether they are of equal size, or from the degrees of freedom in analysis tables. For secondary ESs, enter the N at data analysis for the Ss included in the particular ES. For example, if the total N was 100 with 50 females and 50 males, 100 would be entered for the primary ESs (if they involved all Ss) and 50 each for the treatment-within-gender, male-female secondary ESs. If the number of classrooms, but not the number of students, is given, use 25 Ss/class at the elementary level and 30 Ss/class at the secondary school level to estimate N and n's unless there is an indication that regular classes were not used.

B.1.b. of experimental Ss. This is the number of students in the analysis (see B.1.a. above) for the treatment group for the particular effect size being scored. In a pre-post, one-group design, enter the group N here and put x's in the control group spaces (B.1.c.). For interaction ESs which involve more than two treatment groups, enter x's both here and in the control group spaces.

B.1.c. of control Ss. Again, this is the number of subjects in the control group (or Treatment B or placebo group) for the analysis (see B.1.a. above) for the particular effect size being scored. For pre-post, one-group designs and interaction ESs, enter x's.

B.2. Sample selection. Many reports do not mention how the sample of subjects was obtained, and it may not be obvious from the discussion. In that case, code "0". "Random" sample selection refers to the use of a procedure which insures that each individual has an equal chance of being chosen from a population. In order to code "1" there must be explicit mention of such a process or use of the term "random" in reference to selection. In some studies, the researcher will solicit persons to be involved or will use persons who voluntarily come into a program. In that case, "2" should be coded. Research is often done with "captive/intact" groups, such as school classes or course sections; individuals do not volunteer, but are involved because of group membership. Then, "3" should be coded. If, however, a researcher uses school classes as the selection unit and then asks parents for permission for their children to participate, code "2". If groups, rather than individuals, are randomly selected, code "4", unless groups are the unit of analysis, then code "1". This category is to be coded separately for each effect size and for experimental and control groups, as the selection method is not always consistant across groups.

B.3. % male. As with Ns, enter for those Ss involved in the particular ES. If this information is not available, enter x's. Do not enter a decimal point but simply a three-digit number. For example, 4% would be entered as simply, 004, 90% as 090, and 100% as 100. Do not estimate percentages for treatment or control groups based on total sample information unless there is evidence, such as mention of random assignment, that the subgroups are constituted similarly to the total group. For interaction ESs, enter x's in the experimental and control spaces.

B.4. Treatment context. The point of this item is to code the general milieu of the research, such as whether it was in a public school context. The question is, "What was the treatment a part of?"

"Elementary or secondary schooling" ("1") includes preschool as well as K-12. "College/university education" ("2")includes undergraduate and graduate education in degree-type programs, "Adult education" ("3") refers to noncredit, noninservice type programs. "Inservice" ("4") is job-related, but not on-the-job education. "Work" ("5") refers to interventions conducted on actual work sites while people are working. An example would be providing information about disabled persons to MD's as a part of the staffing of cases. "Community" ("6") refers to studies conducted in general community settings, such as in a shopping mall, through newspaper articles, or through a broadcast over general television (i.e., not closed circuit television). "Recreation" ("7") refers to contexts such as parks, playgrounds, and overnight camps.

B.5. Educational level of Ss. "Preschool" ("1") refers to subjects younger than kindergarten age (age five). "Primary" ("2") includes grades K through 3. "Intermediate" ("3") includes grades 4 through 6. "Middle school" ("4") typically includes grades 6 through 8. While this category may overlap with "intermediate," code the Ss as referred to in the study. That is, if they are in the sixth grade, but referred to as middle school students, enter a "4", but if they are in the sixth grade in an elementary school, code "3". Medical students are graduate students, "9". The point of this categorization is to code educational level as relevant to the intervention. For example, with a program for practicing physicians or for teachers, Ss would be coded as post-professional ("10"). If the program is designed generally for adults who are not in an educational program and who vary in educational level, then "11" should be scored.

B.6. University students according to major. If any category other than "8" or "9" is scored in the item immediately above, code "0=not applicable" for this item. Otherwise, enter the code for the particular major or, if the major is not included, in the list, enter "15" and then write in the major.

B.7. Occupation of Non-student Ss. If the subjects in the study are students, then code "0" for this item. "Community recreation workers" refers to persons who work in such places as playgrounds and parks, including over-night recreational facilities--unless they clearly fall in some other category, such as teachers.

B.8. Prior experiences with disabled persons. If the report indicates that the Ss have had experience with disabled persons prior to the research, code the appropriate category. If the report indicates that disabled persons are in the same setting--e.g., in a special class in a school, or in jobs at a business--but it is not clear whether or how much contact the Ss have had with them, code "6" or "9"; as appropriate. Reports may indicate that the Ss have had prior experience with disabled persons, but not specify what that experience was. In that case, "11" should be coded. If some Ss have had experience with disabled persons and others haven't, code "12", specifying "1" and the appropriate code or codes for the experienced S's.

B.9. <u>Country of subjects</u>. The concern here is not where the research was conducted nor the country where the report was published, but with the country of the Ss. If, for example, a study was conducted in Germany at an American dependent school with the children of U.S. servicemen, then "1" would be coded.

C. TREATMENTS/INTERVENTION

C.1. Basis. Note that the concern here is with the rationale for the type treatment or intervention, not with the rationale for doing research on the topic. Sometimes a treatment or intervention is reported with no indication of how the treatment was arrived at. In that case, score "0". If the treatment is tied explicitly to an attitude change theory, score "1". It will often be helpful to check the titles of citations to see if they are to research reports or theoretical discussions. Also, use of one of the five category labels in Table 1 (pp. C-2 - C-3) is not necessary; they are for organizational purposes. Look rather for the theory names under "A. Theories" in Table 1. If the report simply refers to prior research as the basis for the intervention or the study, but without citations, score "2". If the study/intervention is based on prior research, but only one to three or so studies are cited loosely, then score "3". If the prior research is cited and discussed as part of a well-developed case for the research, score "4". If no research is cited, but the report refers to an experience in general or to experience in a particular school or other setting, score "5".

C.2.a. Attitude change theory. In the prior category, a "1" indicates that a theory was the explicit basis for the study/intervention. If that was the case, then code the theory in one of the five choices for this category. If the theory base was not explicit, then the theory to which the treatment seems most highly related should be coded.

Five kinds of information to use in classifying studies according to attitude change theory are presented in Table 1: (1) The names of attitude change theories which fall within each of the four categories; (2) the names

Table 1. Major Categories of Attitude Change Theories

1. Stimulus-Response & Behavioristic Theories	2. Conditioning Theories	3. Consistency/Equilibrium Theories	4. Social Judgement Theories	5. Functional Theories
A. Theories	A. Theories	A. Theories	A. Theories	A. Theories
. Hull-Spence Behavior Theory . Reinforcement Theory . Yale Communication Program	. Classical Conditioning . Skinnerian Theory . Operant Conditioning	. Affective-Cognitive Consistency Theory . Balance Theory . Belief-Congruence Theory . Cognitive Balancing . Cognitive-Dissonance Theory . Congruity Model or Theory . Consistency Theory . Dissonance Theory . Logical-Effective . Field Theory	. Assimilation-Contrast Theory	. Katz's Theory . Kelman's Theory . Psycnoanalytic Theory . Smith, Bruner, & White's Theory
B. Names	B. Names	B. Names	B. Names	B. Names
. Hovland, Janis, & Kelley . Weiss	. Bem . Doob . Staats . Scott	. Abelson . Brehm & Cohen . Cartwright & Harary . Feather . Festinger . Heider . Lecky . Lewin . McGuire . Newcomb . Osgood & Tannanbaum . Rokeach . Rosenberg & Kelley	. Helson . Sherif & Hovland	. Katz . Kelman . Smith, Bruner, & White . Sarnoff . Stotland . Adler
C. Terminology	C. Terminology	C. Terminology	C. Terminology	C. Terminology
. attention, comprehension, acceptance, retention . persuasive communication . source credibility . incentives . practice/mental rehearsal . effective excitatory potential . communication stimuli . communication effects	. overt stimulus, implicit response, overt behavior . higher order conditioning . reinforcement . cues	. assertions . attitude object . attractions . balance . cognitive elements . cognitive relations . cognitive unit . communication . conceptual arena . congruity . dissonance . equilibrium . object relevance . orientation . psycho-logic . sentiment relation Socratic effect . strain . symmetry . system of orientation . triad . unit relation	. anchors . assimilation . contrast . latitude of acceptance . latitude of rejection . level of adaption . level of noncommitment . reference points . reference scales	. compliance . ego-defensive function . externalization . identification . instrumental/utilitarian function . internalization . knowledge function . object appraisal . social adjustment . value-expressive function

Table 1. Major Categories of Attitude Change Theories (Continued)

1. Stimulus-Response & Behavioristic Theories	2. Conditioning Theories	3. Consistency/Equilibrium Theories	4. Social Judgement Theories	5. Functional Theories
D. Theory Aspects . Changing opinions results in attitude change . Opinions change when a subject has internalized a valuational message . Opinions change when the learner perceives arguments to be reasonable and logical . Variables affecting acceptance of a communication are (a) observable characteristics of the source, (b) the setting in which the communication is presented, and (c) the strength of the arguments or appeals	D. Theory Aspects . New attitudes are learned through systematic control of contingencies	D. Theory Aspects . Attitudes possess psychological structure: a change in the affective component will result in a change in the cognitive component and vice versa . Imbalance/strain/incongruity/dissonance result in tension which forces a change toward equilibrium . Inconsistencies are discovered through thought . Persons who perceive each other as similar should be attracted to each other . Interaction and proximity should result in positive attitudes	D. Theory Aspects . An opinion that is not too discrepant from that held by a person will be accepted . An opinion that is quite discrepant from that held by a person will be rejected	D. Theory Aspects . To change attitudes it is first necessary to know the function of the attitude to be changed . Attitudes change when they no longer serve their original functions
E. Strategies/Techniques . present subject with a persuasive communication (may include neutral material) . use experts or prestigious persons as source . use counter-attitudinal advocacy (e.g., reading a persuasive communication expressing opinion different from one's own with expression and conviction, reading and defending an opinion different from one's own under forced compliance)	E. Strategies/Techniques . use verbal reinforcement . use classical conditioning involving external stimuli	E. Strategies/Techniques . employ a person whom subject respects to present communication . associate person, event, or idea with person, event, or idea subject respects . present opinion as being similar to opinion held by subject . demonstrate to subject that attitude object is relevant to the attainment of certain values . assist subject to recognize conflicts within his/her value system . employ prestigious suggestions . employ counter-attitudinal acts (e.g., role playing, simulation) . present the disabled as being "normal" . provide contact with the disabled so that respect can develop . peer tutoring . cooperative learning	E. Strategies/Techniques . measure a person's point of view and then present a persuasive communication that falls just outside the point of view	E. Strategies/Techniques . provide new information and arguments to show that attitude is not useful . remove perceived threat associated with attitude . identify more appropriate values/attitudes . show that acceptance of attitude is not the best means of achieving social reward . show that acceptance of attitude is not the best way to achieve important values

of authors who are associated with the four categories of attitude change theory; (3) common terminology within each of the four attitude change theory categories; (4) a listing of primary aspects of theories that fall within each category; and (5) a list of the particular strategies and techniques for change that are likely to be used by persons operating from each theoretical position.

C.2.b. Relationship to treatment. In some cases, articles will mention a theory but not go on to draw any connection to the treatment or study. That should be coded "1". If the theory is discussed only briefly in a sentence or two as a basis for the treatment, code "2". If the treatment and study are well related to the theory, code "3". If the author refers to the theory only in interpreting the results obtained, code "4". If a theory is not mentioned and you must infer it, code "5".

If the treatment involves mainstreaming only ("mainstreaming" defined as a systematic, sustained effort to integrate disabled students in regular classrooms for part or all of their instruction, as contrasted with bringing disabled students into a regular classroom temporarily to provide contact as part of a research project), it will be coded in Section C.10. In that case, X-out the coding spaces for Sections C.3.-9. If, however, in preparation for mainstreaming, there is a separate, codable research effort to change the attitudes of, for example, students, teachers, administrators, or parents, code the ESs for that part of the report in Sections C.2.-9., and X-out the coding spaces for Sections C.2.-9. for the mainstreaming ESs.

C.3. Setting. The question here is, "Where, specifically, was the research conducted?" If the research was carried out in an educational context (numbers 1, 2, 3, and 4 in category B.4.) as part of regular instruction (number 1, 2, or 4 in category G.2.), code "1", unless it is clear that the students were not in a classroom for the treatment. Also, if the treatment is a course or part of a course and the experimental Ss are compared against control Ss in another course or class, code "1" for the control group, too. If, however, the control Ss are simply tested and are not part of a structured activity, such as a course, code "13". If the setting was a gymnasium used for regular instruction, code "1". "10=recreation facility" is to be coded for community recreation buildings. "4=institution" refers to a building or set of buildings housing disabled persons on a permanent basis. "5=group home" refers to a situation where a few disabled persons (usually 4 to 5) live in a home-like setting. A hospital, "6", is a facility where medical treatment is provided. If a permanant living facility for mentally retarded persons is referred to as a "hospital", code it as "4" nevertheless. A laboratory ("11") is a special facility for conducting research, not typically used for other purposes. "Individual/small group" ("12") refers to situations in which there is interaction between a small number of people, such as a counselor and parents, or a few parents in a discussion group. If such interaction is the case, code "12" even though the discussion may take place in, for example, a corner of a large classroom. "13=normal life" is used to code studies in which some experience or intervention is provided as the subjects go about their daily activities. For example, if students were asked to simulate disabilities as they went about campus or shopped in stores, or people were shown a film at a Lion's Club meeting, or as a short feature at a movie theater, code "13".

C.4. Treatment/intervention technique(s). The major categories for coding treatments are: "2", information, in which the technique is to provide information about disabled persons as a means of modifying attitudes; "3", direct contact, where subjects are put in personal contact with disabled persons; "4", vicarious experience, in which the technique is to create situations which will lead the subjects to put themselves in the place of handicapped persons and feel what it is like to be handicapped; "5", reinforcement, in which either classical or operant conditioning is used to modify behavior assumed to reflect attitudes; and, "6", persuasive message, in which a message designed with an argument intended to convince people as to what their attitudes toward disabled people should be is presented. If the purpose is not to present an argument, but to test the efficacy of presentation of information through a medium (e.g., a film), code "2". If the study investigated the relative effectiveness of different messages or of using different media to present a persuasive message, code "7=persuasive messages, contrast". Some interventions may be a combination of these techniques, and that should be coded. For combinations not encompassed in "8" or "9", code "10" ("other") and specify the combination. For studies in which the treatment is systematic desensitization (which might involve exposure to disabled persons in imagination or through direct contact), code "11". Typically in that case, "2" ("conditioning") will have been coded in C.2.a. Attitude Change Theory. When "11" is coded, there may be no coding in C.4.a-e. Code "a. Information" only if conveying information is intentional and not incidental. Do code "b. Direct contact" if it is part of the desensitization process. Code "d. Positive reinforcement" only if reinforcement is used; i.e., do not code in that category if only extinction is used.

Usually authors will indicate which technique they intended to use. In some cases, you will need to make a best judgment based on the information provided. The main point is to categorize studies by their intended intervention techniques. For example, if the intended technique was direct contact and the subjects were put in contact with disabled persons through a panel of speakers or through tutoring disabled peers, even though it is likely that they would have picked up information about disabled persons, "2" should not be coded here (although it will be in 4.a. below). On the other hand, if a research report states that the purpose was to convey information about, as well as to put the subjects in contact with, disabled persons, and a panel was used for that purpose, then "8", indicating a combined information-direct contact technique, should be coded.

C.4.a. Information. (1) Type. If the intervention did not provide information about disabled persons, then "0" should be coded. That will be the case for control and placebo groups, and for many studies scored in any treatment/intervention category other than "2" in C.4. However, if it is evident that a treatment designed with a different intent does include information about disabled persons, select the appropriate code.

"1=etiology" refers to information about the causes of disabilities. "2=characteristics" refers to information about the characteristics, including abilities as well as disabilities, aspirations and interests, vocational and social capabilities, and the kinds of activities, such as sports, in which disabled persons can engage. "3=problems" refers to information about the difficulties encountered by disabled persons, including learning problems encountered by learning disabled students. "4" refers to programs in which the characteristics of disabled persons are presented, with emphasis on the similarities with nondisabled persons. "7" refers to information about the legal rights of disabled persons. "8" ("parenting/management") refers to general information about being a parent or managing a classroom or business. "9" ("self") refers to providing the Ss with information about themselves, e.g., how their reactions compare to those of other parents or teachers. "10" ("social relations") refers to general information about how people relate to one another in group/social settings and/or how one might relate to disabled persons in such settings. For a regular course or program (see C.4.a. (2) convention) for which the content is not specified, code "12" and write in "can't tell".

C.4.a. (2) Delivery mode. "0=none" would be coded for control groups. "1", "lecture" refers to lecturing by a regular instructor, while "7" and "8" refer to the use of guest lecturers or speakers. "4", "print" refers to use of textbooks, research reports, other expository types of materials. Case studies presented in print are coded "11". If a film or other material with an identifiable title is used, for the treatment and/or placebo group, write the title here on the coding sheet and on a COMMENTS ON STUDY sheet so that the information can be retrieved easily later, if needed. Simulations may be used primarily to present information rather than to evoke vicarious experiences. If that intent is clear in the report, code "13". If the effects of a regular course (e.g., an abnormal psychology, introductory special education course, or practicum experience) or program (e.g., a master's program for rehabilitation counselors or a nurses' training program) with no special treatment (e.g., a specially selected film, special speakers) is investigated, code "14" or "15", respectively. Indicate what is reported about the mode or modes of delivery or write in "can't tell" if no specifics are given.

C.4.b. Direct contact. Each of the direct contact categories encompasses a means by which contact with disabled persons can be provided. "1" ("as companions") refers to time spent with a disabled person(s) without the structure implied by other categories, such as peer tutoring, the classroom, or supervised playground activities. "3" ("cooperative learning groups") refers to a specific teaching approach in which students work together in groups to attain common goals, minimizing competition among students for grades and other teacher rewards. "5" ("as classmates, behavior modified") refers to direct contact with disabled persons who have gone through a social skills or academic skills training program of some sort in order to make their behavior more acceptable to their nondisabled peers. That is true also for "6", except here the concern was with making disabled students more acceptable to their teachers. "0" is coded for control or placebo groups, or for studies in which direct contact is not an evident part of the treatment.

C.4.c. Vicarious experience. Again, code "0" for control groups, placebo groups, and studies in which vicarious experience is not an evident part of the intervention. "1" refers to situations in which the subjects are asked to role-play situations in which they have contact with disabled persons; "2" refers to situations in which the persons actually role-play being disabled. Role-playing involves, as the term "role" implies, interacting with other persons in an acting-out situation. "3", simulation, involves the use of devices to simulate a disability, such as using a wheelchair, having one's hearing dampened, wearing a blindfold. Projects may also attempt to evoke vicarious experiences through having subjects observe others who are role-playing or undergoing a simulation ("4"), through viewing video tapes or films ("5"), through reading case studies ("6"), through looking at pictures and/or photographs ("7"), through reading works of fiction or biography ("8"), or through playing with disabled dolls or watching plays put on with disabled puppets ("9"). If a technique to provide a vicarious experience cannot be reasonably categorized in the first eight categories, categorize it as "other" and describe it briefly. If the materials used to provide vicarious experiences have an identifiable title, write it here on the coding sheet and on a COMMENTS ON STUDY sheet for possible later retrieval.

C.4.d. Positive reinforcement. If the intent is to keep subjects unaware that they are being reinforced, then "1" should be scored. If there is no intent to conceal reinforcement—for example, students are given tokens based on the number of times or length of time they play with disabled students—then score "2". Use "0" for control groups and for studies with no obvious reinforcement.

C.4.e. Persuasive message. If use of a persuasive message was not part of the study, or if a control group is used in a persuasive message study, code "0". "Expert" ("4") refers to one who has credibility as knowledgeable about disabilities or some other area relevant to the treatment (e.g., attitude theories). "6" refers to the persuasive message strategy in which a subject is asked to present a message in a way so as to convince somebody else, with the intent that they will thereby persuade themselves. When the comparative effects of two different messages are studied, code "7".

C.5. Treatment/intervention to change attitudes toward. Some reports will identify specifically the disabled group of concern; that should be coded, using the "other" ("18") category only if necessary. In some studies, attitudes toward disabled persons in general, rather than toward a specific disability, will be the target and "1" should be coded. Or, attitudes toward the physically disabled in general, or toward those who are "mentally retarded", will be the target; and "2" or "3" should be coded, respectively. Educable mentally retarded persons are moderately retarded ("4") and trainable mentally retarded persons are severely retarded ("5"). Code "6" ("mentally ill") or "7" (emotionally disturbed") according to the terminology used in the report. "Emotionally disturbed" will usually be used in reference to students whose behavior creates classroom problems. Include "behaviorally disturbed" students in "7".

C.6. Treatment/intervention conducted by. Reference here is to the person or persons who carried out the actual treatment, including such activities as introducing the treatment, for example, a film or speaker, and participating in the treatment if it involves, e.g., presenting information or leading a discussion, as contrasted with a "nonperson" treatment such as a film. "1" ("experimenter") refers to a person who is a project director and responsible for the project. "2" refers to people--i.e., research assistants--hired to work on a research project. "3" refers to staff who are regularly employed in the setting where the research is carried out. For example, if regularly employed teachers carry out the treatment/intervention, then "3" should be scored. If the experimenter has a role in orienting Ss and someone else (e.g., a research assistant) conducts the treatment, code "4". In studies of naturally occurring experiences, such as the effects of contact as camp counselors, no treatment is actually "conducted", and "6", "not applicable", should be coded. Code "not applicable" ("6") also for control groups (where no treatment is applied).

C.7. <u>Length of treatment/intervention</u>. Following Category C.7.a., you are asked to enter information in regard to days/week, minutes/day, number of weeks, and total numbers of hours of treatment time. If information on treatment length is not available, enter x's. Length of treatment is potentially an important variable, so try to infer the length if it is not given specifically. For example, if it appears regular class periods were used, assume they were 45 minutes long. Assume the same for a treatment session, unless there is some counterindication. For regular programs (e.g., graduate programs, internships), it will often be possible to determine the number of weeks (e.g., a quarter is assumed to be 10 weeks; a semester, 15 weeks) but not the number of hours per day or perhaps even days per week. Enter the data on number of weeks and enter x's in the spaces for which information is not available.

The length of placebo treatments should be entered if available. For control groups, "0's" will be entered, as they will receive no treatment. When a reading assignment is part of a treatment, do <u>not</u> include that time in the length of treatment unless the time Ss spent reading is clear from the report. Note the reading on the COMMENTS ON STUDY Sheet.

C.7.a. <u>Information available</u>. Some reports will not provide sufficient information for you to determine the length of the treatment. In that case, code "0" and enter x's in b. through e. in the experimental and control spaces.

C.7.b. Number of days per week. Number of days per week does not refer to full days but to the number of days during which there is exposure to the treatment. Round to whole numbers. If, for example, students participate in a program for four and a half days, enter "5". Or, if the treatment is a film presented once (i.e., on one day), enter "1". Note that a control group should be receiving no treatment, so 0's would be entered, unless those spaces are being used for a Treatment B or placebo group.

C.7.c. Minutes/day. Enter here, if available, the number of minutes per day spent in the treatment. If treatment length is discussed in terms of class periods or group presentations, assume a 45-minute period. For overnight camps, use 12 hours is the estimate of contact time; and for day camps, use 6 hours. Estimate length, if possible. If length cannot be estimated, enter x's. With a true control group, enter zeroes if information is available for the experimental group. If information is not available for the experimental group, also enter x's for the control group.

C.7.d. Number of weeks. Enter only whole numbers. If information is given in months, assume 4.3 weeks per month. If the treatment is less than one week insert x's. Assume five days to the week and round as appropriate when treatment does not encompass a full five days but is more than one week. Follow the same convention as in 7.c. for entering information for the control group.

C.7.e. Total # of hours. Multiply the # of actual days by the # of minutes per day and divide by 60 to obtain the total # of hours of treatment. Round to one decimal place. If there are more than 999.9 hours in the treatment, enter 999.9 and note the actual hours on the COMMENTS ON STUDY sheet.

C.8. Verification of treatment implementation. Here the question is whether the researcher made an effort to verify whether the treatment was implemented as intended, including whether the control group actually received no relevant intervention during the treatment as intended. That is, was evidence gathered as to whether the treatment was carried out with fidelity? If no such verification is mentioned, "0" should be coded. On rare instances, it may seem that no verification is necessary because of the way the independent variable is defined, and "1" should be coded. Systematic observation ("2") involves use of a category system to score behavior to determine whether planned behaviors occurred. "3", nonstructured observation, involves more casual observation in which, for example, the researcher or an assistant might simply observe an intervention and make a judgment as to whether it was implemented properly. Subjects might be interviewed ("4") or be given a questionnaire ("5") in order to determine whether they perceived intended variations in treatments as one check on whether those variations occurred. Also, the researcher might debrief the persons who carried out the intervention to determine whether they thought that they had implemented it successfully ("4=intervener follow-up). If a combination of verification techniques is used, score "7" and enter the numbers for the separate techniques.

C.8.a. Reporting. If no verification was attempted, code "0". "1" should be scored if data are presented and simply referred to with no analysis. "2" should be scored if data are presented and some sort of analysis--e.g., the statistical significance of differences between proportions of behaviors for treatments--is reported. If the author or authors simply assert that they gathered data that indicated verification, or that their unstructured observations indicated implementation, then score "3".

C.8.b. Degree of implementation claimed. If "0" or "1" was coded in C.8., code "0" here. Probably no author would indicate that he or she failed to implement the treatment; or, at least, no such article would get published. "1=none", is to be scored if, despite an apparent effort at verification treatment, no claim is made in regard to implementation. It is also not likely that "2=some" will be coded frequently.

C.8.c. Basis for author's conclusion re implementation. If "0" or "1" was coded in C.8.b., code "0" here. Sometimes authors will claim that the treatment was carried out, but give no basis for that conclusion. In that case code "1". If the author simply refers to a judgment made, for example, based on observations, code "2". If statistically significant differences between treatments--for example, in terms of intervener behavior--are reported, code "3". If the author simply reports data and claims that inspection of the data confirms verification, code "4".

C.8.d. Actual implementation. Here you must make a judgment, based on your best reading of the report, as to how well the treatment was implemented.

C.8.e. Description of treatment adequate for replication. The point of concern he is not whether the treatment could possibly be carried out by someone else interested in changing attitudes toward disabled persons, but whether another researcher interested in repeating the study would have enough information about the treatment to repeat it with fidelity. The question to ask is, if I were to attempt to repeat this study, how confident would I be that I could carry out the treatment as it was conducted the first time?

C.9. Treatment validity. Even if a treatment is implemented with fidelity, the subjects may not experience the treatment, intervention, or conditions as intended by the researcher. The subcategories here get at various dimensions of treatment validity, of which fidelity of implementation is one. The question is, if the treatment had an effect, or failed to have an effect, was the outcome likely due to the treatment as intended or due to associated, but unintended, aspects of the treatment?

If secondary ESs involve within-gender treatment comparisons, use the coding for the treatment main effect ESs unless there is reason to think validity differed by gender. For interaction secondary ESs, make an overall judgment of validity for each category in C.9.

Many times authors of research reports will not provide the information necessary to determine whether treatment validity is threatened. For example, often no evidence will be reported to verify that the treatment was presented as intended. Or, the author will not address the question whether the Hawthorne effect or some other effect might have been an unintended part of the treatment or intervention. If there is a lack of information as to whether a threat was present, code "0=can't tell". At the same time, you should use your common sense and knowledge of intervention settings to decide whether a threat is plausible. Cases may arise where a category of treatment validity does not apply. For example, in coding a pre-post, single-group study, "c. John Henry" would not be relevant. In such cases, code "5".

C.9.a. Implementation. Enter here the number which you coded in C.8.d., Actual Implementation, above.

C.9.b. Hawthorne. The Hawthorne effect refers to influences on the behavior of subjects simply because they are aware that they are partic'pating in an experiment. Such an effect might occur, for example, when students in a psychology class m'st volunteer to participate in an experiment for class credit, when it is announced to subjects that they are part of a research project, or when Ss are brought into a laboratory for special treatment. On the other hand, if there is no reason to think that students would be aware that they are part of an experiment--for example, a film is presented by their teacher as a regular part of course instruction--then code "1", indicating that there was no plausible threat.

C.9.c. John Henry. The John Henry effect occurs when members of the control group recognize that a treatment group is getting special treatment, and "work harder" in order to show that they can perform as well without the special treatment. It is probably not a common factor in attitude research, as contrasted with research where some sort of achievement is assessed as an outcome. For pre-post, single-group designs, enter a "5", as "can't tell" is not pertinent and "1" would be a positive design indicator.

C.9.d. Treatment diffusion. Sometimes in educational and psychological research the subjects in different treatment groups will communicate with each other, so that a control or placebo group becomes aware of and/or knowledgeable about the treatment, thereby "washing out" the planned differences between the groups. Code "5" for a pre-post, one-group design.

C.9.e. Dissatisfaction/resentment. Groups of subjects who perceive that they are receiving a less desirable treatment, a more demanding treatment, or an overly demanding treatment may become dissatisfied and/or resentful, adding a characteristic to the independent variable not planned by the researcher. Disruption by the treatment that causes resentment would be scored here; disruption that makes the treatment difficult for the intervenor to handle is coded in C.9.f.

C.9.f. Novelty/disruption. If an intervention is new and exciting to the students, or if it disrupts daily routines, that may be an unplanned treatment characteristic. However, if novelty is a planned aspect of a short-term intervention--for example, that it is unusual to be around disabled persons--then it is not a threat to the treatment validity. Particularly if the treatment is a one-shot affair, novelty may be an important planned characteristic. However, if the treatment is intended to be used over a long period of time (e.g., as a regular part of a class), then novelty/disruption becomes a concern in answering the basic question for this category, Would the effects hold up if the subjects were exposed to the treatment for a longer period of time than reported in the study?

C.9.g. Experimenter effect/expectations. Here the question is whether as a part of presentation of the intervention, experimental expectations were conveyed in such a way that those expectations become part of the treatment, affecting the results. Experimenter effects will usually be a concern when the researcher conducts the treatment. The concern will be especially great when there are no checks on the way in which the treatment is presented, on how the experimenter is perceived, or on whether expectations are projected to the subjects (such as when the introduction of stimuli, such as films, might be affected by the researcher's knowledge that he or she is presenting to a treatment or placebo group).

C.9.h. Treatment/experimenter confounded. When an intervention is carried out by only one person or very few persons, the treatment is confounded with the personal characteristics of the intervener--for example, the enthusiasm or lack of enthusiasm of the intervener. This category is related to experimenter effects. However, that category, as used here, refers to conveying expectations based on knowledge of or desires for the research. The concern here is that any intervener, experimenter or not, brings personal attributes to the intervention situation, regardless of their knowledge. Unless these specific characteristics are "averaged out" over a number of interveners, the treatment effects cannot be separated from the personal characteristics of the intervener.

C.9.i. Test by treatment interaction. Part of the Ss experience with the treatment may be due to the fact that they were sensitized by the content or methodology of a pretest or a posttest. In that case, the testing experience becomes a part of a treatment not anticipated by the researcher, and a threat to treatment validity. This category is related to the later one, D.10.d. Reactivity of measure, and should be scored in that context. That is, the more reactive the test, the more likely it is that there was test-by-treatment interaction.

C.9.j. Multiple treatment interference. There are two types of multiple treatment interference. One occurs when, within a research project, the subjects are exposed to more than one treatment; the other occurs when subjects have participated in one or more prior research projects which affect their reactions to the treatment in the reported project. In both cases, then, the treatment experienced by the Ss is not what the researcher intended.

C.9.k. General Treatment Validity. Given the conception of a treatment (the treatment construct) which the researcher had in mind, what was the general validity of the treatment as carried out? Make your judgment based on your coding of categories C.9.a. through C.9.j. above. Although it is difficult to specify a particular numerical formula for combining those ratings to arrive at a General Treatment Validity score, an excellent rating ("1") would mean that most of the sub-categories received "1's" (not a plausible threat) with not more than one or two "can't tell" ("0") ratings; if any minor problems ("2") or substantial problems ("3") are indicated, the construct validity should be rated "fair" ("2"); if there are more than two or three substantial problems coded and/or one or more major problems ("4"), then "poor" ("3") should be coded.

C.10. MAINSTREAMING. As indicated in C.1., mainstreaming as a treatment/intervention technique, in which disabled students are integrated into regular classrooms for one or more periods a day (see also C.2.b., p. 18), is coded separately in this section. If research on the use of other techniques to modify attitudes toward disabled persons prior to mainstreaming is reported, those treatments should be coded in the prior sections, and the Es columns X'd-out in C.10.

C.10.a. Presence. The purpose of this category is to be able to sort out easily all mainstreaming studies, whether scored in the prior sections or not. "1=mainstreaming only" refers to studies not coded in the prior sections. If an investigation of changing attitudes prior to mainstreaming was coded, code "2" here and X-out the coding categories in Section C.10 for those ESs. If research into the effects of mainstreaming was not part of the study, code "0" and X-out Sections C.10.b.-C.10.o.

C.10.b. Type of study. Sometimes when mainstreaming is implemented in a school or school district, a research study will be planned as part of the implementation. In that case, code "1=planned". In some cases, once the mainstreaming is in place, someone decides that it would be interesting to gather information on the attitudes of mainstreamed vs. non-mainstreamed children, even though it had not been planned prior to the mainstreaming. In that case, code "2=post hoc". In terms of design, post hoc studies will typically be scored as static group designs in the Internal Validity section (E.1., p. 13). Remember, if a post hoc mainstreaming study is correlational--that is, it does not involve comparisons between groups of students categorized as having been involved in mainstreamed vs. nonmainstreamed programs, but instead attempts to obtain indicators of amount of contact and correlates amount of contact with attitudes--then it does not fall within the defined population of studies for this meta-analysis. If you cannot tell from the report whether the study was pre-planned or post hoc, code "0=can't tell".

C.10.c. Instruction in mainstream classes. If you cannot tell what type of classroom instruction was used, code "0". (It is assumed that there was instruction in an integrated classroom or the study would not be coded under "mainstreaming".) Sometimes mainstreaming will occur with no particular adaptations in the typical group-class instruction. If so, code "1". If there is no indication of any special modification in classroom instruction, "1" should be coded. If classroom instruction is conducted using what is called "cooperative learning", in which students work in groups and do not compete with one another for grades and other teacher rewards, code "2". If individualized instruction or peer tutoring is used, code "3" or "4", respectively. "5" should be coded for any combinations of the above instructional approached. With a "5", be sure to specify what the combination is, writing in the numbers, including "6" if an "other" instructional approach is used.

C.10.d. Special personnel support. Often, as part of mainstreaming, teachers will be given special support: Some sort of preparatory workshop is provided to acquaint them with the characteristics of disabled students and how best to teach them (code, "2"); or, consultants are provided, such as special education specialists, to help teachers cope with any problems which might arise (code, "3"). Usually, any special support will be mentioned. If it is not, code "1=none", unless there is some ambiguity in the report, making "0=can't tell" applicable. Remember that this category refers to special support. If, for example, disabled students continue to be in a special education classroom or resource room during the day as they would have been without mainstreaming, that is an on-going program component, not special support.

C.10.e. Special skills training for disabled students. In order to facilitate interactions between disabled students and nondisabled students in a mainstreaming program to, hopefully, thereby improve the attitudes of others toward the disabled students, special social or academic skills training may be provided to the disabled students as part of the mainstreaming. If one or the other or both types of skill training are reported, code "2", "3", or "4", as appropriate. Academic skills training will typically be a regular part of special education; code it here only if something special has been added to enhance attitude change on the part of nondisabled students in the mainstreamed setting. If such training is not mentioned, score "1=none". If there is some indication that such training might have occurred, but you cannot tell for sure, code "0=can't tell".

C.10.f. Type of special skills training. If a study was coded in C.10.d. as "1", not providing training in social, academic, or both, types of skills, or as "0", then code "0" here. Otherwise, code the type of skills training that was provided. Coaching ("1") involves providing instructions to disabled students on how to behave while they are interacting or following interaction with other persons in a real or a role-playing situation. Modeling ("2") involves having the disabled student watch other persons, either in person or via a medium such as a film, perform appropriate behaviors. Counseling ("3") involves one-to-one or small group sessions in which feelings and appropiate reactions are discussed. The use of reinforcement to make behavior more appropriate might be applied either to the individual disabled student ("4") or through the manipulation of group contingencies ("5") in which, for example, the whole group would be rewarded only if all individual members behaved appropriately. Diagnostic/prescriptive training ("6") refers to a process which is particularly likely to be applied to academic skills. It involves careful diagnosis of the disabled student's learning difficulties, with a prescription for specific remediation or action. Because diagnostic/prescriptive teaching is often a regular part of special education that disabled students are likely to be receiving anyhow if they spend part of their day in a special classroom or resource room, be careful to code "6" only if it has been added as special training for the purpose of enhancing mainstreaming. Cognitive control ("7") refers to a technique to help students gain control of their own behavior, for example, by repeating to themselves the steps for carrying out a procedure prior to trying it. Sometimes self-control of reinforcement is included.

C.10.g. Special instruction for nondisabled peers. As part of mainstreaming programs, special interventions may be provided for nondisabled peers. That should be scored here. Note that the categories are identical to categories 2 through 6 in C.4. Treatment/intervention Techniques (p. 5.). If the special instruction for or intervention with nondisabled peers is provided prior to, and in that sense is separate from, the mainstreaming, it should be scored as a separate treatment in the prior sections. If no special training is mentioned, code "1". Code "0" if there is some indication that special training may have occurred but you cannot tell for sure.

C.10.h. Parent education. Sometimes as part of mainstreaming programs, interventions will be carried out with parents to enhance their understanding of mainstreaming and/or disabled children. If only the parents of disabled children are involved in such a program, code "2". If only the parents of nondisabled children are involved, code "3". If both types of parents are involved, code "4". Unless parent education is mentioned explicitly, code "1"; code "0" if there is some indication of training, but it is not certain.

C.10.i. Type of parent education. If "0" is coded in C.10.h., code "1=not applicable". Otherwise, code using the prior definitions for treatment categories.

C.10.j. Handicapped children in mainstreaming classes. Here code the type or types of disabilities of the mainstreamed disabled children.

C.10.k. Number of minutes handicapped children spend daily in mainstreamed class. If no information is available about the amount of time spent in mainstreaming classes each day, enter x's. If the amount of time is reported in terms of class periods, use an estimated 45 minutes per period to compute the number of minutes per day.

C.10.1. Number of days per week in mainstreaming class period. Enter days, rounding to a whole number as necessary. Unless there is some indication to the contrary, assume 5 days per week. Enter x's if no information is available.

C.10.m. Total minutes per week in mainstreaming class. Multiply the number of minutes in C.10.k. times the number of days in 10.1. to obtain the figure for this item, unless the number of minutes per week is stated specifically in the report. Enter x's if the information is not available.

C.10.n. Months in mainstreamed program when outcomes assessed. Here the question is, how long had the students been involved in the mainstreaming program when attitudes toward disabled persons were assessed? The exposure during the school year is of interest. Assume a 9-month school year. So, for example, if the report indicates that mainstreaming had been in effect for three years, multiply 3 x 9 = 27 to get the total number of months-- unless the report includes a more precise indicator of the months. If the number of weeks is reported, assume 4.3 weeks per month.

C.10.o. Outcome measured for. Indicate here the Ss for whom attitudes toward disabled persons were assessed.

D. DEPENDENT MEASURES

Recall (Statement of General Purpose and Populations) that to be included in this meta-analysis, a study must include the assessment of attitudes toward disabled persons as an outcome measure. The assessment may be direct or implicit, with attitudes indicated, for example, by choice of playmates, willingness to associate with disabled persons, or the nature of interactions with disabled persons.

If a measure assesses only attitudes toward children or other persons generally, rather than toward disabled persons specifically, the study does not qualify. (A general scale of attitudes toward children would be acceptable if used to assess attitudes specifically toward disabled children --as indicated, e.g., by instructions to Ss--even if disabilities are not mentioned in the items.) By the same token, measures which assess attitudes toward mainstreaming or some other such intervention strategy do not qualify, because scores on such assessments are not clear-cut indicators of attitudes towards disabled persons. That is, a person may feel positively toward disabled persons and yet believe that mainstreaming, especially under certain circumstances, is not appropriate.

Also, behavioral measures must be clearly related to attitudes toward the disabled. A measure does not qualify if, for example, interactions between disabled and nondisabled S's are scored or reported so that only the quantity of general interaction can be ascertained, with no indication of whether the interactions were positive or negative; or, if the behaviors of disabled and nondisabled students are not analyzed separately; or, if interactions of nondisabled students with disabled and other nondisabled students are not separated for analysis.

D.1. "Attitude" defined. If the report does not contain an explicit definition of "attitude" in its introductory sections or in a Data or Instruments type of section, code "0". Attitudes may be defined affectively ("1")--that is, in terms of feelings toward a referent; cognitively ("2")--that is, in terms in beliefs about, a referent, although rarely will this be the only component of a definition; or, in behavioral terms ("3")--that is, as a manifestation of affect and cognition in approaching or avoiding a referent. Most commonly, a combination of affect, cognitive, and behavioral elements will be used, if any definition is given at all.

D.2. Number of dependent measures. Enter (separately for regular and ES Information Missing CODING INSTRUMENTS) the number of dependent measures that are appropriate for the meta-analysis (which may be different from the total number of dependent measures for which findings are reported). The interest is in the number of dependent measures for which ESs can be computed, not the number of types of measures. For example, if a semantic differential is used with two concepts (e.g., Mental Retardation and Physically Disabled), there are two (2) dependent measures. If the specific names of instruments are given, list the names. If the form of an instrument is reported, list that, too. When you list the names of specific instruments, capitalize and underline, to indicate titles. Also, list names and forms, if available, on the COMMENTS ON STUDY sheet. If names are not available, list the general category of the dependent measure, such as systematic observation of behavior or teacher-made attitude scale.

D.3. Use of common instrument. Listed are acronyms for certain instruments which are likely to be used frequently in the research being reviewed for this meta-analysis. If none was used in the study, code "0". The ATDP ("1") is the Attitudes Towards Disabled Persons Scale. The OMI ("2") is the Opinions about Mental Illness Scale. The ATHI ("4") is the Attitudes Towards Handicapped Individuals Scale. The ATBS ("5") is the Attitude to Blindness Scale. The MRAI ("6") is the Mental Retardation Attitude Inventory. The MTAI, revised ("7"), is the Minnesota Teacher Attitude Inventory. The MTAI will be appropriate for this meta-analysis only if it has been revised so that it assesses attitudes toward disabled persons in particular, rather than toward students in general.

A measure commonly used in attitude studies is "3", the Rucker-Gable Educational Programming Scale (RGEPS). Respondents are presented with brief descriptions of children who exhibit behaviors one would expect of mentally retarded, emotionally disturbed, or learning disabled children and asked to select the most appropriate educational placement for the child, ranging from regular classroom placement to full-time special class placement. The placement selection is considered to be a measure of the degree of social distance the teacher prefers to maintain between himself or herself and such students (Horne, 1985, p. 53-54). As indicated in the introductory paragraph to this section, such placement decisions may be affected by other factors than attitudes toward the disabled. Nevertheless, we will include the Rucker-Gable Scale as a dependent measure for this review when researchers say they are using it to assess attitudes toward disabled persons. It should be coded in D.10.e. as having low validity. However, if a report indicates that the RGEPS was <u>used to assess attitudes toward mainstreaming</u>, or is unclear as to its use, it does <u>not qualify</u> as a dependent measure for this review.

The CMI (Custodial Mental Illness Inventory), which assesses attitudes toward custodial care of the mentally ill, with some items on mental patients, does not qualify as a dependent measure for this review. The AAQ and the CAQ (Client Attitude Questionnaire) are also measures of attitudes toward working with the mentally ill (and, for the CAQ, toward etiology), not of attitudes toward the mentally ill.

D.4. Common instrument modified. On occasion, researchers will modify an instrument listed in D.3. for use in a particular study, either to be more appropriate for a particular population of subjects ("2") or to assess attitudes toward different disabilities than those mentioned in the original instrument ("3"). If some other modification is made, score ("4") and indicate the type of charge. If modifications are made, they will typically be mentioned; so if the report makes no mention of modifications, code "1". If "0" was coded in D.3., code "0".

D.5. Source of data. For this item code the source of the data for the particular instrument used to obtain the specific effect size. If the data were obtained from the individuals themselves as, for example, with an attitude survey or a sentence completion test, score "1". In some studies, attitude change is gauged by asking people to assess the attitudes or attitude change of other individuals. For example, a teacher may be asked to assess the attitudes or attitude change of students in his or her classroom. Categories "2", "3", and "4" are for coding that source of data. Observation ("5") includes not only scoring behavior as it occurs, but the scoring of transcripts of discussions or the scoring of written material, such as essays, with a coding system. A "non-project request" ("6"), involves obtaining responses or behaviors in a situation removed from the treatment, to which it is hoped that the Ss will see no connection. A classic example is Rokeach's use of responses on his Value Survey to create value dissonance in regard to racial discrimination on the part of university students, then later having mailed to them a solicitation to join NAACP, with joining or not the dependent measure.

D.6. Type of assessment. A number of types of instrumentation are listed, most of which are fairly common. Peer assessment, "5", refers to having individuals provide some sort of an assessment or evaluation of other persons as an indicator of attitudes towards those assessed. For example, nondisabled Ss might be asked to list the positive or negative attributes of students, with the number listed the dependent measure. On the other hand, an adjective check list, "15", provides students with a list of adjectives and they are asked to check those which apply to certain labels, such as "mentally retarded". "8=systematic observation" includes the use of a set of categories to score transcripts of discussions or other interactions, as well as live behavior, and the coding of writing, such as essays, produced by the Ss. For instruments that assess behavioral intentions (e.g., intent or willingness to invite a disabled person home, to volunteer to work with disabled persons), code "17" and write in "behavioral intention to . . .", finishing with the intended behavior.

D.7. Source of instrument. Of interest in coding this item is where the researcher obtained the instrument. In some cases, no "instrument" is involved in assessment, as, when following a film or some other treatment, students are asked to volunteer to assist disabled persons in some way and the percentage of volunteering is recorded. In that case "0=not applicable" should be coded. In some cases, no mention of a source is made, and "1" should be coded. Instruments developed by teachers, individually or with little or no specialized assistance, are to be coded "2", while an instrument developed as part of a particular research project would be coded "3". If one of the "common instruments" (D.3.) was used, or if the instrument was developed or used in prior research, code "4". When a semantic differential is used, with reference to Osgood, Suci, and Tennenbaum or a similar source, using Osgood et al.'s adjective pairs with terms (such as "mental retardation" or "mental patient") selected for the project, code "4". Only code "3" if the researcher developed his/her own set of adjective pairs. If an instrument was modified for use in the study being coded, code "5" and specify "instrument modified".

D.8. Development by project. If "2", "3", or "5" ("instrument modified") is coded in D.7., then the adequacy of description of the development process should be indicated by scoring "1", "2" or "3". "Adequate description" is defined as sufficient information so that you could replicate the development process as a researcher.

D.9. Reliability of scores. a. Mentioned? If there is no mention in the research report of the reliability of the scores for the dependent measure in an effect size, code "0". (Although "test reliability" is technically incorrect, code reference to it as score reliability.) If reliability is mentioned, but no coefficient is reported--e.g., a statement is made that "Reliability was found to be adequate" or "Researchers have generally found the instrument to possess adequate reliability"—code "2". Code "1" only if one or more coefficients are reported for the specific dependent measure.

D.9.b. Source. If "0" or "2" was coded in D.9.a., code "0" here. If a coefficient or coefficients are reported, indicate whether they were computed using data from the sample for the particular research study ("1") or reported from other research ("2"), or a combination of the two.

D.9.c. Method. Of concern here is the method used to estimate the reliability coefficient reported in D.9.d. If no reliability coefficient was reported, code "0". If a reliability coefficient is reported, but you cannot tell what type it is, score "1". Corrected split-half, Kuder-Richardson, and alpha coefficients are all internal consistency methods ("3"). Note that both inter-observer agreement ("5" and "6")--i.e., between observer--and intra-observer agreement ("7" and "8")--scores obtained by the same observer--are included. "Categorization reliability" ("9") should be coded when the reliability of observational scores is reported, rather than simply reporting inter-observer or intra-observer agreement.

D.9.d. Coefficient. If a reliability coefficient or percentage of agreement is reported for a dependent measure, insert it here without the decimal point. A reliability co-efficient of .80 would be entered, 80. If coefficients obtained with more than one method are reported, choose one to enter using the following order of preference: (1) Test-retest, (2) internal consistency, (3) alternate forms. "4=inter-observer--%" and "6=intra-observer--%" refer to percentage of agreement in categorizations. "5=inter-observer--r" and "7=intra- observer--r" refer to coefficients for the correlation between categorizations. If both percentage of agreement and a correlation coefficient are available for inter-observer or intra-observer agreement, enter the correlation coefficient. If both inter- and intra observer agreement are available, enter the inter-observer agreement information. If inter- and/or intra-observer agreement are reported along with categorization reliability, record the categorization reliability figure.

If no coefficient is reported, enter x's. If more than one coefficient is reported for the preferred method or in a general reference where method is not indicated (e.g., "Researchers have reported reliability coefficients ranging from .60 to .80"), enter the median of the coefficients reported. That is, in the above example, 70 would be recorded. If coefficients are reported for two forms of a test, one of which is used for a pretest, or for pretest and posttest data for the same test, record the posttest coefficient.

D.9.e. Magnitude. If no coefficient was reported, code "0". Otherwise, enter the number which indicates the range within which the reliability coefficient for the particular ES fell.

D.10. Validity of scores. a. Discussed. If there is no discussion of score validity (or of "test validity") in the report, code "0"; but if validity is mentioned and discussed, even if somewhat superficially, score "1". If a comprehensive discussion of validity is presented--usually encompassing mention and perhaps justification of the type of validity, and presentation of validity evidence from prior studies and/or the current study, code "2=comprehensively". Mentioning a factor analysis is not sufficient alone; the author(s) must indicate specifically that they see the relevance to validity.

D.10.b. Type. Code "0" if "0" was coded in 10.a. Code "1", "general" when validity is discussed but with no indication of what type of validity the author or authors of the report believe is involved. Face validity, "2", refers to a claim in a report that validity is obvious due to the nature of the items or the assessment. If construct validity is alluded to in terms of either discrimination or correlation—i.e., that the means of distinct groups are different as predicted, or that the scores correlate with other non-attitude scores as predicted or yield factors as expected—or if reference is made only to "construct validity", code "3". If the construct validity is estimated by having experts judge the items and/or the total test, code "4". If the validity of the scores is estimated through their correlation with attitude scores obtained on another attitude instrument or assessment procedure, code "5=concurrent". If a combination of types of validity is discussed, code "6", and enter the numbers for the types.

D.10.c. Source. If a "0" has been coded for 10.a., and 10.b., code "0" here. If, however, validity is discussed, but the source of the validity evidence or judgment is not mentioned, code "1". If the discussion of validity is based on general references to the literature without citations of one or more specific validity studies, code "2". If one or more validity studies are cited, code "3". If data from the project relative to validity are presented, code "4". If validity is inferred from inspection of the tests, but without collection of any sort of data and without reference to the literature either generally or specifically, code "5".

D.10.d. Reactivity of measure. The question here is the extent to which it is obvious that the assessment is related to desired treatment outcomes, with a likely effect on Ss' responses to the test (affecting the validity of scores) and to the the treatment (affecting treatment, and external, validity). A scale such as the ATDP or the ATHI so obviously assesses attitudes toward the disab_ d, with the most likely implicit assumption being that positive attitudes are good, that it should be coded "3=high". On the othe: end of the scale, when nondisabled Ss' interactions with disabled Ss are observed <u>without the Ss' knowledge</u>, reactivity is likely to be absent and "1" should be coded. An attitude scale in which items about disabled persons are embedded within a number of other items would be scored "2=moderate"; so would observations where the observers were visible to the Ss, but had observed long enough, and with sufficient unobtrusiveness, that the Ss probably had become acclimated to their presence.

D.10.e. <u>Adequacy of validity</u>. Based on the evidence presented in the article, indicate <u>how valid you believe</u> the dependent measure is for assessing attitudes toward disabled persons in this particular study. You will have to weigh the evidence for validity presented in the report, as well as the information given about the test and its relation to the definition of attitudes for the study, if presented. For example, if the ATDP, an assessment of general attitudes toward the disabled, is used as an assessment instrument in a study aimed at changing attitudes toward persons with a particular disabilility, code low ("1"); even though if used appropriately in a study aimed at modifying attitudes toward disabled persons in general, the ATDP would be coded as having moderate ("2") validity. Remember that the RGEPS is always scored as having low ("1") validity.

D.11. <u>Data collection</u>. a. <u>Type</u>. If you cannot discern from the report how the data were collected, code "0". If data were collected by regular staff of a program within which the research was conducted--e.g., teachers at a school where the research project was carried out—code "1". If the project director/author of the report collected the data himself or herself, code "2". If the data were collected by a paid assistant, code "3". If the data were collected without contact between project personnel and the respondent—e.g., through the mail—code "4". Data collected over the telephone would be coded according to the person who did the telephoning.

D.11.b. Blinded collection. The question here is whether the person or persons collecting the data knew the treatment groups which Ss were in and whether it was a pre or post assessment. "Not applicable" ("1") would be scored only if there were no human interaction involved in the collection, such as data collected via a mail survey or collected automatically as part of a computerized testing program. Code, "no" ("2") for group paper-and-pencil tests and for interview or observational data unless it is clear that blinded collection was done; researchers may overlook the importance of blinded collection of paper-and-pencil data, but are aware of the need for blinded observation and interviewing, and are likely to report it. Code "3" in the situation where partial information was kept from the collectors about the group membership of the subjects or whether a pre or posttest was being administered. Code "4=yes" only if it is clear that blinding was present. Often the lack of blinding will be clear from the way in which the study is presented—for example, if the person who introduced a treatment film also administered the tests following the film.

D.12. Blinded scoring. The same considerations and definitions apply here as for "blinded collection". The "not applicable" ("1") code should be used for attitude scales for which scoring involves no judgment, only the addition of pre-assigned item values.

D.13. <u>Time of Posttest</u>. If you cannot tell from the report how long after end of treatment posttest was administered, code "0". In most studies, a posttest will be administered very soon—immediately, within a day, or at the next class session—following the treatment. In that case, code "1". If the administration of the posttest is delayed--perhaps to obscure the relations between the treatment and the test--code "2". If the first posttest, or another form of it, is administered as a follow-up, code "3". If more that one follow-up administration of a test is reported, code "3" for the ES from each follow-up testing.

D.14. <u>Weeks after intervention to posttest</u>. If you coded "0" in D.13, enter x's here. If you coded "1", "2", or "3", the report should contain the time between the end of treatment and posttesting or you should be able to estimate the time. For same-day testing, enter zeroes. Otherwise, enter the time in weeks, rounded to one decimal place (1 day = .1; 2 days = .3; 3 days = .4; 4 days = .6; 5 days = .7; 6 days = .9). That is, a 2-day time period would be entered as 00.3 weeks. If you coded "3", enter the time in weeks between the end of treatment and the follow-up testing.

E. INTERNAL VALIDITY

E.1. Design. The first three categories ("1", "2", "3") are the Campbell and Stanley experimental group designs. Each assumes random assignment of the units of analysis (whether individual subjects or groups) to the experimental conditions. For a design with random assignment in which a posttest for an experimental group is compared with a pretest used as a simulated control group posttest, code "2". The only one of Campbell and Stanley's quasi-experimental designs included here is "4", the nonequivalent control group design. (If a nonequivalent control group design is used without pretesting, code "8" and write in "4, posttest only".)

It is unlikely that the literature being reviewed will contain many reports of single subject research ("5")—research in which behavioral data are collected on individuals, commonly with a baseline, followed either by a treatment and withdrawal of treatment to see whether the behavior returns to baseline; or using a multiple baseline design in which the treatment is introduced to different individuals at different times, without withdrawal to baseline, to determine if changes in behavior occur with introduction of the treatment as predicted.

Two pre-experimental designs are included: "6", the pre-post, single-group design, in which there is no control group and one group is administered a pretest, given a treatment, and then administered a posttest; and "7", the static group design, in which the researcher identifies already existent groups, one of which has received a treatment of interest and the other one of which has not, and administers a posttest to each to examine the effects of the treatment. This is a design which may turn up in reports of research on mainstreaming.

If "8=other" is coded, enter the Campbell and Stanley design name and number if possible.

E.2. Assignment to groups. The preferred method of assigning subjects to experimental groups is to assign them randomly. If a report states that subjects (or groups, if that was the unit of analysis) were assigned randomly or describes a process of assignment which was random, including random assignment from strata (e.g., gender), code "1". If Ss were matched on relevant variables and then assigned randomly to groups, code "2". Random assignment is a desired attribute, and researchers will be eager to report that they used it. So, code "1" or "2" only if it is obvious from the report that a random process was used. In some cases, researchers will obtain a treatment group and then select control subjects randomly from another larger group or select from a larger group subjects who are matched with individuals in the treatment group on some variable or variables. In either case, code "3". "4" would be coded when, in a nonequivalent control group design, the researcher randomly assigned intact groups but did not use group as the unit of analysis. "5", convenience, refers to what is actually nonassignment--that is, simply using available groups as they are constituted. The assignment is really done by someone other than the researcher. The typical case is the use of one classroom for a treatment group and another for a control group, with no control over who is in which group. In a design such as the pre-experimental, pre-post, one-group design, there is no assignment to groups, and "6" should be coded.

E.3. Threats. The scale used here to code threats to internal validity is the same as that used for coding Treatment Validity (C.9., p. 9). In cases where the design is generally assumed to control a threat and there is no reason to suspect that the threat is not controlled, code "1=not a plausible threat". For example, random assignment is an excellent control for testing--that is, if groups have been randomly assigned and both are administered the pretest, the effect of taking the pretest should be consistent for both and the interest is in whether there is a treatment effect above and beyond the testing effect on both groups. Maturation is also well controlled by random assignment, as are statistical regresson and selection. However, if inspection of the data—for example, pretest data—indicates that, even with random assignment, the groups were initially different on important variables, selection, and by extension, maturation and statistical regression (really selection by maturation interaction or selection by regression interaction), are also plausible threats. When random assignment is not used, selection is a particularly plausible threat.

If there is no evidence, including your knowledge of research settings, as to whether a threat is present or controlled, code "can't tell" ("0"). In some cases, a substantial ("3") threat or major ("4") threat may be present—for example, if convenience groups were used, in the absence of evidence that the groups were similar (and even without evidence that they were different), a substantial selection threat ("3") should be coded; if the groups were not randomly assigned and there are differences between them that might be related to treatment outcomes, or if scores on dependent measures were based on observer inference and the observers were not blind to the treatment conditions and/or to whether or not it was a pretest or posttest condition, 4" should be coded for selection and instrumentation, respectively.

E.3.a. Treatment validity. Whether Ss experience the treatment as intended is an important spect of internal validity. Enter the code from D.9.b. here. For a mainstreaming study for which category D.9.b. is not coded, enter an x.

E.3.b. Maturation. Maturation refers to "biological or psychological processes which systematically vary with the passage of time" (Campbell and Stanley, p. 8), independent of the treatment. Included is not only physiological or cognitive changes due to growing older, but having subjects grow hungry, tired, or bored. Maturation is a strong threat in single-group designs; with treatment and placebo/control groups, the concern is whether there was differential maturation.

E.3.c. History. Here the concern is with events other than the treatment that occurred between the pretest and posttest and might have had a differential effect on results. Of concern is extra-session history (that is, events that occur outside of the treatment essions) and intra-session history (that is, events that occur within treatment sessions, such as distractions from the noise of the lawn-mower, for only one group).

E.3.d. Testing. Simply taking a test may improve Ss' performace when they take the test again (including the taking of an alternate form). Sensitization, as well as practice, is a part of this effect. That is, if the purpose of a test is obvious, as on an attitude scale, people may be more likely to change behavior (positively or negatively) on a second testing as a result of the first testing than if purpose of the test is not obvious to them. Testing is a strong threat ("4") with a pre-post, single-group design; with treatment-control designs, the question is if there is a differential effect between the experimental and control groups.

E.3.e. Instrumentation. "Instrument decay" is another term used to refer to differences or changes in instrumentation which might account for an apparent treatment outcome. Included would be differences in test administration (e.g., a test administrator who is more positive and encouraging with the treatment group that with the control group) or data collection, especially where observers code behaviors or comments and know they are observing different treatment groups or know which is the pre- or the posttest. When there is no information to indicate that the administrator of a paper-and-pencil test is blind to Ss' group memberships, code a "2". Lack of blind testing with observations or interviews would merit at least a "3", and perhaps a "4" depending on the circumstances.

E.3.f. Statistical regression. On retesting, when a less than fully reliable test is used, the means of groups will tend to regress toward the mean of the population from which they came. This effect will be greater for extreme groups. Consequently, when one group initially has a more extreme mean than another group, the regression to the mean will be greater and might be confused with a treatment effect.

E.3.g. Selection. When groups have different characteristics at the beginning of the research study that might account for treatment outcomes, this threat to internal validity is referred to as "selection". Random assignment is an excellent control for this threat, but pretreatment differences can still be large by chance. If there is evidence, such as pre-treatment data, to indicate that the groups did not turn out, by chance, to be quite different on pertinent variables, or if there was random assignment from strata of a relevant variable, code "1" with random assignment.

Otherwise, with random assignment, code "0", unless there is evidence of pretreatment differences; then, "2", "3", or "4" should be coded. Whenever convenience samples are used, selection is likely to be a threat, with a judgment about the degree dependent upon the information that is provided or implicit about the groups.

E.3.h. Experimental mortality. When subjects are lost or drop out differentially from experimental groups, outcomes may be due to this differential loss of subjects rather than to the treatment. Although the basic question is whether the characteristics of those who dropped out or were otherwise lost from different groups were such as to affect outcomes, differential rates of loss or drop-out suggest that treatment effects might be a result of loss of subjects who differed on variables related to outcomes. Even though the author does not mention mortality, you may be able to pick it up through observing discrepencies between the n's reported and the d.f. in reports of analyses.

E.3.i. General internal validity. Based on your coding of the seven threats to internal validity and your general evaluation of the design, assign a General Internal Validity rating. Perhaps the best way to define high ("1"), medium ("2"), and low ("3") validity is by specifying the extremes. A study should receive a "1" if there is evidence that an excellent design was well-executed. In that case, there should be no more than one or two "0" ratings and/or no more than one or two "2" ratings. On the other end of the scale, low ("3") would be coded for a study which had one or more "4" ratings, or two or more "3" ratings, three or more "2" ratings, or four or more "0" ratings. In addition, a design with one "3" rating with two or three "2" ratings should be coded "3". Medium internal validity ("2") would then be coded for a study which had no "4" rating, at the most one "3" rating, fewer than two "2" ratings, and fewer than three "0" ratings.

F. RESULTS

F.1. Statistical significance. Statistical significance will only be coded, if available, in terms of whether results were or were not significant at the .05 level.

F.2. Author's conclusions about effectiveness. a. Qualified? Here the question is whether the author limited his or her interpretation of the results based on the characteristics of the study or the need to replicate the results. Authors may limit their conclusions by reference to the restricted nature of their sample ("1"), by reference to possible interactions between the treatment and personological or ecological population characteristics ("2"), features of the design ("3"), limitations in the methods used to assess outcomes ("4"), the lack or inadequacy of efforts to verify that the treatment was actually implemented ("5"), the need to replicate the study to determine whether the results are reliable or generalizable ("6"), or some other limitation ("7") or combination of limitations ("8"). If the last, enter the numbers of the limitations in the combination. Recommendations for future research are not qualifications *per se*; they are qualifications only if stated in times of caution in interpreting the results of the reported study. For an interaction ES, enter an x.

F.2.b. Treatment effective? This category does not ask for your judgment in regard to the effectiveness of the treatment, but asks you to code the author's conclusion about treatment effectiveness. If the author does not indicate whether he or she believed the treatment was effective, code "0". Otherwise, code "1" if the conclusion is that the treatment had no effect (including the lack of Treatment A vs. B differences); code "2" if the author concluded that the outcomes were equivocal--that is, that a conclusion that the treatment was effective was not clearly supported; code "3" if the conclusion is that the treatment did produce a positive effect on attitudes toward disabled persons; code "4" if it was concluded that the result was a negative effect--that is, the result was less positive attitudes toward disabled persons. Remember, what is sought is the author's conclusions in regard to a treatment's effectiveness relative to another, and this judgment may vary across dependent measures and, thus, ESs. For an interaction ES, enter x.

For Prior Contact ESs, the question is, did prior contact have an effect on treatment outcomes and, if so, what effect?

F.3. <u>Effect size(s)</u>. a. <u>Available</u>. If information to compute an ES is not available even though information on statistical significance is, but the direction of the difference cannot be determined, code "0". If information is available to calculate or estimate an effect size--a <u>D</u>, correlation coefficient, or both--code "1", unless the effect size involves a negative attitude change or no change by a treatment group. Then code, "2", "negative change". The "2" code is not applicable to posttest-only designs, but is applicable to pre-post designs. In a "Treatment A vs. B" comparison, code "2" if the group exposed to either treatment (anticipated to have a positive effect on attitudes) has a decline or no change in assessed attitudes. Use the "2" code, if appropriate, with follow-up as well as immediate or delayed pre-post comparisons, and with secondary effect sizes. If an ES cannot be computed, but statistical significance is available and the direction of difference is negativ , code "3". If an ES cannot be computed, but statistical significance is available and the direction of difference is positive, code "4".

F.3.b. D. 1) ± D. If a D can be computed or estimated--see the COMPUTATION OF EFFECT SIZES section--code it here. Include a + or a - sign to indicate the direction of the effect size, with a + indicating a difference in the direction of more positive attitudes for the treatment groups or from the pre- to posttest. When you enter a - sign, always circle it as a reminder to the key entry operator that a - is a value to be punched, not an indicator of missing data. With a Treatment A vs. B comparison, assume that Treatment A (the first group) is the treatment and assign a + or - accordingly. With an interaction effect, no D can be computed, and x's should be entered.

In using ES Information Missing CODING INSTRUMENTS, enter a + or - sign (in this case, circling either so that it will stand out for the key entry operator) if the direction of the group difference can be determined even though an ES cannot be computed. Enter x's in the remaining spaces. If direction of difference cannot be determined, enter an x in the first space and zeroes in the remaining spaces.

F.3.b.2) Source. If no D is available, as for an interaction ES, code "0". If an appropriate standardized mean difference is reported, code "1". If you calculate a D using any of the means and standard deviations indicated in 3.b.3) and 3.b.4) below, and discussed in the COMPUTATION OF EFFECT SIZES section, code "2". If the D is estimated, based on a t-test for comparison between means or the F from a one-way analysis of variance with only two groups, code "3". If the D is estimated from a t-test for correlated means for which the correlation is available, code "4". If the correlation is not available and an estimate of .50 was used, code "5". If D is estimated from the F for the treatment effect in a factorial analysis variance in which there are only two levels of treatment, code "6". If the D is estimated from an F for a one-way analysis of covariance for which there is only two groups, and the correlation between scores on the dependent measure and the covariate is available, code "7"; if the dependent measure-covariate correlation coefficient is not available and .50 is used as an estimate, code "8". If the data were reported in proportions and chi-squared or Z^2 is used to estimate D, code "9". If the corresponding t value for a Mann-Whitney U or some other nonparametric statistic other than chi-squared is used to estimate D, code "10". If an r_{pb} is the only information reported, and you use that to estimate D, code "11". If the only information provided in the report is an exact significance level from which you determine the value of a statistic (t, F, Z, chi-square) and then use that to estimate D, code "12".

F.3.b.3) Scale of $\bar{X}$ difference. In F.3.b.3) and F.3.b.4), you are to provide information about the elements used in computing the D for the particular effect size, if "2" was coded in F.3.b.2) above. If any category other than "2=calculated" was coded in F.3.b.2), code "0=not applicable". The categories of mean differences correspond to those discussed in COMPUTATION OF EFFECT SIZES. If the means used in computing D were the unadjusted posttest means, code "1". If a posttest mean is compared with the pretest mean of another group to simulate a control group comparison, code "1". If pretest and posttest means were available so that you could subtract the pretest from the posttest means and obtain mean gains for each group to use in the computation of D, or if change means were reported, code "2". If means adjusted as part of an analysis of covariance were used, code "3". If the researcher used the correlation between the pretest and the posttest to predict posttest scores and then computed residuals (the differences between actual and predicted posttest scores), and the mean residuals were used to compute D, code "4".

F.3.b.4) Standard deviation. Again, unless "2" was coded in F.3.b.2, indicating that you were able to calculate a D directly, code "0". If the standard deviation used to compute D was based on posttest data for a control or placebo group (or pooled standard deviations in the case where there was more than one control or placebo group), or if it comes from a pretest which was treated as a posttest for the purpose of a simulated control group comparison, code "1". If the standard deviation used in computing D came only from pre-test scores--as in pre-post, one-group designs or when control group posttest standard deviations aren't reported--code "2". If both a posttest standard deviation for the control group and pre-test standard deviations on the dependent measure for the control and experimental groups are available to be pooled (using the formula presented in the COMPUTATION OF EFFECT SIZES section), code "3". In the case that the standard deviation must be obtained from the pooled posttest sums of squares based on all of the groups in the design (see the COMPUTATION OF EFFECT SIZES section), code "4", "5", "6", or "7", as appropriate. If the only standard deviation available is the control group standard deviation for covariance adjusted, residual, or gain scores and you are able to estimate the unadjusted control group standard deviation using one of the formulae in the COMPUTATION OF EFFECT SIZES section, code "8" for an adjusted covariance or residual standard deviation or "9" for an adjusted gain score standard deviation.

F.3.b.5) Mdn primary D for type of assessment. In order to have one D for each type of assessment (as coded in D.6) which figures in the primary ES(s) for a research report, enter the median D here in the columns for primary ESs. (Enter x's in the secondary ES columns.) If the number of D's is odd, enter the middle one. If the number of D's is even, take the mean of the two (or more) middle values. If there is only one D for a type of assessment, re-enter it here. If the report contains ESs for attitude change efforts prior to mainstreaming and for mainstreaming, compute separate median D (and correlation coefficients, F.3.c., and variance ratios, F.3.d) for the "prior" primary ESs and the "mainstreaming" primary ESs. Also, if follow-up means are reported, compute median D's, correlation coefficients, and variance ratios separately for the immediate (or delayed) assessments and the follow-up assessment or assessments. Compute separate median ESs also if there is a a "true" within study replication (e.g., Experiment I and II) or if results from different populations of Ss are reported in what amount to two different studies (e.g., use of a curriculum with attitude assessments of the students who study it and the teachers who teach it). If the "within study replication" coded in category A.9.b. is a pseudo-replication (i.e., you have coded it because of the different levels of subjects, such as grade levels), compute one median.

F.3.b.6) Mdn primary D, overall. Here the purpose is to indicate the average D for all dependent measures that figure in primary ESs for the study. (Compute the median as indicated in 3.b.5.) If there is only one D for the study, re-enter it. Enter x's in secondary ES columns. Compute median primary overall D's separately in the same circumstances as in F.3.b.5.

F.3.c. Correlation. 1) Type. If no correlation coefficient can be obtained, code "0". For any effect size which involves the comparison of two group means on a dependent measure, obtain the point biserial correlation (r_{pb}), as described in the COMPUTATION OF EFFECT SIZES section, and code "1". For an interaction effect, Eta^2 is the appropriate correlation coefficient ("2"). When data are reported in terms of proportions of responses for two groups being compared, the Phi coefficient will be the appropriate correlation coefficient ("3").

Recall that when a F-ratio is reported for multiple means but the means are not reported, no effect sizes can be obtained for pairs of means. By the same token an analysis of proportions which includes more than two groups will not yield effect sizes, unless contingency tables are reported so that you can break out the data by pairs of groups and compute your own chi-squares and, then, effect sizes. Similarly, if an analysis of proportions for two groups has more than two levels of response (e.g., a greater than 2-by-2 contingency table, in which there are two columns, or rows, for treatment groups but three rows, or columns, for attitude assessment items), it will usually not be possible to obtain an effect size. You might be able to break such a table into a number of two-by-two tables to get separate effect sizes. Or, there may be occasions when the various levels of the dependent measure are equivalent indicators of attitudes and all of the results are in the same direction. In that case, Cramer's V, an extension of the Phi coefficient can be computed (see the COMPUTATION OF EFFECT SIZES section) and "4" should be coded.

F 3.c.2) ± coefficient. With Eta^2 for an interaction effect, no direction of relationship can be specified, and an x should be coded in the first space.

For the other coefficients, a direction can be specified through inspection. Enter a + if the treatment group has the higher mean or proportion and a - if the control group does. Check back to item F.3.b., as the + or - signs should be consistent for that item and this one. With Treatment A vs. B comparisons, assign a direction as indicated in F.3.b.1). Be certain to circle a - when you enter it.

F.3.c.3) Source. If no correlation was available for F.3.c.2), code "0". Indicate here whether the correlation which you recorded in F.3.c.2) came from the research report ("1"), from your calculations ("2") (as when a Phi coefficient is calculated from a chi-square), or was estimated from a D ("3").

F.3.c 4) Mdn primary coefficient for type of assessment. The purpose here is the same as for D's in F.3.b.6). Compute the median coefficient in the same way; enter in primary ES columns, with x's in secondary ES columns.

F.3.c.5) Mdn primary coefficient, overall. The purpose here is the same as for D's in F.3.b.7). Compute the median coefficient in the same way.

F.3.d. Variance ratio. 1) Ratio (S_E^2/S_C^2). Of interest here is whether there is any indication that after treatment the experimental group has less or greater variability than an untreated population or a differently treated population. The variance ratio is obtained for primary ESs by dividing the variance for the treatment group by the variance for the control or placebo group for the particular effect size being coded. In the case where the effect size involves a Treatment A vs. B comparison, put in the numerator the variance for the group which you have coded as the experimental group

(Treatment A) on the earlier pages of the coding instrument and put the variance for other group (Treatment B) in the denominator. The variance used must not be pooled--i.e., it is not to come from pooling the control group posttest variance with pre-test variances, nor can a variance ratio be calculated when D is computed based on an estimate of the variance obtained from a sum of squares that includes the treatment posttest scores. For secondary ESs other than within-gender treatment comparisons, enter x's .

F.3.d.2) Median primary variance ratio for type of assessment. The purpose here is the same as for D's in 3.b.6). Compute the median coefficient and enter in the same way.

F.3.d.3) Median primary variance ratio, overall. The purpose here is the same as for D's in 3.6.7). Compute the median coefficient and enter in the same way.

G. SUPPLEMENTAL INFORMATION

G.1. Information gain. This item is relevant for studies in which the presentation of information is the treatment/intervention to change attitudes. The question is whether the intervening variable of information gain was established. For reports for which "2=information" was not scored in C.4. Treatment/Intervention Technique, categories G.1.a., b., & c. should be coded as "not applicable" and x's inserted in G.1.d. If another intervening variable is the object of the treatment (such as anxiety), is assessed, and ESs can be computed, code it here and make a note on the COMMENTS ON CONVENTIONS Sheet.

G.1.a. Reported? Code here whether information gains were reported for relevant studies. Code "0" if information was not scored in C.4. as the intended treatment/intervention technique. Code ES's separately. That is, if information gains are reported for the groups in some ESs but not others, code accordingly in the ES columns.

G.1.b. Type of report. "1=verbal" refers to verbal descriptions of information gain outcomes with no reference to statistical significance and no reporting of means or standard deviations. If the author refers to "significant" or "nonsignificant" results, or reports levels of P, code "2". If means (or medians) alone, or with standard deviations, are reported, code "3".

G.1.c. Conclusions. Code here the author's conclusions in regard to information gains. If information gain is not reported (i.e., 0 or 2 was coded in G.1.a.), code "0", "not applicable". If no conclusion is stated, but there is a report of gains (i.e., "1=yes" was coded in G.1.a.), select a code based or how the reported results would typically be interpreted in educational research reports. For example, if the result is statistically significant, or without statistical significance, if an ES is 1 or larger, code "1" ("clear gain"). If there is not enough information to decide on "1", "2", or "3", or if the evidence is inconclusive (e.g., a small ES), code "4".

G.1.d. Effect size--D. If a D for information gain can be computed or estimated (see COMPUTATION OF EFFECT SIZES) for a primary or secondary attitude ES, enter it here, with a + or - sign. If D's are available for more than one information test, enter the median D. In general, if "0" or "2" were coded in G.1.a., if a D cannot be computed or estimated, or if the attitude ES is an interaction ES, enter x's.

G.2. Type of study. The purpose of this category is to identify reports of evaluations of ongoing, in-place courses, "1" (such as a regular college abnormal psychology or special education course or an inservice course), or programs, "2" (such as mainstreaming or a master's or other degree program), which involve no special treatment, as contrasted with reports of studies of treatments or interventions introduced in order to investigate the effects experimentally, "3". In some cases, this category will be redundant with the code "14" or "15" under category C.4.a. Information (2) Delivery mode (p. 5 of the Coding Instrument) when the course or program conveys information, but it will pick up the evaluation of regular program components that are, for example, of a contact nature such as practica. On the other hand, the categories in 10. Mainstreaming do not provide a place to enter variations in time of mainstreamed contact. Studies that investigate variations in contact time should be coded "4" for later identification.

H. CODING SUMMARY

1. Minutes spent coding. On the cover of the coding sheet, you will have indicated the time when you started to read the research report and when you completed the coding instrument, whether in one or more settings. Enter here the total number of minutes for reading and coding.

2. Coder. Enter here your code number as indicated on the coding instrument.

RETURN TO PAGE 1 AND COMPLETE CHECK LIST.

CHECKLIST

1. Citation checked. Check the citation for the report as entered in the alphabetical listing of reports to be sure that all information is correct. Indicate any changes and give to the secretary. Once done, put a check in the space provided.

2. References checked. If the reference list for the report has been checked to determine whether it contains citations for reports to be included in the review of literature and any such citations have been taken off to be followed up on, place a check in the space provided.

3. Every space marked. To be certain that no available information is left uncoded, it is important that every coding space be filled in, with a code number or an x to indicate the lack of information. Skim the Coding Instrument to be certain all spaces are filled and that all numerals are legible. Then place a check in the space provided.

4. ES data available. If the data to compute all of the relevant ES(s) were available, write "yes" in the space provided. If not, write "no" or "requested", as appropriate.

5. ESs computed. If computation of ESs is complete, check space provided. Otherwise, leave space blank.

6. ESs checked. Leave space blank until someone other than the person who originally coded the report checks on the appropriateness and computations of any ES(s). Then the person who checked the ESs is to enter his/her initials in the space provided.

7. Comments on Conventions Sheet. If a COMMENTS ON CONVENTIONS sheet was used, write "yes" in the space. If not, enter "no".

8. Comments on Study Sheet. If a COMMENTS ON STUDY sheet has been completed, write "yes" in the space. If none was needed, enter "no'.

9. Scoring log completed. The SCORING LOG provides a record of the reports each scorer has coded, including the serial order and whether they were scored as agreement checks. Once the LOG item is completed, check the space provided.

10. Report disposition sheet completed. The last thing to be done is to complete the REPORT DISPOSITION sheet. Once it is filled in, check the space provided in the checklist.

I. PRIOR CONTACT CODING SHEET

The particular interest is in whether there were prior contact effects on treatment outcomes with the dependent measures and samples identified as relevant for the primary ESs already coded. (Control of prior contact as an antecedent variable is a secondary interest.)

For studies where the effects of prior contact on treatment outcomes were assessed and effect sizes can be obtained, enter the code to identify the type of comparison in each secondary prior contact ES in A.6.d. on the original Coding Instrument. Enter "13" if there is an analysis of prior contact levels within the treatment group; enter "14" if there is analysis of treatment effects within prior contact levels across groups (as treatment effects within levels of gender were coded); enter "15" if the interaction of prior contact and treatment is analyzed. If "15" can be coded, do not code "13" or "14". If "14" can be coded, do not code "13". If more than one definition of prior contact (e.g., number and frequency) are analyzed separately, or if quality of prior contact is analyzed separately, each is the basis for an ES.

Complete a PRIOR CONTACT EFFECT SIZES sheet for prior contact ESs coded in A.6.d., based on the types of prior contacts in I.12., and for any quality (I.13.) ESs that could be computed. Enter the quality ESs last. Comp[illegible] on a PRIOR CONTACT EFFECT SIZES COMPUTATIONS sheet. Use the standard deviation used for computing treatment Ds. In computing Ds for within treatment group ESs ("13" in A.6.d.), treat "no contact" or "neutral (in the case of quality), as the control condition. That is, place the dependent variable mean for that classification second in the D; compare the other means with it. In the case of quantity of contact, compare means only against "none", not with each other. In the case of quality, also compare dependent variable means for positive versus negative groups.

For coding alternatives, see Table 2.

I.5. Prior contact assessed. If there is no mention of the extent of the Ss' prior contact with disabled persons, code "0". If prior contact is assessed, code "1". If there is no explicit assessment but prior contact is implicit (e.g., the Ss are either experienced psychiatric nurses or nurses in training) and prior contact is considered by the author(s) to be a relevant variable (i.e., pertinent to treatment outcomes), code "2". If "0" is coded, cross out items I.6. through I.7. and I.9. through I.15.

I.6. Use—selection/assignment. This category refers to the use made by the researcher(s) of the prior contact information. If "0" was coded in I.5., code "0". If the report indicates that prior contact was assessed, but no further mention of that information is made, code "1". If information about prior contact was only used to describe the sample, with no further use made of it, code "2". E.g., if the data are also used for assignment, or are only reported incidentally in reporting analyses, do not code "2". If prior contact differences between groups prior to treatment are checked as an assurance of comparability, code "2" and indicate the result(s) on the PRIOR CONTACT COMMENTS sheet. If coding of another category supersedes the coding of "2" when such comparisons are made, note the comparisons and results on the COMMENTS sheet. In some cases, researchers will use information about prior contact to determine who will be in the sample. For example, the researcher may exclude anyone who has had prior contact. If persons with prior contact were excluded from the sample, code "3". If the researcher

Table 2

Prior contact coding alternatives.

Situation	Effect Sizes: Primary	Effect Sizes: Secondary, Non P.C.	Effect Sizes: Secondary, P.C.
1. Prior contact not assessed.			
Code:	On P.C. Coding Instrument, #2, 4, 5, & 8. Others, "O" (not applicable) except x's in 9	On P.C. Coding Instrument, #2, 8, 9; rest x's	
2. Prior contact assessed, no prior contact ES(s).			
Code:	Entire P.C. Coding Instrument x's in 9	On P.C. Coding Instrument, #2, 8, 9; rest x's	
3. Prior contact assessed, prior contact ES(s).			
Code:	Entire P.C. Coding Instrument x's in 9	On P.C. Coding Instrument, #2, 8, 9; rest x's	Original coding instrument--A.6.a., b., c., d.; B.1.a., b., c.; B.3.; D.3., D.4., D.5., D.6.; F.1., F.2.b., F.3. (all). Copy others from previous coding. Entire P.C. Coding Instrument.

uses a category such as "no and minimal contact", code "3". If only persons with prior contact were included, code "4". If the researcher(s) categorized people by amount of contact and then selected subjects from those strata, code "5". If the researcher(s) stratified by amount of contact and assigned to groups from the strata, code "6". If prior contact is used as a covariate, including being entered first in a multiple regression equation, in analyzing posttest scores, code "8".

I.7. Use--outcome analysis, correlation. If the correlation of prior contact with posttest scores on the dependent measure for a treatment group is reported, or can be computed, code "2"; if with change scores, code "3". If the correlation between prior contact and posttest scores on the dependent measure is reported for treatment and control (or treatment B) groups, code "4"; with change scores, code "5". If "4" or "5" can be coded, do not code "2" or "3". If "5" can be coded, do not code "4".

I.9. N of added secondary ESs. Enter the number of new prior contact secondary ESs. Enter x's for primary ESs, and "0's" for secondary ESs if no previous contact ESs are coded for the study.

I.10. Type of assessment. If "0" was coded in I.5., code "0". If the report only mentions prior contact, but does not tell how it was assessed, code "1". If the researcher used a demographic or biographical questionnaire, code "2"; if the information was obtained during an interview, code "3". If the researcher assumed contact because of the setting in which the Ss had been prior to the research, such as having been in a mainstreamed classroom or school, working as a psychiatric aide in a hospital, or being a parent of a disabled child, code "4". If the researcher takes information that students have provided on a questionnaire or interview and recategorizes it or reclassifies it to come up with contact categories, code "5". If "5" is coded, then do not code "2", "3", or "4".

I.11. Type of prior contact setting. Unless "4=setting" was coded in I.6., code "0". Categories "1" and "2" are for students, and categories "3" and "4" are for teachers. If the prior setting was as aides or nurses in a mental hospital or institution for mentally retarded persons, code "5". If the information provided is that the contact was in a family setting, for example, as parents, code "6".

I.12. Definition of degree of contact. If "0" was coded in I.5., or if the ES is for quality (I.11.), code "0". If the survey or questionnaire used to assess prior contact is available, examination of the items on it is the preferred way to determine the definition of contact. If "5" was coded in I.6., code I.8. in terms of the definition of the classification levels. If contact is assessed in terms of the general amount of contact, such as none, some, a great deal, code "2". If contact is assessed in terms of the number of disabled persons with whom the person n had contact, code "3". If contact is assessed in terms of frequency of contact, that is, how many ti es the person has had contact with disabled persons, score "4". If contact is assessed in terms of length, such as number of hours, days, years, of contact, code "5". If contact is coded in ter s of intensity, such as "no" contact and "close" contact, code "6". If contact is assessed in terms of a dichotomy that cannot be coded in 2-6, such as "knows disabled persons or not", code "7". Or, if contact is assessed in terms of the type of relationship—such as friend, family member, code "8".

I.13. Quality of contact. Here the question is whether the intent was to assess whether the Ss' prior contact was a positive or negative experience for them. Code "0" if "0=No" was coded in I.5.

I.14. Direction of quality comparison. This category is applicable only to within treatment prior contact ESs (coded "13" in A.6.d.) which involve an assessment of the quality of prior contact. For all other ESs, code "0" (not applicable). Code "1", "2", or "3", depending on the comparison in the within treatment ES.

I.15. Type of disability. Here, code the disability or disabilities of the persons with whom the prior contact of the Ss was assessed. The categories are the same as for C.4.d., and conventions are also the same.

J. CONTACT CODING SHEET

The purpose is to code studies in which contact is the treatment to determine if attributes of contact are related to outcomes.

J.5. Status. This set of categories is included to get at the variable of relative status for the nondisabled and disabled persons involved in the contact.

J.5.a. Age. The question here is whether there is sufficient age difference to affect status as perceived by the nondisabled participants. For example, for young children in particular, a grade level difference is likely to be significant (e.g., 4th vs. 5th grade). If both disabled and nondisabled Ss appear to be about the same age or age differences are distributed evenly--e.g., all are campers or college students--code "2". In mental hospitals, patients may be a variety of ages, so, for example, it cannot be assumed that they are older than or the same age as college Ss. In such cases, unless specific age indicators are provided, code "0". If the ages are specified but there is a variety of age relationships (e.g., some nondisabled college Ss work with children, some with adults), code "4".

J.5.b. Educational-vocational prestige. Here code the relative status accorded by obvious educational achievement (as distinct from age) and/or vocational differences, as perceived by the nondisabled Ss. For example, if a disabled professor lectures to a group of students, code "3". If the nondisabled person is a nurse working with disabled patients, code "1". If nondisabled Ss are tutoring or otherwise in contact with disabled students, code "1" only if there is an obvious educational distinction, such as college students tutoring high school or younger students, or the disabled persons being tutored or in other situations are obviously educationally hampered (e.g., severely/profoundly retarded--TMR) so as to be at a lower level than the nondisabled Ss. That would not be true, e.g., for kindergarten Ss and TMR peers. If tutors are different in age, but close in grade level, code "2". In settings, such as a mental hospital where patients may come from a variety of backgrounds, code "0", unless specific indicators are provided.

J.5.c. Helping relationship. This category is intended to address whether the nondisabled persons are providing help of a personal sort, in an one-to-one relationship, to disabled persons. Excluded would be situations where the relationship is of a general, group nature, not one-to-one, such as cooperative learning groups (code "0", and pick up the nature of cooperation in 7.c.). For the "helping professions" (medical, psychological counseling, rehabilitation, special education teachers or other teachers working closely with disabled student) code "1". For trainees in a professional field in, e.g., a practicum, code "2". If professional persons such as psychiatric nurses, were in an inservice program, code "1". Coded "0" are, e.g., educational relationships, such as a disabled professor teaching a class or a disabled person coming into an elementary school to tell stories to the children. Nonprofessional ("3") includes helping relationships such as volunteers (including companions) in mental hospitals, volunteers in recreational programs who provide assistance to the disabled persons rather than just participating in recreational activities with them, and tutors.

J.5.d. Overall. Based on your coding of J.5.a.-J.5.c. and any other information in the report--for example, indications of low social status based on socio-gram data or on other elements in the situation such as special treatment due to the disability--decide what the relative status of the disabled and nondisabled persons was in the contact situation and code accordingly. Code "0=can't tell" only as a last resort.

J.7. Type of contact. a. Voluntariness. Here the question is whether or not the nondisabled Ss volunteered to have contact with disabled persons. For example, if students are arbitrarily assigned to a contact group or are in a group chosen to be a contact group, code "1". If there is only one group in a pre-post design and students do not volunteer to have contact, code "1", unless "2" or "4" is appropriate. If persons volunteered for a professional role that they knew would include contact (such as deciding to be a psychiatric nurse, an employee in an institution for mentally retarded persons, or a special education teacher) or for a preprofessional role that would involve contact, such as signing up for a college course that they knew included a practicum with disabled persons, code "2". If the design included Ss who voluntarily chose to have contact, for example, choosing to be companions to mentally ill or mentally retarded persons when that is not a part of a professional or preprofessional role, code "3". If within an assigned group, Ss may or may not have had contact, depending on their own volition--such as in a playground situation in which some nondisabled children may choose to play with disabled children while others may choose not to--code "4".

J.7.b. Intimacy. Contact can vary in how personal it is. For example, observing mentally retarded persons in an institution or being lectured to by a disabled person typically involves no interaction and should be coded "1".

Tutoring, playing together on the playground, studying together, or participating in a discussion led by a disabled person should be assumed to be "casual" contact (code "2"), unless the report indicates otherwise. Having a disabled person into one's home or sharing a cabin at a camp would be coded "3", unless there is evidence in the report to the contrary. "Varied" contact ("4") is coded when all Ss likely had some contact, but of varying types, such as in a typical mainstreamed classroom or a dormitory situation in which some Ss have disabled roommates and others don't. In a situation where there is potential for contact but no assurance that all Ss had contact with disabled persons, much less interacted with them, as on an integrated university campus, code "5".

J.7.c. Cooperation-competition. In some contact situations, typically those that lack interaction, there is no opportunity for cooperation or competition--for example, in observing mentally retarded persons in an institution or in being lectured to by a disabled person. In such cases, code "1". Some treatment situations in which there are opportunities for interaction do not necessarily call for cooperation or competition between the disabled and nondisabled persons, and either or both might have occurred. Examples would include, living together in a dorm, simply being together in a classroom, free play on the playground (recess), or care for disabled persons in a hospital or institution. In such cases, code "2". In some contact studies, cooperation is necessary for treatment implementation, but it is not the focus of the treatment. Examples include tutoring, classroom discussions, being a companion, or volunteers helping with a recreation program. Code such situations, "3". If cooperation is an explicit part of the treatment, as when the Ss are told to cooperate as part of the intervention--e.g., groups are told that their rades depend on a shared effort--code "4". If there is both implied and explicit (assigned) cooperation, code "4". If competition between the nondisabled and disabled persons is part of the natural setting--e.g., playing a sport to win or competition for grades--code "5". If Ss are told to compete as part of the intervention--e.g., points are assigned for performance, with a prize for the one with the most points--code "6". If there is both implicit and explicit (assigned) competition, code "6". If there is any combination of "3" and/or "4" with "5" and/or "6", code "7". Code "can't tell" only as a last resort.

J.7.d. Reinforcement. If there was likely some shared intrinsic reward for common performance of a task or participation in an activity (evidenced, for example, by mention of satisfaction with a job done), code "2". If the "contactees" share a reward given by someone else, such as a prize, code "3". If only the nondisabled person is reinforced, e.g., by praise, for having contact, code "4". Because external rewards are more easily identifiable, but are likely to also involve intrinsic satisfaction from achievement, "3" takes precedence over "2" in coding. Follow that convention in coding "5" or "6".

J.7.e. Pleasantness. Here the point is whether the nondisabled persons found the contact situation, not only the contact with disabled persons, to be pleasant or not. For example, if college students indicated that a practicum was "an excellent learning experience" that is evidence for coding "2"--unless they also indicate that what was excellent was learning to handle one's negative feelings about a situation. Don't score "1" or "2" without some indicator, such as responses on a questionnaire or during an interview.

Don't simply assume without evidence, e.g., that contact with a profoundly mentally retarded person is unpleasant or that common participation in a cooperative group is necessarily pleasant.

J.7.f. Modeling. The question here is whether others model positive interactions with disabled persons for the nondisabled Ss as part of the treatment or treatment setting, whether the modeling was in a natural setting or staged. A "significant other", coded "3", would be, e.g., a teacher. Score "0" ("can't tell) unless there are indicators that modeling took place or that modeling was not a part of the treatment.

J.8. Institutional/authority/peer support. Code here whether the norms that governed interactions, the persons in authority where the contact occurred, or the statements and behaviors of peers likely promoted positive attitudes toward disabled persons. Look for specific evidence that the nondisabled persons were aware of the norms (such as reports of explicit statements to them by teachers) or peers' negative attitudes (such as statements by the Ss themselves). Or look for evidence that such norms were not likely, as in a carefully controlled piece of research in which the Ss are kept blind as to the purpose of attitude modification. If such evidence is not available, code "0".

J.9. Characteristics of disabled persons. a. Disability. As type and severity of disabi... may be a factor in contact effects, the purpose here is to code disabilities as specifically as possible. Use the "can't tell", "0", category only as a last resort. If several types of disabilities were present and are specified, code "1" and enter the code numbers for the disabilities. See Table 3 for the li... of disabilities. If there are

Table 3. List of disabilities for Category 9.a.

0 = can't tell
1 = combination
2 = multiple disabilities
3 = MR, can't tell
4 = MR, general
5 = MR, mild/moderate (e.g., IQ 35 or 40 to 70); EMR
6 = MR, severe/profound (e.g., IQ below 35 or 40); TMR
7 = severe multiple impaired
8 = physical, general
9 = wheelchair
10 = amputee
11 = paraplegic
12 = facial disfigurement
13 = visually impaired
14 = blind
15 = hearing impaired
16 = deaf
17 = mentally ill
18 = emotionally disturbed
19 = learning disabled
20 = cerebral palsy
21 = speech impaired
22 = spinal bifida

several types of unspecified disabilities, e.g., in a special education classroom, code "2". If reference is to mentally retarded or developmentally disabled persons, but you cannot discern the level of retardation, code "3". If there is a range of mental retardation, e.g., moderately to severely retarded, code "4". If reference is made only to "physical disabilities", code "8". If the disabled person or persons are in a wheelchair, code "9", unless one of the other categories, e.g., "10=amputee" or "20=cerebral palsy", can be coded to give a more specific indication of disability. Note that there is not an "other" category. If you encounter a disability not included in the list on page 3, add it to the list after consultation.

J.9.b. <u>Negative stereotype</u>. A concern in the literature is whether the disabled persons with whom there is contact look or behave such as to reinforce negative stereotypes about disabled persons. Don't infer simply from the disability, e.g., blindness, that negative stereotyping was present, but look for clues in the descriptions of Ss' reactions, the disabled persons (e.g., paraplegics playing basketball or running races in wheelchairs are nonstereotypic and should be coded "2", unless there is specific information to the contrary), the treatment (e.g., mannerisms to be displayed by the disabled persons), and the situation (e.g., those in an institution for the mentally retarded are likely to be severely/profoundly retarded and reflective of stereotypes, so code "1" for visits to such institutions. The picture is not so clear for mental institutions. Some persons are likely to be there because their behavior is stereotypic; others may appear very normal. Unless there is evidence of severe psychiatric cases, code "0".). If disabled persons were selected purposely to avoid negative stereotypes, code "2".

J.9.c. Competence. The question here is whether the person's disability was likely perceived by the nondisabled Ss as affecting competence in regard to either (1) a specific, identifiable role which is part of the treatment (e.g., a physically disabled lecturer or a paraplegic athlete) or (2) general functioning. Code based on (1), if present; if not, code based on (2). If the disabled person(s) clearly lacked competence, code "1". For example, in a (1) situation, mentally retarded persons given an experimental role requiring verbal abilities they lacked or students being tutored who lacked the skills for the task being learned would be coded "1"; in a (2) situation, mentally retarded students in self-contained classrooms who interacted with nondisabled Ss, mentally retarded persons observed in institutions, or patients in psychiatric wards, or people in mental institutions would be coded "1". However, if in a (1) situation there was clear acknowledgement and acceptance by the disabled and nondisabled participants of any disability limitations in performing a task, code "2". If the disabled person is likely perceived as competent, either because the disability was irrelevant to a task at which the disabled person is competent (such as a lecturer in a wheelchair) or because the disabled person performed well a task for which the disability would ordinarily be considered an impairment (e.g., a wheelchair athlete), code "3". In some situations, it may be difficult to discern whether the disablity was relevant to the nondisabled persons' perceptions of the competence of the disabled persons. Examples would be a physically disabled elementary student in a classroom setting, or mentally retarded persons in a recreation program in which they may perform some activities well but not others. In such cases, code "0", "can't tell".

J.10. Characteristics of nondisabled. a. Personality related. If potentially relevant personality traits, such as I.Q. or authoritarianism, were not assessed, code "0". If a personality trait was assessed, but its relation to attitude outcomes was not tested, code "1". If the relation was tested (e.g., in a factorial ANOVA or through a correlation with posttest scores on the dependent measure(s) for a treatment group) and the author(s) indicated there was none--or, in the absence of that, the relation was not statistically significant at the .05 level--code "2". If there were different results on multiple personality measures or different results across groups, code "3". If there was a relationship, judged as above, code "4".

J.10.b. Prior attitudes related. Code as with 10.a., now attending to pre-post dependent measure(s).

META-ANALYSIS: MODIFYING ATTITUDES TOWARD DISABLED PERSONS

COMPUTATION OF EFFECT SIZES

Two types of effect sizes (ESs) will be used in this meta-analysis: (1) the standardized mean difference, the result of dividing the difference between two means by an appropriate standard deviation; and, (2) correlation coefficients which express the degree of relationship between the independent and dependent variables.

The Standardized Mean Difference (D)

At least two types of standardized mean difference have been presented in the literature. One is Glass's Delta, in which the difference between a treatment and a control mean is divided by the control group standard deviation; the other is Cohen's d, which is the difference between a treatment and control mean divided by the pooled standard deviation for the treatment and control group. We use the symbol, D, on the coding instrument for the standardized mean difference to be used in this study.

The rationale behind the standardized mean difference as an effect size is to compare the mean of subjects in one group (e.g., the treatment group) to the mean of the subjects in another group (e.g., the control group) relative to the dispersion of subjects who have not received any treatment. The control group standard deviation is one legitimate basis for estimating the dispersion in a non-treated population. On the other hand, if the pretest standard deviations for the treatment and control groups are also available, use of this additional information will yield a more stable estimate of the standard deviation in the untreated population. The guiding rule is to obtain the most stable estimate possible, excluding treatment effects if possible. Once the estimate of the standard deviation is obtained for a dependent measure, use it in computing all D's for that measure.

In this meta-analysis, the first choice for the standard deviation in computing a standardized mean difference will be the posttest standard deviation for the control or placebo group(s) pooled with the pretest standard deviations for the treatment and control (or placebo) groups. If pretest standard deviations are not available, then the control group standard deviation (sd) on the post assessment will be used, if available. In a COVAR analysis (or when residual gain means are used) the only sd available may be for the adjusted (or residual) scores. If r_{XY} is available, use it to estimate the sd for unadjusted scores. If not, use $r = .50$ to estimate the sd for unadjusted scores.

When the only information available is from an analysis of variance table, a pooled standard deviation including the treatment group(s) will be used. With a pre-post, one-group design, the pretest standard deviation should be used. And, with a Treatment A vs. B design, a pooled standard deviation for the pretest is preferred; if that cannot be computed, then a pooled Treatment A and B posttest standard deviation will be used. (See Table 4 for a summary of sd choices and formulae.)

Whatever standard deviation is chosen, based on the availability of information, the same standard deviation is to be used for computing all standardized mean differences for the particular dependent variable in the study.

The question of interest is whether the status of the groups on the posttest indicates any evidence of a treatment effect. However, a frequent contaminating factor will be the existence of pre-treatment differences

Table 4. Choice of standard deviation for computing D

ORDER OF PREFERENCE	RELEVANT FORMULAE
For treatment vs. control or placebo:	
1. Posttest sds for placebo and/or control group(s) pooled with pre-treatment sds for all groups.	1. #2
2. Posttest sd(s) for placebo and/or control group(s).	2. #2
3. Sd estimated from control group(s) sd(s) for gain, adjusted, or residual scores.	3. #3, 4
4. Pooled sd from an ANOVA including the treatment group(s).	4. #5, 6, 8
5. Estimated pooled sd from a COVAR including the treatment group(s).	5. #7
For pre-post, one-group design: pretest sd.	
For Treatment A vs. B:	
1. If part of a design which includes a control and/or placebo group(s), the same as for the treatment vs. control or treatment D (1-5 above)	1. See 1-5 above.
2. If a Treatment A vs. B only design, use:	
a. Pooled pretest sd.	2a. #2
b. Pooled posttest sd.	2b. #2
c. Estimated sd, if adjusted or residual scores used, as in 3 and 5 above.	2c. See 3 and 5 above

487

between the groups, even with random assignment. For that reason, when possible, the means used in computing the standardized mean difference should be adjusted for pre-treatment differences. The first preference, if the necessary information is available in the research report, is to compute and use raw gain means (that is, the posttest mean minus the pretest mean for each group). If the means of change scores are reported, they equal the raw gain means. The second preference will be adjusted means from an analysis of covariance. The third preference will be residual gain scores. The final preference will be the simple unadjusted posttest mean scores. (With each type of mean difference, use the standard deviation as defined above, or an estimate of it. Do not use the reduced standard deviation for the adjusted scores.)

Whenever possible, the standardized mean difference (D) should be computed directly from information provided in the report. When that is not possible because of a lack of information, D can be estimated from statistics which are frequently reported, such as t- or F- ratios, and failing that, from precisely reported significance levels.

To summarize, standard deviations are preferred in the following order: (see Table 4 for more detail)

1. Pooled post-control and pre-treatment and control standard deviations.
2. The post-control standard deviation.
3. A pooled standard deviation which includes the treatment group.

The order of preference in regard to means to be used is:

1. Raw gain means, or mean change scores.
2. Covariance adjusted posttest means.
3. Residual gain means.
4. Unadjusted posttest means.

The preference is always to compute the standardized mean difference (D) using means and one of the three standard deviations; if such information is not available in the report, then estimate the standardized mean difference using statistics such as t- or F-ratios or chi-square if they are reported; or, if such statistics are not reported but a precise significance level is, use it.

Calculation of D

$$D = \frac{\bar{X}_1 - \bar{X}_2}{sd} \qquad (\#1)$$

1. Computation using means and sd.

   a. Means

      1) If pretest and posttest means are available, subtract to obtain mean gain; or, if change score means are available, use them.

      2) If pre-post means are not available, use, in order of preference: (a) covariance adjusted means; (b) residual means (i.e., scores were predicted, usually based on pretest scores, and the obtained minus the predicted score used as the dependent measure); or, (c) unadjusted posttest means.

   b. Standard deviations

      1) If posttest control group and pretest treatment and control groups sd's are available (with multiple treatments, more than 3 sd's may be available), obtain a pooled sd, using the following formula (note that sd's are changed to variances for this computation):

$$sd = \sqrt{\frac{s_1^2(n_1-1) + s_2^2(n_2-1) + s_3^2(n_3-1) + \ldots + s_k^2(n_k-1)}{(n_1+n_2+n_3+\ldots+n_k)-k}} \qquad (\#2)$$

      where s_1^2 is the variance of group 1, n_1 is the number in Group 1, and k is the total number of groups. If you have multiple sd's to pool and one is an extreme outlier, do not include it in the pooled sd.

      2) If sd's are not available as in 1) above, use the control group posttest sd. You may have to estimate it in the following ways:

a) If gain scores are used and a standard deviation for gain scores (sd_g) for the control group is provided along with the pre-posttest correlation, then

$$S = S_g \Big/ \sqrt{2(1-r_{xy}^2)} \qquad \text{(G, M, S, p. 118)*} \qquad (\#3)$$

If r_{xy} is not reported, use $r_{xy} = .50$.

b) With residual or covariance adjusted scores and a standard deviation for control group residual scores (sd_{res}) and the pre-posttest correlation available:

$$S = S_{res} \Big/ \sqrt{1-r_{xy}^2} \qquad \text{(G, M, S, p. 118)} \qquad (\#4)$$

If the r_{xy} is not available, use $r_{xy} = .50$.

3) If the necessary standard deviations are not available to get a "treatment-free standard deviation, use the following:

a) With a one-way ANOVA:

$$S = \sqrt{MS_w} \qquad (\#5)$$

b) With a one-way ANOVA, if only F (and not a source of variance table) and group means are provided, compute:

$$MS_B = \frac{\Sigma(\bar{X}_j - \bar{X})^2}{k-1} \quad \text{and} \quad MS_w = MS_B / F \qquad \text{(G, M, S, p. 128)} \qquad (\#6)$$

If $\bar{X}$ (the grand mean) is not reported, compute it as a weighted average of the group means:

$$\bar{X} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2 + \ldots + n_j\bar{X}_j}{N} \qquad (\#6A)$$

c) With a one-way COVAR, and one covariate:

$$S = \sqrt{\frac{MS_w \quad (d.f._w - 1)}{(1-r_{xy}^2) \quad (d.f._w - 2)}} \qquad (\#7)$$

If there is more than 1 covariate, the $d.f._w$ terms must be decreased by 1 more for each additional covariate. If r_{xy} is not available, use $r_{xy} = .50$.

---

*Sources of special formulae are included. G, M, S = Glass et al., 1981; Hays = Hays, 1973; White = Karl White, unpublished materials.

d) With an n-way ANOVA (a factorial design) or COVAR, all sources but that for the treatment must be pooled; so, for a 2 x 2 design, with the treatment the A factor:

$$S = \sqrt{\frac{SS_B + SS_{AB} + SS_w}{df_B + df_{AB} + df_w}}$$ (G, M, S, p. 119) (#8)

2. Using an inferential statistic to estimate $\underline{D}$ when an sd cannot be obtained.

Sometimes when a mean difference is not statistically significant, the report will not indicate which group had the higher mean. In that case, even though $\underline{D}$ might be obtained, no sign could be attached to the $\underline{D}$; it would be meaningless for our purposes, and should not be coded.

a. t-test for independent groups:

$$\underline{D} = t\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$ (G, M, S, p. 126) (#9)

b. t-test for correlated groups (matched pairs):

$$\underline{D} = t\sqrt{\frac{2}{n}(1 - r_{xy}^2)}$$ White (G, M, S, p. 126) (#10)

If r_{xy} is not available, use $r_{xy} = .50$.

c. t-test comparing gain scores:

$$\underline{D} = t\sqrt{2(1 - r_{xy}^2)(1/n_1 + 1/n_2)}$$ (G, M, S, p. 127) (#11)

If r_{xy} is not available, use $r_{xy} = .50$.

d. F from a one-way ANOVA with 2 levels:

with $n_1 = n_2$, use formula #9, or

$$\underline{D} = 2\sqrt{\frac{F}{n_1 + n_2}}$$ (White) (#12)

With $n_1 \neq n_2$, take $F = t$ and use formula #9.

e. F from a one-way COVAR with 2 levels and 1 covariate:

$$\underline{D} = 2\sqrt{\frac{F'(1 - r_{xy}^2)(df_w - 1)}{(n_1 + n_2)(df_w - 2)}}$$ (G, M, S, p. 127) (#13)

If r_{xy} is not available, use $r_{xy} = .50$.

If there is more than 1 covariate, $d.f._w$ terms must be adjusted by 1 more for each additional covariate. If r_{xy} is not available, use $r_{xy} = .50$.

f. n-way ANOVA with 2 levels of the treatment factor:

All sources of variation <u>other than the treatment</u> must be collapsed, an adjusted MS_e computed (add SSs and add d.f.s., and divide SSs by d.f.s. to get adjusted MS_e), and an adjusted F computed, which is the MS_B divided by the adjusted MS_e. Then, if $n_1 = n_2$, use formula #9 or

$$\underline{D} = 2\sqrt{\frac{F_{1,d.f.}}{n_1 + n_2}} \qquad (\#14)$$

If $n_1 \neq n_2$, take $\sqrt{F} = t$ and use formula #9.

3. Using level of statistical significance when an sd or test of significance is not available and an exact p value and n's or d.f. are reported:

Find the exact level of statistical significance for the specific d.f. in the appropriate table and read off the corresponding statistic (e.g., t), then proceed as in 2. above.

4. Using nonparametric test statistics, other than chi-square, when 2 groups are compared:

Substitute for the nonparametric statistic (e.g., the U from a Mann-Whitney test) the value of t with the equivalent level of significance, and proceed as in 2. above.

5. When results for two groups are available in proportions or percentages, with the statistical significance of the difference in proportions tested using either Z or chi-square:

Probit transformations (finding the standard normal deviates of the proportions and taking the difference between them as an estimate of $\underline{D}$) have commonly been used in this situation. An alternative which seems to produce more realistic results (i.e., fewer high $\underline{D}$'s, out of line with the $\underline{D}$'s from means and standard deviations and/or t's) is to compute Phi and use it to estimate $\underline{D}$.

In the case of proportions for two groups, the statistical significance of the difference can be tested using Z or chi-squared, with identical results, because with 1 d.f., the square root of chi-squared is a normal deviate, so Z^2=chi-squared.

First, compute Phi, using either of the following formulae:

$$\phi = \sqrt{\frac{\chi^2}{N}} = \sqrt{\frac{\chi^2}{N}} \qquad (\#15)$$

Then, compute $\underline{D}$, using the following formula:

$$\underline{D} = \frac{2\phi}{\sqrt{1-\phi^2}} \qquad (\#16)$$

Note: If the chi-squared has more than one d.f., neither phi or $\underline{D}$ can be computed.

6. When a point biserial r is available,

$$\underline{D} = \frac{2r}{\sqrt{1-r^2}}$$

Correlation Coefficients

Effect sizes can be expressed in terms of the degree of association between group membership (e.g., treatment and control group membership) and scores on the dependent variable. For all instances except interaction effects, ESs for this meta-analysis will be recorded both as $\underline{D}$'s and correlation coefficients.

For the straightforward design which involves the comparison of two groups on a posttest, the point biserial correlation (r_{pb}) is the coefficient of choice. The r_{pb} is also appropriate for pre-post or post-delay comparisons. When the findings are in the form of percentages (e.g., the % of the treatment group with favorable attitudes, as compared to the % of the control group), a Phi coefficient is appropriate.

For interaction effects, a $\underline{D}$ will not be available, but a correlation coefficient can be computed—Eta^2. Cramer's V, an extension of Phi, may be appropriate for some X^2 tables that are larger than 2-by-2. (V has the advantage over the contingency coefficient, C, in that it can attain a value of 1.00 and can be compared across contingency tables with varying d.f.'s.)

Calculation of correlation coefficients

1. r_{pb}

If the r_{pb} for the particular ES is included in the report, use it. If r_{pb} is not reported, it can be obtained in the following ways:

a) With a D as an effect size for a comparison of means:

$$r_{pb} = \frac{D}{\sqrt{D^2 + 4}} \qquad (\#17)$$

b) With an ANOVA with only 2 levels of the treatment, when a source of variance table is provided, divide the sum of squares for the treatment by the total sum of squares to obtain an Eta^2 (with 2 groups, $Eta^2 = r_{pb}^2$), and take the square root:

$$E^2 = r_{pb}^2 = \frac{SS_B}{SS_{Total}}; \quad r_{pb} = \sqrt{r_{pb}^2} \qquad (\#18)$$

2. Phi

With two proportions, a Phi can be computed using either Z or chi-squared:

$$\phi = \sqrt{\frac{Z^2}{n}} \quad \text{or} \quad \phi = \sqrt{\frac{\chi^2}{N}} \qquad (\#19)$$

3. Eta^2 for an interaction effect

As noted in 1. above, Eta^2 is a ratio of the sum of squares for a source of variance to the total sum of squares. So,

$$E^2 = \frac{SS_{AB}}{SS_{Total}} \qquad (\#20)$$

4. Cramer's V can be computed using Phi or chi-squared:

$$V = \sqrt{\frac{\phi^2}{L-1}} = \sqrt{\frac{\chi^2}{N(L-1)}} \qquad \text{(Hays, p. 745)} \qquad (\#21)$$

with L the smaller of R, the number of rows, or C, the number of columns.

Calculation of Variance Ratios

The point of calculating a variance ratio (S^2_E/S^2_C) is to determine whether the treatment has had an effect on variability. The experimental group posttest variance (remember always to square the standard deviation) is divided by the control group posttest variance. In Treatment A vs B ESs, divide the Treatment A posttest variance by that for Treatment B. With a pre-post, single-group design ES, divide the posttest variance by the pretest variance. Variance ratios should be computed only if the standard deviation or variance is available for each group. Estimates should not be used.

META-ANALYSIS: MODIFYING ATTITUDES TOWARD DISABLED PERSONS

CONVENTIONS ADDENDA

1. Omnibus F. If an omnibus F (for 3 or more means) is computed and reported as statistically nonsignificant, then post hoc comparisons of all pairs of means would be statistically nonsignificant. Consequently, "1" ("not significant at .05 level") would be coded for "F. Results. 1. Statistical Significance", for each pair of means. (Note that the overall F would not provide the information necessary to compute ESs.) However, if an omnibus F for 3 or more means is statistically significant, nothing can be concluded about the significance of differences between specific pairs of means, and "0" would be coded. If means and standard deviations, or means and the F, are available, ESs can be computed. Otherwise, unless more information can be obtained from the author(s), ESs cannot be computed.

2. Dependent measures.

Interactions. The number or nature of interactions of nonhandicapped Ss with handicapped persons may be a dependent measure in a study of attitude change, on the premise that changed attitudes will be manifested in changed (e.g., more frequent, more positive) behavior by nonhandicapped Ss. However, studies in which the aim is simply to change behavior *per se*, rather than to assess behavior as a possible outcome of an attitude change intervention or to change behavior as a means of changing attitudes, as separately assessed, are not relevant for the meta-analysis.

Social status. As with interactions, the assessment of the social status of handicapped persons may yield a dependent measure in a study of attitude change, on the premise that changes in attitudes will be reflected in changed estimates of social status. However, studies in which the aim is to change social status _per se_, rather than to assess social status as an outcome of an effort to change attitudes or to change social status as a means of affecting attitudes, as separately assessed, are not relevant for the meta-analysis.

Other measures. In general, a measure--such as the attractiveness of disabled persons or volunteering to work with disabled persons--that is not a straightforward assessment of attitudes is _not_ to be included in the meta-analysis as a dependent measure unless the research report indicates the measure was considered by the investigators to be an attitude assessment. Then, of course, its validity must be judged (D.10.e). This is consistent with the treatment of the Rucker-Gable Educational Programming Scale (RGEPS) specified on page D-3 of the CONVENTIONS FOR USE OF CODING INSTRUMENT.

4. _Static group program evaluations_. Some of the reports to be coded will involve, rather than planned treatments or interventions, post hoc evaluations of instructional programs, e.g., courses of study for prospective special education teachers or rehabilitation counselors. These studies will typically involve a static group design with assessment at the end of the program. Program students and students from other programs who are at comparable points in their education will typically be the Ss. In the Coding Instrument, Category C. _Treatment/Intervention_, 1. _Basis_, calls for a judgment as to the basis for the treatment or intervention. In this particular kind of research, as with many mainstreaming studies, there is not a planned treatment or intervention but an effort to determine whether an ongoing program is having some effect. For that reason, and to be able to sort out this type of program evaluation projects during the analysis, code

"6=other" for Category C.1. and then specify the type of program being evaluated--e.g., graduate program for counselors. The same line of reasoning will apply to mainstreaming studies which have been implemented as the result of public policy and then studied on a post hoc basis. If, however, there is a clear rationale laid out for the program and/or expectations for program outcomes, score the study appropriately. For example, if there is a discussion of a theoretical basis for the program, or if the type of program or the expectations of outcomes, are based on the citation of prior research, code as "3" or "4", as appropriate.

5. Interaction ES's. When an interaction is coded (A.6.d.), only an Eta^2 will be available as an ES. Moreover, experimental and control group information is not codable if there are more than two groups. In that case, for categories for which experimental and control group spaces are to be filled in, enter x's or the code for "not applicable", if there is one. For the other categories, enter codes as usual.

6. Incidental dependent measures. Include in ESs only measures defined or clearly intended by the researcher(s) as dependent measures which assess attitudes. That is, if results are reported on measures which were not considered to assess attitudes (e.g., an observational measure which is not discussed as an assessment of the behavioral aspect of attitudes is included as well as an attitude scale) or which were not the aim of the treatment (e.g., the purpose was to change authoritarian attitudes toward the mentally ill, as assessed by the OMI Authoritarianism Factor, but information is also given incidentally for other OMI factors, including the Unsophisticated

Benevolence factor), do not include them. As noted in the STATEMENT OF GENERAL PURPOSE AND POPULATIONS, measures such as sociometric scales, friendship choices, or positive interactions are relevant only if the researchers consider them to be assessments of attitudes.

7. Primary ES for Solomon 4-group designs. With a Solomon 4-group design, compute a D_a for the pretest-posttest part of the design (using raw gains) and a D_b for the posttest-only part of the design and pool the two $\underline{D}$'s (weighting by $\underline{n}$ if the design is not balanced).

With equal $\underline{n}$'s:

$$\underline{D} = \frac{D_a + D_b}{2}$$

With unequal $\underline{n}$'s:

$$\underline{D} = \frac{(D_a)(n_a) + (D_b)(n_b)}{n_a + n_b}$$

If pretest means are not reported for the pretest-posttest part of the design, pool the means and standard deviations for the two design parts and then compute a D.

8. ESs with the OMI. Unless the authors indicate they scored their data otherwise, high scores on both the OMI Authoritarianism (A) and Unsophisticated Benevolence (B) Scales are "bad" and low scores "good".

9. Intra-coder Agreement Checks. After you have coded 30 studies in sequence, excluding reliability checks, go back to the beginning of that sequence and select the first article with four or fewer effect sizes to recode as an intra-coder reliability check.

10. ATDP--Forms A and B. If Forms A and B of the ATDP are used for the pretest and posttest, and pretest (Form A) scores are used as the covariate in a COVAR, use the adjusted posttest means (Form B) to compute ESs. However, if adjusted means are not reported, or if an ANOVA or t-test is used on the posttest means, compute raw gains using the two forms (difference between Form A, pretest, mean and Form B, posttest, mean, or vice versa) to use in computing ESs.

11. Studies with ES information missing. Use the ES Information Missing CODING INSTRUMENT to code studies that lack the necessary information to compute ESs only if information on the statistical significance of results is available. If the study is relevant, but ESs cannot be computed and information is not available to choose between "1" and "2" in category F.1 on page 14 of the CODING INSTRUMENT, the study should be placed on the "Lack of Information" discard list and in the "Lack of Information" discard pile. For example, if ESs could not be computed and pretest-post gains for a control and an experimental group were tested separately for statistical significance, but no comparison was made of experimental-control posttest means, the study should not be coded.

If a report has one or more ESs which can be computed and one or more which cannot, code the two sets of ESs separately using the appropriate CODING INSTRUMENT (i.e., the ES Information Missing instrument for the second set). Code as if you had two separate studies. For example, for the first set, the N of ESs (A.6.a. of the CODING INSTRUMENT) will be the number which can be computed; for the second set, the number for which information is missing. Where Report ID# is recorded on coding sheets, enter the ID# without an asterisk for the first set and with an asterisk for the second set. Also, change the Alphabetic and Numerical lists so that the ID# is followed by an asterisk within parentheses, to indicate that the study has been coded using both instruments.

APPENDIX D

LISTS OF DISCARDED REPORTS

(1) Irrelevant Reports

(2) Reports Lacking Information

Discarded Reports—Irrelevant

Agness, Phyllis J. (1980). Unpublished doctoral dissertation, Ball State University.

Alper, Sandra & Retish, Paul M. (1972). Training School Bulletin, 69, 70-77.

Aloia, Gregory F., Beaver, Robert J., & Pettus, William F. (1978). American Journal of Mental Deficiency, 82(6), 573-579.

Anderson, Shirley Seaton. (1982). Unpublished doctoral dissertation, George Peabody College for Teachers.

Anthony, William A., & Cannon, John. (1969). Rehabilitation Counseling Bulletin, 12, 239-40.

Appolone, Carol; Romeis, James; Gibson, Patricia; McLean, William; & Howard, George. (1979). Epilepsia, 127-132.

Armstrong, Barbara, Johnson, David W., & Balow, Bruce. (1981). Contemporary Educational Psychology, 6, 102-109.

Armstrong-Hugg, Robin Lee. (1982). Unpublished doctoral dissertation, The American University.

Asher, Nancy Weinberg. (1973). Rehabilitation Psychology, 20(4), 156-164.

Babow, I. & A. Johnson. (1969). American Journal of Learning Disabilites 74, 116-124.

Baker, Sheldon R. (1964). Nursing research, 13, 345-347.

Ballard, Maurine; Gottlieb, Jay; Corman, Louise; & Kaufman, Martin J. (1977). Journal of Educational Psychology, 69(5), 605-611.

Barclay & Kehli. (1979). Journal of Research and Development in Education, 12, 80-92.

Barnes, Ellen Beard. (1978). Unpublished doctoral dissertation, Syracuse University.

Barr, James. (1985). Unpublished doctoral dissertation, East Texas State University. DAI, 46(08), 2262-A.

Becker, Sidney Graves. (1979). Unpublished doctoral dissertation, Queens College.

Begab, Michael J. (1968). Unpublished doctoral dissertation, Catholic University of America.

Begab, Michael J. (1968). The effect of differences in curricula and experiences on social work student attitudes and knowledge about mental retardation, ED 029 408.

Begab, Michael J. (1970). American Journal of Mental Deficiency, 74, 801-808.

Bell, Steven. (1981). The Journal of Special Education, 17, 371-375.

Benson, Gwendolyn T. (1982). Unpublished doctoral dissertation, University of Kansas.

Bernnardt, Karl S.; Johnston, Frances L.; Foster, Nan; & Brown, Marguerite. (1959). Mental Hygiene, 43, 394-399.

Berrigan, Carol Rizzo. (1979). Unpublished doctoral dissertation, Syracuse University.

Bilings, H.K. (1963). Journal of Experimental Education, 31, 381-387.

Boeck, Debora G., & Foster G. G. (1975, October). Psychology in the Schools, 12, 409-411.

Bond, J. H. (1965). Mammouth, OR: Teaching Res. Division, Oregon State System of Higher Education.

Borton, Deborah H. (1981). Unpublished doctoral dissertation, University of Pennsylvania.

Boucher, Carol Robin. (1979). Unpublished doctoral dissertation, University of Minnesota.

Bradfield, Robert H., Brown, Josephine, Kaplan, Phyllis, Rickert, Edward, & Stannard, Robert. (1973). Exceptional Children, 40, 384-390.

Britton, Jean O.; Thomas, Kenneth R. (1972). American Archives of Rehavilitation Therapy. 112-13.

Brooks, Beatrice R. (1959). Unpublished doctoral dissertation, New York University.

Bruininks, Rynders & Gross. (1974). American Journal of Mental Deficiencies, 78, 377-382.

Budoff, Milton, & Siperstein, Gary N. (1978). American Journal of Mental Deficiency, 82 (3), 474-479.

Bursor, Davele E.; Marcon, Rebecca; & Coon, Robert C. (1981). Journal for Special Educators, 17, 166-173.

Buttery, Thomas J. College Student Journal, 74-78.

Buttery, Thomas J. (1979). The Southern Journal of Educational Research, 13(2), 65-69.

Cable, Lawrence. (1977). Unpublished doctoral dissertation, Boston University.

Cantrell, R. P., & Cantrell, M. L. (1976). Exceptional Children, 42, 381-386.

Carlson, Laurance B. & Potter, Robert E. (1972). Journal of School Psychology, 10(2), 147-151.

Carroll, Andrea, & Purdy, Joseph. (1978). Report to the California State Department of Education

Cary, J. A. (1978). Dissertation Abstracts International.

Cavallaro, S.A. & R.H. Porter. (1980). American Journal of Mental Deficiencies, 84, 357-366.

Centers, Louise, & Centers, Richard. (1963). The Journal of Social Psychology, 61, 127-132.

Chafin, Peggy, & Peipher, Richard A. (1979). American Annals of the Deaf, 73, 468-471.

Chaires, Madelyn Chennault. (1966). Unpublished doctoral dissertation, Indiana University.

Chapman, Marjorie Ellen. (1975). Unpublished doctoral dissertation, Texas Woman's University.

Chasey, William C.; Swartz, John D.; Chasey, Carol C.; Brogman, Ruth; & Sandler, Howard. (1975). Therapeutic Recreational Journal, 9, 68-74.

Chennault, Madelyn. (1967). American Journal of Mental Deficiency, 72, 455-458.

Chueca, Y. Mora. (1985). Unpublished doctoral dissertation, University of California. DAI, 46(08), 2263-A.

Clark. (1976). Education & Training of the Mentally Retarded, 11, 333-335.

Cleland, Charles C., & Chambers, William R. (1959). American Journal of Mental Deficiency, 64, 124-130.

Cleland, C. & I. Cochran. (1960). American Journal of Mental Deficiencies, 64, 817-822.

Cohen, A.R. (1959). Journal of Personality, 27, 386-396.

Cohen, J. (1983). Teaching Exceptional Children, 16(1), 23-25.

Cohen, J.S. (1963). American Journal of Mental Deficiencies, 67, 705-713.

Cohen, Michael Paul. (1983). Dissertation Abstracts International, 45 (5).

Cohen, Shirley. (1977). Children Today, 6, 7-12.

Cohen, S. (1977). Educational Forum, 42, 9-20.

Cohen, S. (1978). Educational Digest, 43, 16-19.

Combs & Harper (1967). Exceptional Children, 33, 399-403.

Comer, Marilyn Scott. (1977). Unpublished doctoral dissertation, University of Mississippi.

Conine, Tali Alborz. (1968). Unpublished doctoral dissertation, Indiana University. (DAI, 29, 4102-A)

Cook, J. William, & Wollersheim, Janet P. (1976). The Journal of Special Education, 10, 187-198.

Cooper, Lucille, Johnson, David W., Johnson, Roger & Wilderson, Frank. (1980). The Journal of Social Psychology, 111, 243-252.

Cormack, Peter H. (1967). Rehabilitation Counseling Bulletin, 10, 106-109.

Crandell, John M. (1969). The Training School Bulletin, 66(2), 72-79.

Custer & Osguthoyse (1983). Exceptional Children, 50, 173-174.

Daar, Lisa. (1985). Unpublished doctoral dissertation, State University of New York at Stony Brook.

Dancer, Jesse E., & Keiser, T. Hope. (1981). Journal of Self-help for Hard-of-Hearing People, July/August.

Davis, Frances L. (1977). Unpublished doctoral dissertation, University of Missouri-Columbia.

Day, Harry P. (1959). The Journal of Teacher Education 10 (3), 326-328.

Dehouske, Ellen Jane. (1981). Unpublished doctoral dissertation, University of Pittsburgh.

Del-Val, Patricia B. (1981). Quincy Public Schools.

DeWar (1982). Teaching Exceptional Children, 14(5), 188-189.

Dewling. (1985). Academic Therapy, 20, 347-351.

Dhir, Arpan, & Shirali Kishwar Ahmed. (1978). Indian Journal of Clinical Psychology, 5, 173-178.

Diggs, Eugene. (1963). Unpublished doctoral dissertation, Colorado State College.

Dinnerstein, Albert J., & Lowenthal, Milton. (1968). Archives of Physical, medicine and rehabilitation, 49(3), 167-169.

Distefano, M. K., Jr., Craig, Stella H., Henderson, George L., & Pryer, Margaret W. (1975). Psychological Reports, 37, 1321-1322.

Distefano, M. K., Jr., & Pryer, Margaret W. (1978). Psychological Reports, 42, 190.

Donder, Daniel, & Nietupski, John. (1981). Education and Training of the Mentally Retarded, 16, 270-276.

Doubrava-Harris, M. (1982). Unpublished doctoral dissertation, Bowling Green State University.

Douglas, Dorothy Ff., Westley, Bruce H., & Chaffee, Steven H. (1970). Journalism Quarterly, 47, 479-87, 92.

Dowds, Murdo M., Jr., Kulik, James A., & Scheibe, Karl E. (1969). Psychological Reports, 25, 35-40.

Downes S. C. (1967). Dissertation Abstracts International, 29, 120-A.

Drabman, Ronald, & Lahey, Benjamin. (1974). Journal of Applied Behavior Analysis, 7, 591-598.

Drabman, Ronald; Spitalnik, Robert; & Spitalnik, Karen. (1974). Journal of Applied Behavior Analysis, 7, 93-101.

Dresang, Eliza T. (1982). Paper presented to the Convention of the Association for Educational Communications and Technology, Dallas. (ERIC #ED 223 215)

Dunlop, Kathleen H., Stoneman Zolinda, & Cantrell, Mary Lynn. (1980). Exceptional Children, 47, 132-141.

Edelmann, Anne M. (1966). A Pilot Study in Exploring the Use of Mental Health Consultants to Teaches of Socially and Emotionally Maladjusted Pupils in Regular Classes. EDd 026 292.

Efron, Rosalyn E., & Efron, Herman Y. (1967). American Journal of Mental Deficiency, 72, 100-107.

English, R. William. (1966). Unpublished master's thesis, Southern Illinois University.

English, R. William. (1971). Rehabilitation Research and Practice Review, 2, 19-27.

Eron, Leonard D. (1955). Nursing Research, 4, 24-27.

Farina, Amerigo; Fisher, Jeffrey D.; Getter, Herbert; & Fischer, Edward H. (1978). Journal of Abnormal Psychology, 87(2), 272-279.

Fink, Judith. (1981). Dissertation Abstracts International, 41(9), 3981-A.

Fiorentino, Michael Jr. (1978). Unpublished doctoral dissertation, University of Massachusetts.

Fishler, Karol; Koch, Richard; Sands, Russell; & Bills, Jack. (1968).

Journal of Medical Education, 43, 64-68.

Flanders & Havumaki. (1960). Journal of Educational Psychology, 51, 65-68.

Fletcher, Grace Marjann. (1978). Unpublished doctoral dissertation, Arizona State University.

Franzman, Burton W. (1976). Unpublished doctoral dissertation, University of Massachusetts.

Gaier, Eugene L.; Linkowski, Donald C.; & Jaques, Marceline E. (1968). The Journal of Social Psychology, 74, 117-126.

Garrett, Mary K., & Crump, W. Donald. (1980). Learning Disability Quarterly, 3, 42-48.

Gay, Nancy. (1976). Unpublished doctoral dissertation, East Texas State University. (DAI, 37, 4278-A)

Gelfand, Sidney, & Ullmann, Leonard P. (1961). The International Journal of Social Psychiatry.

Gelles, H. (1976). Unpublished doctoral dissertation, Rutgers University.

Genskow & Maglioni. (1965). Journal of Social Psychology, 67, 329-341.

Gerjuoy, Herbert, Rosenberg, B. G., Bond, James G., McDevitt, Robert, & Balogh, Joseph K. (1963). The Journal of General Psychology, 68, 173-180.

Giordano, Jo-Ann Marie. Unpublished doctoral dissertation, The University of Toledo.

Glass, Raymond M., & Meckler, Roy S. (1972). Exceptional Children, 39, 152-156.

Glazzard, P. (1979). Teaching Exceptional Children, 11, 101-104.

Goldfried, Marvin R., Dworkin, Earl, & Adinolfi, Allen A. (1969). Journal of Abnormal Psychology, 74(4), 482-485.

Gonwa, James Nicholas. (1981). Unpublished doctoral dissertation, Western Michigan University.

Goodman, Gottlieb & Harrison. (1972). American Journal of Mental Deficiencies, 76, 412-417.

Gorn, Gerald J., Goldberg, Marvin E., & Kanungo, Rabindra N. (1976). Child Development, 47, 277-280.

Gosse, Verena F., & Sheppard, Glenn. (1979). Canadian Counsellor, 13, 131-135.

Gottlieb, Jay. (1974). American Journal of Mental Deficiency, 79 (3), 3,268-273.

Gottlieb & Budoff. (1973). American Journal of Mental Deficiencies, 78, 15-

Gottlieb, Cohen & Goldstein. (1974). Training School Bulletin, 71, 9-16.

Gottlieb, Jay. (1972). Unpublished doctoral dissertation, Yeshiva University.

Gottlieb, J. A., & Siperstein, G. N. (1976). American Journal of Mental Deficiency, 80, 376-381.

Grab, Susan G. (1985). Unpublished masters thesis, Pacific Luthern University.

Grantham, Emily V., & Block, Marvin J. (1983). Journal of Dental Education, 47, 681-684.

Gunderson. (1978). Academic Therapy, 14, 99-102.

Guralnick, M.J. (1976). American Journal of Orthopsychiatry, 46, 236-245.

Guralnick, Michael J. (1980). Exceptional Children, 46, 248-253.

Guralnick, Richardson & Heiser. (1982). Exceptional Children, 48(4), 338-346.

Gustin, Samuel, L. (1973). Viewpoints in teaching and learning, 49, 85-95.

Haefner, James E. (1976). Journalism Quarterly, 53, 211-215.

Hagino, J.L. (1980) Childhood Education, 57, 97-100.

Hall, Edwin P. (1969). Unpublished doctoral dissertation, University of Alabama.

Hall, Ernest L., & Mueller, Betty S. (1968). Nursing Research, 17, 172-174.

Hanson, Marvin L. (1968). Final Report to U. S. Department of Health, Education, and Welfare. (ERIC No. ED 020 659)

Harasymiw, S.J. Horne, M.D., & Lewis, S.C. (1976). Rehab Literature, 37, 98-102.

Harasymiw, Stefan J., & Horne, Marcia D. (1975). Education, 96, 153-158.

Harold, J. Gary. (1980). Unpublished doctoral dissertation, The University of North Carolina at Greensboro.

Harth, Robert. (1966). Exceptional Children, 33, 119-120.

Hartlage, Lawrence C. (1974). Interamerican Journal of Psychology, 8, 25-28.

Hastorf, Albert H.; Wildfogel, Jeffrey; & Cassman, Ted. (1979). Journal of Personality and Social Psychology, 10, 1790-1797.

Haynes, Robert Clayton. (1969). Dissertation Abstracts International, 30, 4307-A.

Hazzard, Ann. (1983). The Journal of Special Education, 17(2), 131-139.

Hedberg, Sally. (1981). Academic Therapy, 16, 559-562.

Hennigan, Donald James. Dissertation Abstracts International.

Herr, David E. (1975). Exceptional Children, February.

Herr, David E., Algozzine, Bob, & Eaves, Ronald C. (1976). Journal of Education Research, 69, 261-264.

Hersch, Paul D., Kulik, James A., & Scheibe, Karl E. (1969). Journal of Consulting and Clinical Psychology, 33(1), 30-34.

Hess, Mari Anne. (1985). Unpublished master's thesis, Pacific Lutheran University. (Masters Abstracts, 23(3))

Hewett, Frank M. (1980, March). Los Angeles: California University, Graduate School of Education.

Higgs, R. W. (1971). Dissertation Abstracts International, 32, 4450-A.

Higgs, Richard W. (1975). Exceptional Children, 41, 496-497.

Himes, Joseph S. (1958). The New Outlook. 330-335.

Home, M. (1982). Exceptional Children, 29(2), 79-86.

Hoover, J.J. (1984). Journal of Research and Development in Education, 18, 33-39.

Hopkins. (1980). Elementary School Journal, 81 (1), 34-39.

Horne, Marcia D. (1983). The Exceptional Child, 30, 93-98.

Horowitz, Leola S., & Rees, Norma S. (1962). The Volta Review, 64, 180-188.

Huddle, D. D. (1971). Mental Retardation, 9(2), 33-35.

Hurtado-Portillo, Jose. (1980). Unpublished doctoral disseration.

Husek, T. R. (1965). Journal of Personality and Social Psychology, 2(1), 125-128.

Iano, Richard P., Ayers, Dorothy, Heller, Howard, McGettigan, James F., & Walker, Valaida S. (1974). Exceptional Children, 267-271.

Iguchi, Melvin T., & Johnson, Ronald C. (1966). The Journal of Social Psychology, 68, 107-111.

Jackson, Robert Donald. (1976). Unpublished doctoral dissertation, University of Virginia.

Jacobson, J. W., et. al. (1983). Education, 104(1), 29-33.

Jacobson, William H. (1984). Unpublished doctoral dissertation, University of Arkansas.

Jaffe, Jacob. (1967). Journal of Counseling Psychology, 14(5), 482-484.

Jamison, Harry Joseph, Jr. (1978). Unpublished doctoral dissertation, University of Pennsylvania.

Johnson, Alex B. (1979). The Journal of Special Education, 13(4), 453-462.

Johnson, G.O. (1950). AJMD, 55, 60-89.

Johnson, David W., & Johnson, Roger T. (1982). The Journal of Social Psychology, 118, 257-268.

Johnson, David W., & Johnson, Roger T. (1981). Contemporary Educational Psychology, 6, 344-353.

Johnson, David W., & Johnson, Roger T. (1984). The Journal of Social Psychology, 124, 85-94.

Johnson, Roger T., & Johnson, David W. (1981). American Educational Research Journal, 18(4), 415-423.

Johnson, Roger T., & Johnson, David W. (1983). Exceptional Children, 49, 323-329.

Johnson, Roger T., & Johnson, David W. (1982). The Journal of Social Psychology, 116, 211-219.

Johnson, Roger, Johnson, David W., DeWeerdt, Nancy, Lyons, Virginia, & Zaidman, Brian. (1983). American Journal of Mental Deficiency, 87, 611-618.

Johnson, Roger; Rynders, John; Johnson, David W.; Schmidt, Bernie; & Haider, Sajjad. (1979). American Educational Research Journal, 16(2), 161-167.

Jones, Reginald L., Lavine, Karen, & Shell, Joan. (1972). The New Outlook for the Blind, 66, 75-80.

Jones, Reginald L., Marcotte, Marcia, & Markham, Karen. (1968). Exceptional Children, 34, 309-315.

Jordan, John & Proctor, Doris I. (1969). The Journal of Special Education, 3(4), 433-439. 3A

Karniski, Marsha A. Perkins. (1978). Rehabilitation Counseling Bulletin. 22(1), 16-20.

Kearney, Nolan C., & Rocchio, Patrick D. (1956). Journal of Educational Research, 49, 703-8.

Keith-Spiegel, Patricia. (1967). Paper presented at the California Psychological Association, Coronado.

Keith-Spiegel, Patricia, & Spiegel, Donald E. (1970). Journal of Clinical

Psychology, 26, 387-388.

Kilburn (1984). Teaching Exceptional Children, 16(2), 124-127.

Kimbrell, Don L., & Luckey, Robert E. (1964). American Journal of Mental Deficiency, 69, 21-22.

Kitano, Margie K. (1976). Dissertation Abstracts International, 37, 5038-5039.

Kitano, Margie, & Chan, Kenyon S. (1978). American Journal of Mental Deficiency, 83, 37-39.

Kolmer, Marietta Babcock, & Kern, Howard M. (1968). American Journal of Psychiatry, 125, 698-702.

Kushman, Karen Michelle. (1976). Unpublished doctoral dissertation, Boston University.

Lapp, E.R. (1957). American Journal of Mental Deficiencies, 62, 254-262.

Larrivee, Barbara. (1981). Exceptional Children, 48, 34-39.

Larrivee, Barbara & Cook, Linda. (1979). The Journal of Special Education, 13, 315-324.

Lazar, A., Gensley, J., & Gowan, J. (1972). Gifted Child Quarterly, 16, 27-31.

Lazar, Alfred L., Haughton, Donna, & Orpet, Russel. (1977). Behavioral Disorders, 2(2), 85-88.

Leggett, Stephen L. (198[illegible]). Unpublished doctoral dissertation, The University of Florida.

Lehrer, Ariella. (1983). The Journal of Educational Research, 77, 95-99.

Lewis, Juliet. (1972). Mental Retardation, 10(6), 14-15.

Lewis Rena B. & Johnson Marilyn Kay. (1982). Teacher Education and Special Education, 5, 15-18.

Leyser, Y. & Abrams, P.D. Clearing House, 57(6), 250-255.

Leyser, Yona, Abrams, Peter, & Lipscomb, Edra. (1972). Journal for Special Educators, 18(4), 1-10.

Leyser, Yona, & Gottlieb, Jay. (1980). Exceptional Children, 46, 459 [illegible]61.

Leyser, Yona, Johansen, John H., & Abrams, Peter D. (1984). Teacher Education Quarterly, 11, 64-71.

Lieberth, Ann K. (1982). Volta Review, 84, 291-295.

Lillis, Linda, & Wagner, Robert M. (1977). Rehabilitation Literature, 38, 358-363.

Lilly, Stephen M. (1971). Exceptional Children, 37, 341-347.

Linkowski, Donald C., & Dunn, Marilyn A. (1974). Rehabilitation Counseling Bulletin, 28-33.

Lombardi, Thomas, P., Meadowcroft, Pamela, & Strasburger, Richard. (1982). Exceptional Children, 48, 544-545.

Long, Ralph S., Jr. (1963). Journal of Clinical Psychology, 19, 338-341.

Lundstedt, Sven, & Lillibridge, John. (1966). Mental Hygiene, 50, 432-438.

Madden, Nancy A., & Slavin, Robert E. (1983). The Journal of Special Education, 17, 171-182.

Mandel, David Marc. (1968). Dissertation Abstracts International, 28, 4004-A.

Manning, M. Lee. (1982). Middle School Journal, 13(2), 12-13.

Marlowe, Michael J. (1978). Dissertation Abstracts International.

Marlowe, Michael J. (1979). Education and Training of the Mentally Retarded, 14, 262-268.

Martin, David S. (1984). Volta Review, 86, 152-157.

Martino, Linda, & Johnson, David W. (1979). The Journal of Social Psychology, 107, 177-183.

Mason, Walter C. (1983). Unpublished doctoral dissertation, George Peabody College for Teachers of Vanderbilt University.

Matovu, H. L. (1974). Social Science & Medicine, 8, 47-50.

Mauer. (1979). Exceptional Children, 45(5), 326-330.

Mayers, Friedericka. (1976). The New Outlook for the Blind, 70(2), 72-77.

Mazer, Gilbert E. (1969). The Journal of Educational Research, 63(3), 116-120.

Mba, P. O. (1978). Eric. No. ED 158 539.

McAleer, Charles A. (1972). Unpublished doctoral dissertation, University of Missouri-Columbia.

McDaniel, Clyde O., Jr. (1970). Education and Training of the Mentally Retarded, 5, 4-14.

McGinnies, Elliott. (1958). Mental Hygiene, 42, 409-422.

McGinnies, Elliott; Lana, Robert; & Smith, Clagett. (1958). Journal of Applied Psychology, 42, 40-46.

McRainey, Gay. (1981). Unpublished doctoral dissertation, George Peabody College for Teachers of Vanderbilt University.

Melton, Gary B. (1977). *Journal of Clinical Child Psychology*, *6*(3), 89-90.

Melton, G.P. (1980). *Journal for Special Educators* *16* (2), 198-204.

Meltzer, M. L., & Grigorian, H. M. (1972). *Psychiatry*, *35*, 195-204.

Miller, K. M., & Biggs, J. B. (1958). *Journal of Educational Psychology*, *49*, 224-228.

Miller, Louise B., & Erwin, Edmond F. (1961). *Journal of Medical Education*, *36*, 422-428.

Miller, Sidney L., & Heil, James. (1976). *Journal of Dental Education*, *40*, 740-744.

Mims, Aquilla Ann Kirklan. (1981). Unpublished doctoral dissertation, University of Alabama.

Mittler, P. (1970). *British Journal of Mental Subnormality*, *16*(1), 3-8.

Monroe, Jay Donald. (1968). *Dissertation Abstracts International*, *29*, 1771-A.

Monroe, J. Donald. (1971). *Education and Training of the Mentally Retarded*, *6*, 20-24.

Morgan, B. & Leung, P. (1980). *Journal of Counseling Psychology*. *27* 209-212.

Monson, Dianne. (1979). *Early Years*, *9*, 38-39.

Morrison, James K. (1976). *Journal of Community Psychology*, *4*, 181-185.

Morrison, James K. (1977). *Journal of Clinical Psychology*, *33*, 549-551.

Morrison, James K., & Becker, Roy E. (1975). *Journal of community psychology*, *3*, 281-84.

Morrison, James K., Becker, Robert E., & Bourgeois, Christine A. (1979). *Psychological Reports*, *44*, 855-859.

Morrison, James K., & Nevid, Jeffrey S. (1976). *Journal of Family Counseling*, *4*(1), 43-48.

Morrison, James K., & Teta, Diana C. (1977). *Journal of Clinical Psychology*, *33*, 1128-1131.

Morrison, James K., & Teta, Diana C. (1980). *Journal of Clinical Psychology*, *36*, 275-276.

Mortz, M. (1977). University of Michigan. *DAI* *38*, 4110A-4111A.

Mortz, D. M. J., & Sister, Mary Eileen. (1977). *An investigation of a faculty controller inservice training model in attitudes and knowledge* related to handicapped children in regular classrooms. Unpublished doctoral dissertation, University of Kansas.

Murray, Joan Fell. (1958). Mental Hygiene, 42, 402-408.

Neale, Marie D., Cuming, Jocelyn M., & Thompson, G. B. (1981). B. C. Journal of Special Education, 5, 237-258.

Nevin, Ann, Johnson, David W., & Johnson, Roger. (1982). The Journal of Social Psychology, 116, 41-59.

Nolan, M. (1980). Previews, 8(9), 43.

Norlander, Kay Annette. (1984). Unpublished doctoral dissertation, University of Connecticut.

Odom, Samuel L. (1982). Unpublished doctoral dissertation, University of Washington.

Orlansky, Michael D. (1977). Unpublished doctoral dissertation, University of Idaho.

Orlansky, Michael D. (1979). Exceptional Children, 46 49-52.

Overline, H. M. (1972). Unpublished doctoral dissertation, Temple University.

Palmer, Douglas J. (1980). Mental Retardation, 18, 171-175.

Palmerton, Keith E.; Frumkin, R. M. (1969). Perceptual and Motor Skills. 28, 489-490.

Parrish, Linda H. (1982, December). Paper presented at the American Vocational Association Convention.

Patterson, David L. (1977). Unpublished doctoral dissertation, Georgia State University.

Pavone-Kennedy, Margaret. (1983). Unpublished doctoral dissertation, Rutgers University The State University of New Jersey.

Payne, Reed, & Murray, Charles. (1974). Exceptional Children, 40, 123-125.

Perry, Howard Louis. (1979). Unpublished doctoral dissertation, George Washington University.

Philipp, Constance. (1977). Unpublished doctoral dissertation, University of California, Berkeley.

Pinckney, George A. (1962). Journal of Educational Psychology, 53, (6), 275-278.

Pohl, Sheila. (1978). Dissertation Abstracts International, 39(10), 6071-A.

Popp (1983). Teaching Exceptional Children, 15(2), 78-81.

Posner, Bernard. (1972). Mental Retardation, 23(1), 15-18.

Pratt, George W. (1977, November). Paper presented at the Annual National Conference on Urban Education, Norfolk, VA.

Pratt, Sarah Evelyn. (1978). Dissertation Abstracts International, 39(4), 2040-A.

Preininger, David R. (1968). Social Work, 13, 75-77.

Prillaman, Douglas. (1981). Journal of Learning Disabilities, 14, 344-346.

Rapfogel, Richard Gerson. (1976). Dissertation Abstracts International, 37(5), 2523-B.

Reese-Dukes, Judson ., & Stokes, Elizabeth H. (1978). Education and Training of the Mentally Retarded, 13, 356-361.

Reichel, Elizabeth A. (1975). Journal of Applied Rehabilitation Counseling, 6, 188-192.

Richmond, Naomi Long. (1984). Unpublished doctoral dissertation, University of Urbana-Champaign. (DAI, 45(11), 3324-A)

Roeher, Godfrey Allan. (1959). Unpublished doctoral dissertation, New York University. (DAI, 20, 3134-A)

Rose, C. D. (1979). B. C. Journal of Special Education, 3, 399-411.

Rosenberg, Harry. (1976). Special Education Forward Trends, 3(2), 8-9.

Rucker, Chauncy Newton. (1968). Dissertation Abstracts International, 28, 3038-A.

Rucker, Chauncy N., Howe, Clifford E., & Snider, Bill. (1969). Exceptional Children, 617-623.

Rucker, Chauncy, & Vincenzo, Filomena M. (1970). Exceptional Children, 36, 679-680.

Rusalem, Herbert, & Rusalem, Roslyn. (1964). The New Outlook for the Blind, 58, 260-263.

Ruud, J. B. (1971).

Salend, Spencer J. (1981). Education Unlimited, 3, 40-42.

Salkind, Isadore. (1962). Journal of Rehabilitation. 28.

Salvia, J., Clark, G., Ysseldyke, J. (1973). Exceptional Children, 39, 651-652.

Sanche, Robert, Chapman, Don, & Dineen, Tana. (1976). Saskatchewan Journal of Educational Researchand Development, 7, 25-28.

Sanche, Robert, Haines, Leonard, & Van Hesteren, Frank. (1982). B.C.

Journal of Special Education, 6 (1), 69-77.

Sandler, Allen, Coren, Andrea, & Thurman, S. Kenneth. (1983). Exceptional Children, 49, 355-358.

Santomier, James, & Kopczuk, Walter. (1981). Learning of the Mentally Retarded, 16, 20-23.

Scheffelin, Margaret Merrick. (1982, February). Paper presented at the International Conference of the Association for children with Learning Disabilities.

Schorn, Frank Roger. (1976). Unpublished doctoral dissertation, University of Massachusetts.

Scofield, Pape, McCracken & Maki. (1980). Journal Applied Rehab Counseling. 11, 183-187.

Scranton, T., & Ryckman, D. (1979). Journal of Learning Disabilities, 12, 402-407.

Seitz, Sue B., & Cleland, Charles C. (1967). Psychcological Reports, 20, 51-54.

Sellin, Donald, & Mulchahay, Ray. (1965). American Journal of Mental Deficiency, 70, 408-412.

Seltzer, Mildred M. (1977). Educational Gerontology: An International Quarterly. 2 183-189.

Semmel, Melvyn I., & Baum, R. Bruce. (1973). Viewpoints, 49(6), 59-69.

Shaw, Stan F., & Gillung, Tom B. (1975). Mental Retardation, 13, 3-6.

Sheare, Joseph B. (1974). American Journal of Mental Deficiency, 73, 678-682.

Sheare, Joseph B. (1978). Psychology in the Schools, 15, 406-412.

Shewell, Betty. (1974). Dissertation Abstracts International, 35, 199-200.

Shum, Katherine T. I. (1982). Unpublished doctoral dissertation.

Simpson, Richard L. (1980). Focus on Exceptional Children, 13(3), 1-11.

Singleton, Karin. (1976). Final Report to the California State Department of Education, Sacramento Division of Special Education.

Siperstein, Gary N., Bak, John J., & Gottlieb, Jay. (1977). Journal of Educational Research, 70, 131-134.

Siperstein, Gary N., & Chatillon, Alice C. (1982). American Journal of Mental Deficiency, 86, 453-458.

Siperstein, G. N., & Goding, M. J. (1985). Journal of Learning Disabilities, 18, 139-144.

Slavin, Robert E., Maddin, Nancy A., & Leavey, Marshal. (1982). Eric. No. ED 223 046.

Smith, Gloria J. (1983). Unpublished doctoral dissertation, Michigan State University.

Smith, Karl, Johnson, David W., & Johnson, Roger. (1982). The Journal of Social Psychology, 116, 277-283.

Smith, Sterling, & Ogley, Peter. (1981). Australian Journal of Developmental Disabilities, 7(2), 87-89.

Smyth, Veronica, & Thomas, David R. (1971). Psychological Reports, 28, 34.

Soloway, Michael M. (1974). Unpublished doctoral dissertation, University of California, Los Angeles.

Spencer, Lona Davis. (1977). Dissertation Abstracts International, 38(5), 2350-B.

Stainbach & Stainbach. (1982). Education and Training of the Mentally Retarded. 17 88-92.

Stainbach, Stainbach, Strathe, & Didrich. (1983). Ed. and Training of the Mentally Retarded, 18, 204-209.

Stauffer. (1974). Physical Therapy, 54, 1084-1085.

Steer, Michael. (1979). Education Canada, 19, 26-31.

Steingisser, Edith R. (1954). Unpublished master's thesis, University of New Hampshire.

Stiefel, Doris J. & Edmond L. Truelove. Journal of Dental Education, 49, 85-90.

Stiehl, J. Dissertation Abstracts International.

Stotsky, Bernard A., & Rhetts, John E. (1966). Nursing Research, 15, 175-177.

Strasburger, Richard J. (1983). Unpublished doctoral dissertation, West Virginia University.

Sumlin, Donna Lee. Dissertation Abstracts International, 39(7), 4186-A.

Sutherland, John H., Algozzine, Bob, Ysseldyke, James E., & Freeman, Sheryl. (1983). Journal of Learning Disabilities, 16, 217-220.

Szuhay, J.A. (1961). Dissertation Abstracts International

Taft, Susanne. (1980). Unpublished doctoral dissertation, University of Cincinnati.

Thatts, W. (1967). Journal of Personality and Social Psychology, 5, 4-15.

Threlkeld, R. M., et al. (1979). Changing attitudes to the disabled; the invisible barriet; Project match. Report (Grant No. 12-P-59142 from NIH).

Threlkeld, Robert M., & DeJong, William. (1982). Rehabilitation Counseling Bulletin, 25, 282-85.

Udin, Richard D., & Kuster, Curtis G. (1985). Journal of Dental Education, 49(2), 111-114.

Vander Kolk, Charles J., & Bright, Bobra C. (1983). Journal of Visual Impairment and Blindness, 77(2), 49-51.

Vaughan, Richard William. (1971). Unpublished doctoral dissertation, University of Northern Colorado.

Voeltz, Lranna M. (1980). Education Unlimited, 2(1), :~-11.

Waggoner, Louadah. (1976). Unpublished doctoral dissertation, Texas Tech University.

Waggoner, L. G. (1976). Unpublished doctoral dissertation, Texas Tech. DA-38:741 A.

Walker, C.E. Journal of Social Psychology, 78, 255-261.

Walsh, Joan E. (1971). Nursing Research, 20(6), 522-529.

Warren, Sue Allen, Turner, Dale R., & Brody, David S. (1964). Mental Retardation, 2(4), 235-242.

Webster, Sanford, Freedman. (1955). Journal of Psychology, 40 73-84.

Weinberg, Nancy. (1976). Rehabilitation Psychology, 23(4), 115-124.

Weinberg, Nancy. (1978). Rehabilitation Counseling Bulletin, 22, 114-124.

Weinberg, Nancy. (1978, August). Paper presented as part of a symposium on Critical Dimensions of Interaction between able-bodied and disabled at the National Convention of the American Psychological Association. Toronto, Canada.

Wesolowski, Michael D., & Deichmann, John. (1980). Rehabilitation Counseling Bulletin, 23, 218-226.

West, Jessie Moore. (1976). Unpublisned doctoral dissertation, East Texas State University.

West, M.A. (1980). Paper presented at the Council for Exceptional Children's Annual Conference in Philadelphia.

Westervelt. (1983). Journal of Abnormal Child Psychology, 11, 327-343.

Westervelt, Van Dien, & Turnbull, Ann P. (1980). Physical Therapy, 60, 896-900.

White, J. A. (1972). Dissertation Abstracts International, 32, 4457-A.

White, W. (1973). Unpublished doctoral dissertation, Boston University.

Whiteman, Martin, & Lukoff, Irving F. (1962). New Outlook for the Blind, 56(5), 153-8.

Willard, Derek, H., Logan, Henrietta, & Simpson, Roger. (1978). Journal of Dentistry for the Handicapped, 3(2), 5-8.

Winzer, Margret. (1984). B.C. Journal of Special Education, 8, 273-282.

Wright, Beatrice. (1973). Rehabilitation Literature, 12.

Wright. Rehabilitation Psychology.

Wyder, Frank T., Wilson, Milton E., Jr., & Frumkin, R. M. (1967). Perceptual and Motor Skills, 25, 188.

Wyeth. (1965). American Annals of the Deaf, 110, 479-482.

Yates, James R. (1973). Exceptional Children, 39, 471-472.

Zanna, M., Kiesler, C., & Pilkonis, P. (1970). Journal of Pers. and Psych., 14, 321-328.

Zuckerman, Robert A. (1975). Unpublished doctoral dissertation, Indiana University.

Discarded Reports—Lack of Information*

Allen, Benjamin H.; Shaw, Carl Neil; and Foshee, James G. (1967). Student work programs in mental retardation facilities. (Report to the Division of Mental Retardation, Social Rehabilitation Services, Grant No. MR1014A67[T]. Tallahassee, FL: Department of Special Education and Rehabilitation, Florida State University.

Altrocchi, John, & Eisdorfer, Carl. (1961). Changes in attitudes toward mental illness. Mental Hygiene, 45, 563-570.

Anderson, Luleen S.; Del-Val, Patricia B.; Griffin, Carol Lee; & McDonald, Joan. (1983, April). Understanding handicaps: Changing students' attitudes toward the disabled. Paper presented at the Annual Meeting of the American Orthopsychiatric Association, Boston. (ERIC No. ED 232 366)

Appell, Melville J.; Williams, Clarence M.; & Fishell, Kenneth N. (1964). Changes in attitudes of parents of retarded children effected through group counseling. American Journal of Mental Deficiency, 68, 807-812.

Baran, Stanley J. (1979). Television programs about retarded children and parental attitudes toward their own retarded children. Mental Retardation, 17, 193-194.

Brummel, Colette. (1984). Changes in non-handicapped children's attitudes toward trainable retarded children using two methods: Buddy interacting and peer tutoring. Unpublished doctoral dissertation, University of Northern Colorado.

Cleland, Charles C., & Cochran, Irene L. (1961). The effect of institutional tours on attitudes of high school seniors. American Journal of Mental Deficiency, 65, 473-481.

Cloerkes, Gunther, & Wohrl, Hans-Georg. (1982). Modification of young people's attitudes toward handicapped persons by means of information presented at a theatrical performance. International Journal of Rehabilitation Research, 5, 543-552.

Fox, Robert A.; Lopuch, Wanda R.; & Fisher, Eve. (1984). Using volunteer work to teach undergraduates about exceptional children. Teaching of Psychology, 11, 113-115.

Gelfand, Sidney, & Ullmann, Leonard P. (1961). Attitude change with psychiatric affiliation. Nursing Research, 10, 200-204.

---

*These studies were relevant to the review topic, but data to compute effect sizes and information on the statistical significance of results were missing from the report.

Gibson, Beryl S., & Reed, John C. (1974). Training nurses in MR. Mental Retardation, 12, 19-22.

Godejohn, Carol Jean; Taylor, Jacqueline; Muhlenkamp, Ann; & Blaesser, Willard. (1975). Effect of simulation gaming on attitudes toward mental illness. Nursing Research, 24(5), 367-370.

Gorelick, Molly C.; Joseph, Marjory, L.; & Friedman, Loretta J. (1975). Careers in integrated early childhood programs. (Report to the U. S. Department of Health, Education, and Welfare, Social and Rehabilitation Service, Grant No. 55-P-45144/9-04). Northridge, CA: California State University. (ERIC No. ED 112 628)

Gutierrez, J. L., & Ruiz, J. Saiz. (1978). A comparative study of the psychiatric nurses' attitudes towards mental patients. International Journal of Social Psychiatry, 24, 47-52.

Hersh, Alexander; Carlson, Raymond W.; & Lossino, David A. (1977). Normalized interaction with families of the mentally retarded--To introduce attitude and behavior change in students in a professional discipline. Mental Retardation, 15(1), 32-33.

Iverson, Maynard J., & Davis, Paul D. (1980, March). The effects of selected inservice education methods on the attitudes of vocational education instructors regarding teaching students with physical disabilities. Auburn University, Department of Vocational and Adult Education, Auburn, AL.

Iverson, Maynard J., & Davis, Paul D. (1981). Effects of inservice education methods on improving the attitudes of vocational agriculture teachers toward physically handicapped students. The Journal of Vocational Education Research, 6(1), 1-16.

Jaffe, Jacob. (1972). The effects of work conferences on attitudes towards the mentally retarded. Rehabilitation Counseling Bulletin, 15, 220-227.

Koch, Audrey Straus. (1975). Changing attitudes toward blindness: A role-playing demonstration for service clubs. The New Outlook for the Blind, 69(9), 407-409.

Kuchan, Anthony Mark, Jr. (1964). Psychiatric nursing experience and group discussion as variables affecting attitudes toward the mentally ill. Unpublished doctoral dissertation, Purdue University.

Marsh, Velma, & Friedman, Robert. (1972). Changing public attitudes toward blindness. Exceptional Children, 38, 426-428.

Monson, Dianne, & Shurtleff, Cynthia. (1979). Altering attitudes toward the physically handicapped through print and non-print media. Language Arts, 56, 163-170.

Murthy, R. Srinivasa, & Arora, Manju. (1976). Attitude changes in medical postgraduates following "short-term" training in psychiatry. Indian Journal of Clinical Psychology, 3, 165-169.

Nunnally, Jum C., & Bobren, Howard M. (1959). Attitude change with false information. Public Opinion Quarterly, 23, 260-266.

Ovide, Christopher R. (1977). Affective development training and disability simulation with sighted children and its effects on interaction strain and attitudes toward visually handicapped peers. Unpublished doctoral dissertation, The College of William and Mary.

Penn, William W. (1972). The effect of a training program on the attitudes of elementary school administrators and teachers toward children with learning disorders. Unpublished doctoral dissertation, University of Pittsburgh.

Scheffers, Wenda L. (1977). Sighted children learn about blindness. Journal of Visual Impairment and Blindness, 71, 258-261.

Slater, Sherwood B., & Chyatte, Samuel B. (1970). Evaluation of effects on medical students of an introductory course in rehabiliation medicine. Archives of Physical Medicine and Rehabilitation, 5, 558-564.

Staffieri, Robert, & Klappersack, Bernard. (1960). An attempt to change attitudes toward the cerebral palsied. Rehabilitation Counseling Bulletin, 3, 5-6.

Steinzor, Luciana Visentini. (1966). School peers of visually handicapped children. The New Outlook for the Blind, 60, 312-314.

Vacher, Carole Doughton. (1973). Changes in knowledge, attitudes, and skills as a function of mental health consultation to physicians. Unpublished doctoral dissertation, North Carolina State University - Raleigh.

APPENDIX E

LIST OF REPORTS CODED

RESEARCH REPORTS CODED

0050[1]
Abernathy, Catherine L. (1980). Sex role and locus of control orientations, and attitude change. Unpublished doctoral dissertation, The Catholic University of America.

0245*
Aldridge, Jerry Titus II. (1978). Cognitive and affective gains of regular elementary educators from inservice education concerning nature of exceptional children. Unpublished doctoral dissertation, University of Alabama.

0001
Alese, Joseph A. (1973). Operation awareness. Mental Retardation, 11(5), 38-39.

0272
Allen, Benjamin H., & Allen, Jerry Carlton. (1968). Student work programs in mental retardation facilities. (Report to the Division of Mental Retardation, Social Rehabilitation Services, Grant No. 43970-01-68). Athens, GA: Department of Special Education, University of Georgia.

0269
Allen, Benjamin H., & Foshee, James G. (1966). Student work programs in mental retardation facilities. (Report to the Mental Retardation Division of the United States Public Health Service, Contract Nos. PH 2-1639, PH 201751, and PH 108-66-125 [P]). Tallahassee, FL: Department of Special Education and Rehabilitation, Florida State University.

0082*
Anastasiow, Nicholas J.; Everett, Michael; O'Shaughnessy, Thomas E.; Eggleston, Patricia J.; & Eklund, Susan J. (1978). Improving teenage attitudes toward children, child handicaps, and hospital settings: A child development curriculum for potential parents. American Journal of Orthopsychiatry, 48(4), 663-672.

0005
Anthony, William A. (1969). The effects of contact on an individual's attitude toward disabled persons. Rehabilitation Counseling Bulletin, 12, 168-171.

---

[1]Identification numbers were assigned arbitrarily for recordkeeping purposes as reports were queried for coding.

*Asterisks indicate reports which lacked data for computing effect sizes, but contained information on the statistical significance of results. An asterisk within parenththeses--(*)--indicates that at least one effect size and at least one "missing data" result were coded.

0007A[2] (Anthony & Carkhuff, 1970)
Anthony, William A. (1968). The effects of rehabilitation counselor training on attitudes toward the disabled and on the ability to communicate and discriminate the levels of facilitative conditions. Unpublished doctoral dissertation, State University of New York at Buffalo.

0007B (Anthony, 1968)
Anthony, William A., & Carkhuff, Robert R. (1970). Effects of training on rehabilitation counselor trainee functioning. Rehabilitation Counseling Bulletin, 13, 333-342.

0119*
Apolloni, Tony; Ruskus, Joan; & Triest, George. (1982). The effects of project interdependence. (Report to The Office of Special Education, California State Department of Education, Grant No. 49-03561-3066-82). Rohnert Park, CA: Sonoma State University, California Institute on Human Services. (ERIC No. ED 229 962)

0174*
Ashmore, Steven Michael. (1981). Inservice training of administrators: a study of attitude change. Unpublished doctoral dissertation, University of Wisconsin-Madison.

0137
Aukayanagul, Benja. (1979). Comparing effectiveness of the use of selected instructional techniques in modifying negative or stereotypic attitudes of nondisabled individuals toward the disabled in Thailand. Unpublished doctoral dissertation, University of Northern Colorado.

0068*
Austin, David R.; Powell; Lou G.; & Martin, Donald W. (1980). Modifying attitudes toward handicapped individuals in a classroom setting. The Journal for Special Educators, 17, 135-141.

0222
Avery, Russell M., & Davis, Paul D. (1983). The effect of anxiety producing simulation tasks on nondisabled preparatory vocational teachers' attitudes toward the physically handicapped vocational student. Journal of Vocational Education Research, 8(4), 1-10.

0128
Baker, Amanda Sirmon. (1974). Attitudes of nursing students toward mental retardation before and after curricular experience with mentally retarded children. Unpublished doctoral dissertation, University of Florida.

0203
Baker, Peter J.; Bussard, Lee; Johnson, Stella A.; & Rhodes, Clarence H. (1981). Student attitudes toward the handicapped. Edmonton: Planning and Research Branch, Alberta Department of Education.

---

[2] A letter following an identification number indicates that there were multiple reports of the same study, all of which were coded as one report. The author(s) and date(s) of the other report(s) are included in parentheses following the ID numbers.

0101*
Baran, Stanley J. (1977). TV programming and attitudes toward mental retardation. Journalism Quarterly, 54, 140-142.

0013
Bateman, Barbara. (1964). The modifiability of sighted adults' perceptions of blind children's abilities. The New Outlook for the Blind, 58, 133-135.

0040A (Beardsley, 1981-82)
Beardsley, Donna A. (1979). The effects of using fiction in bibliotherapy to alter the attitudes of regular third grade students toward their handicapped peers. Unpublished doctoral dissertation, University of Missouri-Columbia.

0040B (Beardsley, 1979)
Beardsley, Donna A. (1981-1982). Using books to change attitudes toward the handicapped among third graders. Journal of Experimental Education, 50, 52-55.

0017
Bitter, James A. (1963). Attitude change by parents of trainable mentally retarded children as a result or group discussion. Exceptional Children, 30, 173-177.

0032
Boehm, Helen L. (1978). The effect of contact with, and knowledge concerning, the severely and profoundly mentally retarded on the attitudes of special education majors. Unpublished doctoral dissertation, Teachers College, Columbia University.

0112A (Bond, 1980)
Bond, Carole L. (1979). Attitude change toward the physically disabled through the use of short stories. Unpublished doctoral dissertation, Arizona State University.

0112B (Bond, 1979)
Bond, Carole, L. (1980). Attitude change toward the physically disabled through the use of short stories. (Report to Arizona State University and Scottsdale Public Schools). Memphis, TN: Memphis State University, Department of Curriculum and Instruction. (ERIC No. ED 195 978)

0022
Brady, John Charles II. (1966). An investigation of the effects produced by a simulated experience with blindness on attitudes toward the blind. Unpublished master's thesis, Southern Illinois University.

0083*
Braff, Michael H., & Nealon, Linnea. (1982). Dental hygiene students and the developmentally disabled. Journal of Dental Education, 46, 709-712.

0208A (Brantlinger, 1983)
Brantlinger, Ellen A. (1978). Effects of in-service training on attitudes of residential care staff about the sexual needs and rights of mentally retarded residents. Unpublished doctoral dissertation, Indiana University.

0208B (Brantlinger, 1978)
Brantlinger, Ellen. (1983). Measuring variation and change in attitudes of residential care staff toward the sexuality of mentally retarded persons. Mental Retardation, 21(1), 17-22.

0024
Brighi, Robert J. (1978). The effects of integration on children's attitudes toward the physically handicapped. Unpublished doctoral dissertation, University of Iowa.

0111*
Brooks, Benjamin L., & Bransford, Louis A. (1971). Modification of teachers' attitudes toward exceptional children. Exceptional Children, 38, 259-260.

0268
Bryant, Cynthia A. (1985). The effects of a lecture training program and independent study on the knowledge and attitudes of law students toward the mentally retarded offender. Unpublished doctoral dissertation, University of Alabama.

0195
Bucich-Naylor, Denise A. (1978). The comparative effectiveness of a direct contact program and a didactic program in effecting changes in attitude of non-disabled children toward disabled children. Unpublished doctoral dissertation, Hofstra University.

0063
Butler, Joan M. (1981). Attitude change of second-grade students toward the handicapped through the use of children's books about the handicapped. Unpublished doctoral dissertation, Mississippi State University.

0066B (Elliott & Byrd, 1984)
Byrd, E. K., & Elliott, T. R. (1984). Attitude change toward disability through television portrayals with male college students. Paper presented at the Annual Meeting of the Southeastern Psychological Association.

0221
Cautela, Joseph R.; Walsh, Kenneth J.; & Wish, Peter A. (1971). The use of covert reinforcement in the modification of attitudes toward the mentally retarded. The Journal of Psychology, 77, 257-260.

0190
Cerreto, Mary C. (1976). The effects of empathy training on children's attitudes and behaviors toward handicapped peers. Unpublished doctoral dissertation, University of Washington.

0086*
Chinsky, Jack M., & Rappaport, Julian. (1970). Attitude change in college students and chronic patients: A dual perspective. Journal of Consulting and Clinical Psychology, 35, 388-394.

0142
Chow, Stanley H. L.; Rice, Carol F.; & Whitmore, Lynn A. (1976, April). Effects of a mediated training course on teachers and students in

mainstreaming programs. Paper presented at the meeting of the AERA, San Francisco. (ERIC No. ED 123 822)

0224
Clark, Evelyn Jane Manning. (1978). The efficacy of systematic desensitization as a strategy to affect attitude change in teachers toward severely handicapped children. Unpublished doctoral dissertation, University of Houston.

0113B* (Clore & Jeffrey, 1972)
Clore, Gerald L., & Jeffery, Katharine. (1971, May). Emotional role playing, attitude change, and attraction toward a disabled person. Paper presented at the Midwestern Psychological Association Convention, Detroit. (ERIC No. ED 050 393)

0113A* (Clore & Jeffrey, 1971)
Clore, Gerald L., & Jeffery, Katharine. (1972). Emotional role playing, attitude change, and attraction toward a disabled person. Journal of Personality and Social Psychology, 23(1), 105-111.

0156
Cobun, John M. (1972). Attitude changes in vocational rehabilitation counselors related to the physically disabled during induction preparation. Unpublished doctoral dissertation, The American University.

0220
Cohen, S[illegible]n F. (1978). The effect of systematic desensitization and information on attitudes toward the physically disabled. Unpublished doctoral dissertation, University of Rhode Island.

0183
Colasuonno, Thomas F. (1981). The use of contact with a deaf adult to alter the attitudes of teachers of the deaf and teachers in training toward deaf persons and toward the educational integration of hearing impaired persons. Unpublished doctoral dissertation, New York University.

0198*
Cole, Francis Conrad. (1970). Contact as a determinant of sighted persons' attitudes toward the blind. Unpublished doctoral dissertation, Florida State University.

0033
Cortez, Donna M. (1983). A study of the effects of an inservice program for postsecondary instructional faculty on mainstreaming handicapped college students. Unpublished doctoral dissertation, New Mexico State University.

0025
Costin, Frank, & Kerr, William D. (1962). The effects of an abnormal psychology course on students' attitudes toward mental illness. Journal of Educational Psychology, 53(5), 214-218.

0130
Cray-Andrews, Robert. (1981). A handicap awareness program helping nonhandicapped students develop more positive attitudes toward learning disabled individuals. Unpublished doctoral dissertation, Vanderbilt University.

0186
Cronk, Mildred S. (1978). Attitude change toward trainable mentally retarded: "Mainstreaming in reverse". Paper presented at the World Congress on Future Special Education, First, Stirling, Scotland. (ERIC No. ED 158 509).

0217
Crow, C. M.; Mowbray, R. M.; & Bloch, S. (1970). Attitudes of medical students to mental illness. Journal of Medical Education, 45, 594-599.

0026
Dahl, H. G.; Horsman, K. R.; & Arkell, R. N. (1978). Simulation of exceptionalities for elementary school students. Psychological Reports, 42, 573-574.

0184A (Dailey & Halpin, 1981)
Dailey, Jane, Leslie. (1978). Modifying undergraduates' attitudes toward the handicapped by means of video tapes. Unpublished doctoral dissertation, Auburn University.

0184B (Dailey, 1978)
Dailey, Jane L., & Halpin, Gerald. (1981). Modifying undergraduates' attitudes toward the handicapped by videotapes. The Journal of Special Education, 15(3), 333-339.

0246A (Dampier, Dancer, & Keiser, 1985)
Dampier, Kim. (1982). Comparison of an empathy tape to a lecture tape in changing the perceptions/feelings of college students toward older persons with hearing loss. Unpublished report, University of Arkansas, Medical Sciences, Little Rock.

0246B (Dampier, 1982)
Dampier, Kim; Dancer, Jess; & Keiser, Hope. (1985). Changing attitudes toward older persons with hearing loss: Comparison of two audiotapes. American Annals of the Deaf, 130, 267-271.

0147*
Daniels, Lloyd K. (1976). Covert reinforcement and hypnosis in modification of attitudes toward physically disabled persons and generalization to the emotionally disturbed. Psychological Reports, 38, 554.

0244(*)
Dickson, Patsy J. (1974). Project Understanding. Final Report to Hilroy Fellowship Program. Saskatoon, Saskatchewan: Children's Rehabilitation Centre, University Hospital.

0087*
Distefaro, M. K., Jr., & Pryer, Margaret W. (1970). Stability of attitudes in psychiatric attendants following training. Mental Hygiene, 54, 433-435.

0095*
Distefano, M. K., Jr., & Pryer, Margaret W. (1975). Effect of brief training on mental health knowledge and attitudes of nurses and nurses' aides in a general hospital. Nursing Research, 24, 40-42.

0029B (Pryer & Distefano, 1978)
Distefano, M. K., Jr., & Pryer, Margaret W. (1979). Follow-up of mental health attitudes of psychiatric aides after training. Psychological Reports, 45, 801-802.

0098*
Dixon, Calvert R. (1967). Courses in psychology and students' attitudes toward mental illness. Psychological Reports, 20, 50.

0031A (Donaldson, 1976; Donaldson & Martinson, 1977)
Donaldson, Joy. (1974). Effects of live, video, and audio presentations by a panel of physically disabled individuals on attitudes toward disabled persons. Unpublished doctoral dissertation, University of Kentucky.

0031B (Donaldson, 1974; Donaldson & Martinson, 1977)
Donaldson, Joy. (1976). Channel variations and effects on attitudes toward physically disabled individuals. Audio-Visual Communication Review, 24, 135-144.

0031C (Donaldson, 1974, 1976)
Donaldson, Joy, & Martinson, Melton C. (1977). Modifying attitudes toward physically disabled persons. Exceptional Children, 43, 337-341.

0185
Drake, Gregory, A. (1977, April). A comparative study of pre and post-semester attitudes toward the handicapped of students in introductory special education classes. Paper presented at the Annual International Convention of the the Council for Exceptional Children, Atlanta, GA. (ERIC No. ED 139 221)

0052
Dresang, Eliza T. (1981). Communication conditions and media influence on attitudes and information uses: The effects of library materials selected in response to student interests about mainstreaming and disabilities. Unpublished doctoral dissertation, University of Wisconsin-Madison.

0210A (Dubin & Fisher, 1982)
Dubin, Martin S. (1981). Attitude influence and attitude formation of nonhandicapped elementary school students as related to the mainstreaming of emotionally disturbed students. Unpublished doctoral dissertation, Vanderbilt University.

0210B (Dubin, 1981)
Dubin, Martin S., & Fisher, Maurice. (1982). Attitude influence and attitude formation of non-handicapped elementary school students as related to the mainstreaming of emotionally disturbed students. Paper presented at the Annual Meeting of the American Educational Research Association. (ERIC No. ED 217 663)

0197A* (Kierscht & DuHoux, 1980)
DuHoux, Mary A. (1979). Changing children's attitudes toward the disabled: Intervention and assessment. Unpublished master's thesis, Moorhead State University.

0034A (Dye, 1980)
Dye, Celeste A. (1978). Effects of persuasion and autotelic inquiry methods on attitude change. Perceptual and Motor Skills, 47, 943-949.

0034B (Dye, 1978)
Dye, Celeste A. (1980). Autotelic inquiry: A learning approach to attitude change. Educational Gerontology, 5, 239-248.

0056
Dyer, Larry L. (1970). Attitude changes in university students as a function of direct classroom contact with handicapped children. Unpublished doctoral dissertation, Indiana University.

0247*
D'Zamko, Mary E. (1979). Relationship of a program of diagnostic-prescriptive inse-- ice education and regular classroom teacher attitudes toward handicapped pupils. Unpublished doctoral dissertation, The University of Florida.

0201
Elliott, Timothy R., & Byrd, E. Keith. (1983). Attitude change toward disability through television portrayal. Journal of Applied Rehabilitation Counseling, 14(2), 35-37.

0066A (Byrd & Elliott, 1984)
Elliott, Timothy R., & Byrd, E. Keith. (1984). Attitude change toward disability through television: Portrayal with male college students. International Journal of Rehabilitation Research, 7, 320-322.

0037
Elliott, Timothy R., & Byrd, E. Keith. (1984). Video depictions of blindness and attitudes toward disability. Journal of Rehabilitation, 50(1), 49-52.

0151A (Emerton & Rothman, 1978)
Emerton, R. Greg, & Rothman, Gail. (1975). Attitudes of hearing students toward deafness on a deaf-hearing college campus. (Report to the U. S. Department of Health, Education, and Welfare). Rochester, NY: National Technical Institute for the Deaf, Rochester Institute of Technology.

0151B (Emerton & Rothman, 1975)
Emerton, R. Greg, & Rothman, Gail. (1978). Attitudes towards deafness: Hearing students at a hearing and deaf college. American Annals of the Deaf, 123(5), 588-593.

0039A (Esposito & Peach, 1981; Esposito & Reed, 1986)
Esposito, Beverly G., & Peach, Walter J. (1983). Changing attitudes of preschool children toward handicapped persons. Exceptional Children, 49(4), 361-363.

0039B (Esposito & Peach, 1983; Esposito & Reed, 1986)
Esposito, Beverly G., & Peach, Walter J. (1981). Changing attitudes of preschool children toward handicapped persons as a result of contact with severely handicapped peers. Unpublished manuscript.

0039C (Esposito & Peach, 1981, 1983)
Esposito, Beverly G., & Reed, Thomas M., II. (1986). The effects of contact with handicapped persons on young children's attitudes. Exceptional Children, 53(3), 224-229.

0124
Euse, Franklin John. (1975). An application of covert positive reinforcement for the modification of attitudes toward physically disabled persons. Unpublished doctoral dissertation, Auburn University.

0036
Evans, Joseph H. (1976). Changing attitudes toward disabled persons: An experimental study. Rehabilitation Counseling Bulletin, 19(4), 572-579.

0018
Farrell, Ann Marie. (1985). An investigation of attitudes toward the handicapped learner. Unpublished doctoral dissertation, Loyola University of Chicago.

0014
Feldman, David, & Feldman, Brian. (1985). The effect of a telethon on attitudes toward disabled people and financial contributions. Journal of Rehabilitation, 51(3), 42-45.

0038
Felton, Gary S. (1975). Changes in attitudes toward disabled persons among allied health paraprofessional trainees in an interdisciplinary setting. Perceptual and Motor Skills, 40, 118.

0226A (Lange & Fender, 1980)
Fender, Marilyn J. (1979). Modifying prospective teachers' attitudes toward the handicapped in a secondary pre-service program. Unpublished doctoral dissertation, Colorado State University.

0213
Fenrick, Nancy J., & Petersen, Todd K. (1984). Developing positive changes in attitude towards moderately/severely handicapped students through a peer tutoring program. Education and Training of the Mentally Retarded, 19, 83-90.

0253*
Fenton, Thomas Ransford. (1974). The effects of inservice training on elementary classroom teachers' attitudes toward and knowledge about handicapped children. Unpublished doctoral dissertation, University of Northern Colorado.

0117
Finell, Janet; Nathanson, Robert; & Lambert, Jeffrey. (1980). Regular Education Inservice training program. (Report to Division of Personnel Preparation, Bureau of Education for the Handicapped, U. S. Office of Education, Grant No. G007901258). Brooklyn, NY: Long Island University, Department of Guidance & Counseling. (ERIC No. ED 201 088)

0107
Finn, Therese E. (1979). Teacher inservice education to ease the mainstreaming process and enhance student development. Unpublished doctoral dissertation, Loyola University of Chicago.

0042
Fisher, Elaine B. (1975). An evaluation of a student teaching center approach in special education to developing teacher characteristics and attitudes at Kean Colle'e in New Jersey. Unpublished doctoral dissertation, Teachers College, Columbia University.

0094
Forader, Alvin Thomas. (1969). Modifying social attitudes toward the physically disabled through three different modes of instruction. Unpublished doctoral dissertation, University of Massachusetts.

0030
Fox, Robert A.; Lopuch, Wanda; & Rotatori, Anthony. (1984, August). Undergraduate attitudes towards the dev-lopmentally disabled: Impacts of volunteering. Paper presented at the 92nd Annual Convention of the American Psychological Association, Toronto, Ontario, Canada. (ERIC No. ED 253 988)

0062
Frazier, Robert S. (1975). Changing attitudes toward mental illness through systematic desensitization and information. Unpublished doctoral dissertation, University of Rhode Island.

0015
Freeman, Stephen W. (1970). Affective and cognitive changes in teacher trainees of the emotionally disturbed as a result of field experience with hospitalized adolescent psychiatric patients. Unpublished doctoral dissertation, The University of Tennessee.

0188
Friedel, Janice Ann Nahra. (1980). Changes in knowledge and attitudes of university students in secondary education after completion of a module in special education. Unpublished doctoral dissertation, University of Iowa.

0115
Friedman, Ronald S. (1975). The peer-peer program: A model project for the integr..ion of severely physically handicapped youngsters with nondisabled peers. (Report to New York State Education Department, Division of Drug and Health Education Services). Albertson, NY: Human Resources School. (ERIC No. ED 142 004)

0162*
Frith, Greg H., & Lindsey, Jimmy D. (1981). The effects of inservice training on regular educators' attitudes toward handicapped students. Education Unlimited, 29-31.

0041
Gelfand, Sidney, & Ullmann, Leonard P. (1961). Change in attitudes about mental illness associated with psychiatric clerkship training. International Journal of Social Psychiatry, 7, 292-298.

0260
Gerstein, Lawrence H. (1976). Attitudes toward disabled people as a function of open and closed mindedness and exposure to persuasive communications. Unpublished masters thesis, University of Georgia.

0126*
Giannotti, Thomas J., & Doyle, Robert E. (1982). The effectiveness of parental training on learning disabled chi. .ren and their parents. Elementary School Guidance and Counseling, 17, 131-136.

0044
Glauberman, Nancy R. (1980). The influence of positive T.V. portrayals on children's behavior and attitude toward the physically disabled. Unpublished doctoral dissertation, Teachers College, Columbia University.

0047
Gottlieb, Jay. (1980). Improving attitudes toward retarded children by using group discussion. Exceptional Children, 47, 106-111.

0048A (Graham, no date)
Graham, John R. (1968). Effects of introductory and abnormal psychology courses on students' attitudes toward mental illness. Psychological Reports, 22, 448.

0048B (Graham, 1968)
Graham, John R. (no date). The effects of introductory and abnormal psychology courses on students' attitudes toward mental illness. Unpublished paper, Lake Forest College, Lake Forest, IL.

0104
Granofsky, Jack. (1955). Modification of attitudes toward the visibly disabled. Unpublished doctoral dissertation, Yeshiva Univesity.

0096
Gronberg, Gary William. (1982). Attitude responses of nonhandicapped elementary students to specific information and contact with the handicapped. Unpublished doctoral dissertation, University of Northern Colorado.

0173
Grove, Michael Stephen. (1978). Social work intervention to reduce the stigma attached to mental illness by the families of psychiatric patients in Western Jamaica. Unpublished doctoral dissertation, Bryn Mawr College.

0131
Guldager, Lars. (1973). The impact of spot television announcements on attitudes towards deaf blind children. Unpublished doctoral dissertation, Boston College.

0069
Haddle, Harold William Jr. (1973). The efficacy of automated group systematic desensitization as a strategy to modify attitudes toward disabled persons. Unpublished doctoral dissertation, Georgia State University.

0051
Hafer, Marilyn, & Narcus, Margery. (1979). Information and attitudes toward disability. Rehabilitation Counseling Bulletin, 23(2), 95-102.

0061
Hagton, Elizabeth V. (1980). The effects of an inservice program upon a secondary school faculty's attitude toward and knowledge about handicapped students. Unpublished doctoral dissertation, University of Cincinnati.

0016
Hamilton, Edward J., & Anderson, Stephen C. (1983). Effects of leisure activities on attitudes toward people with disabilities. Therapeutic Recreation Journal, 50-57.

0102A* (Handlers & Austin, no date)
Handlers, Adele, & Austin, Katherine. (1980). Improving attitudes of high school students toward their handicapped peers. Exceptional Children, 47, 228-229.

0102B* (Handlers & Austin, 1980)
Handlers, Adele, & Austin, Katherine. (no date). Improving attitudes of high school students toward their handicapped peers. New Milford, CT: New Milford High School. Unpublished paper.

0105*
Harasymiw, Stefan J., & Horne, Marcia D. (1976). Teacher attitudes toward handicapped children and regular class integration. Journal of Special Education, 10, 393-400.

C179A (Haring, Stern, & Cruickshank, 1958)
Haring, Norris G. (1956). A study of the attitudes ot classroom teachers toward exceptional children. Unpublished doctoral dissertation, Syracuse University.

0179B (Haring, 1956)
Haring, Norris G.; Stern, George G.; & Cruickshank, William M. (1958). Attitudes of educators toward exceptional children. New York: Syracuse University Press.

0060
Harrell, Rona L. (1982). Attitude change toward the handicapped through participation in a short-term affective and cognitive program for nonhandicapped adolescents. Unpublished doctoral dissertation. University of Southern California.

0196
Hawisher, Margaret Frame. (1977). An evaluation of an experimental early childhood curriculum designed to create an accepting scholastic environment for the mildly physically handicapped youngster. Unpublished doctoral dissertation, University of South Carolina.

0045A (Hazzard & Baker, 1982)
Hazzard, Ann Petesch. (1981). Enhancing children's attitudes toward disabled peers using a multi-media curriculum. Unpublished doctoral dissertation, University of California, Los Angeles.

0045B (Hazzard, 1981)
Hazzard, Ann P., & Baker, Bruce L. (1982). Enhancing children's attitudes toward disabled peers using a multi-media intervention. Journal of Applied Developmental Psychology, 3, 247-262.

0073
Heitzman, Judy P. (1982). The effects of reading literature and teaching concept development on ninth graders' attitudes towards the mentally handicapped. Unpublished doctoral dissertation, Boston University.

0057
Hicks, Jack M., & Spaner, Fred E. (1962). Attitude change and mental hospital experience. Journal of Abnormal and Social Psychology, 65(2), 112-120.

0214
Hogan, Lois Ann Helton. (1983). Attitudinal development of parents with handicapped preschoolers. Unpublished doctoral dissertation, East Texas State University.

0167
Holmquist, Donna. (1981). Changing attitudes toward the handicapped. Delta Pi Epsilon, 23(2), 9-16.

0058
Holzberg, Jules D., & Gewirtz, Herbert. (1963). A method of altering attitudes toward mental illness. Psychiatric Quarterly Supplement, 37, 56-61.

0175
Horan, Diane Bunnel. (1981). Attitudes of non-disabled preschool and kindergarten students toward visibly-disabled and obese children and an investigation of an intervention to modify those attitudes: An empirical study. Unpublished doctoral dissertation, University of Iowa.

0020
Hourcade, Jack. (1981). Recreational setting as an agent for attitudinal change toward the mentally retarded. The Journal for Special Education, 17, 361-365.

0243
Hus, Yvette. (1979). The socialization process of hearing-impaired children in a summer day camp. Volta Review, 81, 146-156.

0204A (Ibrahim & Herr, 1982)
Ibrahim, Farah A. (1979). A comparison between an experimental technique and an information technique in modifying attitudes toward the disabled of undergraduate students in speech pathology and audiology. Unpublished doctoral dissertation, The Pennsylvania State University.

0204B (Ibrahim, 1979)
Ibrahim, Farah A., & Herr, Edwin L. (1982). Modification of attitudes toward disability: Differential effect of two educational modes. Rehabilitation Counseling Bulletin. 26(1), 29-36.

01??*
Ingram, Sylvia H. (1976). An assessment of regular classroom teachers' attitudes toward exceptional children subsequent to training in mainstreaming. Unpublished doctoral dissertation, University of Michigan.

0231
Jackson, Micheal Benoy. (1? 6). Comparison of attitudes and effects of brief academic exposure on attitudes toward mental retardation of northern, southern and southwestern Blacks. Unpublished doctoral dissertation, Ohio State University.

0092
Jacobs, Anita I. (1977). The efficacy of a university course on exceptional children and youth for non-special education majors in fulfilling the intent of Georgia House Bill 671. Unpublished doctoral dissertation, Georgia State University.

0165
January, Michael O. C. (1978). The effects of single-image and multi-image presentations upon attitudes toward disabled people. Unpublished doctoral dissertation, East Texas State University.

0110*
Johannsen, Walter J.; Redel, Sister M. Constance; & Engel, Ronald G. (1964). Personality and attitudinal changes during psychiatric nursing affiliation. Nursing Research, 13, 334-345.

0067
Jones, Thomas W.; Sowell, Virginia M.; Jones, Julie K.; & Butler, Lester G. (1981). Changing children's perceptions of handicapped people. Exceptional Children, 47(5), 365-368.

0157
Kauffman, Daniel J. (1976). Effects of contact and instruction on regular classroom teachers' attitudes toward the mentally retarded. Unpublished doctoral dissertation, University of Missouri.

0084
Kiernan, William Eugene. (1974). Attitude change in beginning education students resulting from exposure to the vocational and social capabilities of the disabled. Unpublished doctoral dissertation, Boston College.

0197B* (DuHoux, 1979)
Kierscht, Marcia S., & DuHoux, Mary A. (1980). Preparing the mainstream: Changing children's attitudes toward the disabled. School Psychology Review, 9, 279-283.

0009
Kisabeth, Kathryn L., & Richardson, Diane B. (1985). Changing attitudes toward disabled individuals: The effect of one disabled person. Therapeutic Recreation Journal, 19, 24-33.

0079
Klein, Jenny W. (1969). Maternal attitude changes in nursery schools for mentally retarded children. Unpublished doctoral dissertation, University of Maryland.

0122*
Klein, Malcolm W. (1963). An experimental evaluation of audio-visual methods: Changing attitudes toward education. (Report to U. S. Department of Health, Education, and Welfare, Grant No. OEG-704002). Los Angeles, CA: John Tracy Clinic. (ERIC No. ED 003 526)

0023
Kologinsky, Eileen Frances. (1981). An evaluation of an intervention designed to increase children's acceptance of physically handicapped peers. Unpublished doctoral dissertation, State University of New York at Stony Brook.

0116
Krieger, Mary M. W. (1978). Selected instructional techniques to create change of attitude toward blindness of pre- and inservice vocational home economics teachers. Unpublished doctoral dissertation, University of Michigan.

0261*
Kromer, Keith Alyn. (1976). The comparative effects of two staff development approaches upon regular class teachers' attitudes toward and their knowledge of handicapped students. Unpublished doctoral dissertation, University of Minnesota.

0074
Kulik, James A.; Martin, Robert A.; & Scheibe, Karl E. (1969). Effects of mental hospital volunteer work on students' conceptions of mental illness. Journal of Clinical Psychology, 25, 326-329.

0226B (Fender, 1979)
Lange, Robert R., & Fender, Marilyn J. (1980). Adult handicapped as resource persons for secondary teacher preservice programs. Journal of Teacher Education, 31, 49-54.

0159
Landis, Laura Gebiner. (1981). First semester special education majors, exposure and attitude change toward mentally retarded pupils. Unpublished doctoral dissertation, Fordham University.

0019
Lapp, Bernard. (1974). The effects of behavior modification inservice training on teacher behavior, teacher attitudes, and knowledge of behavior modification. Unpublished doctoral dissertation, University of Connecticut.

0076
Lazar, Alfred L.; Gensley, Juliana T.; & Orpet, Russel E. (1971). Changing attitudes of young mentally gifted children toward handicapped persons. Exceptional Children, 37, 600-602.

0077A (Lazar, Orpet, & Demos, 1976)
Lazar, Alfred L.; Orpet, Russell; & Demos, George. (1973, November). The impact of class instruction in changing student attitudes. Paper presented at the 52nd Annual Conference of the California Educational Research Association. (ERIC No. ED 086 702)

0077B (Lazar, Orpet, & Demos, 1973)
Lazar, Alfred L.; Orpet, Russell; & Demos, George. (1976). The impact of class instruction on changing student attitudes. Rehabilitation Counseling Bulletin, 20(1), 66-68.

0012B (Sippel, Lazar, & D'Alonzo, 1976)
Lazar, Alfred L.; Sippel, Harry; & D'Alonzo, Bruno J. (1977). Changing attitudes toward the TMR through an integrated recreation program. Mental Retardation, 15(4), 64.

0255
Lazar, Alfred L.; White, Ralph; & Sengstock, Wayne. (1975, April). A study of attitude change in special education majors in three university training programs. Paper presented at the International Council for Exceptional Children, 53rd Annual Conference. (ERIC No. ED 125 234)

0004
Lennington, Harold L. (1973). A study of a teacher development program and change in teacher attitudes toward exceptional children. Unpublished doctoral dissertation, University of Houston.

0262*
LeUnes, Arnold; Christensen, Larry; & Wilkerson, Dennis. (1975). Institutional tour effects on attitudes related to mental retardation. American Journal of Mental Deficiency, 79(6), 732-735.

0257
Leung, Esther Kau-To. (1979). Evaluation of a children's literature program designed to facilitate the social integration of handicapped children into regular elementary classrooms. Unpublished doctoral dissertation, Ohio State University.

0078
Levinson, Marc, & Distefano, M. K., Jr. (1979). Effects of brief training on mental health knowledge and attitudes of law enforcement officers. Journal of Police Science and Administration, 7(2), 241-244.

0064
Levinson, William S. (1973). The effects of a brief, selected, simulated experience with blindness on the attitudes of college students toward blindness. Unpublished doctoral dissertation, Southern Illinois University at Carbondale.

0125*
Lewis, I. Lauvine, & Cleveland, Sidney E. (1966). Nursing students' attitudinal changes following a psychiatric affiliation. Journal of Psychiatric Nursing, 4, 223-231.

0152
Leyser, Yona, & Abrams, Peter D. (1983). A shift to the positive: An effective programme for changing pre-service teachers' attitudes toward the disabled. Educational Review, 35(1), 35-43.

0010A (Lipsky, 1981)
Lipsky, Dorothy K. (1978). Modification of students' attitudes toward disabled persons. Unpublished doctoral dissertation, Hofstra University.

0010B (Lipsky, 1978)
Lipsky, Dorothy K. (1981, April). The modification of students' attitudes toward disabled persons. Paper presented to the Annual Meeting of the American Educational Research Association, Los Angeles. (ERIC No. ED 207 287)

0227
Lory, Nancy Savino. (1983). A study to improve the attitudes and behavior of young children toward handicapped persons. Unpublished doctoral dissertation, George Peabody College for Teachers of Vanderbilt University.

0055
Lowther, C. Leigh. (1978). Changing and maintaining attitudes of peers toward the integrated retarded pupil. Unpublished doctoral dissertation, University of Missouri-Columbia.

0118
Lundstrom, Karen V. (1979). Measuring attitude change toward special needs learners. Fargo: North Dakota State University, College of Home Economics. (ERIC No. ED 200 535).

0263
Maas, John M. (1983). Reducing prejudice by increasing intellectual and emotional understanding of human differences: Effects of generic and specific programs on fourth grader's attitudes. Unpublished doctoral dissertation, Boston University.

0163
MacIntyre, Rennett Domack. (1981). Modifying children's attitudes toward the handicapped through personal interaction. Unpublished doctoral dissertation, University of San Francisco.

0264
Marcus, Diana. (1979). The effect of a program of daily oral reading about the handicapped on the attitude and self-concept change of mainstreamed handicapped and non-handicapped fifth- and sixth-grade students. Unpublished doctoral dissertation, University of Connecticut.

0206
Margo, Bonnie C. (1983). Modifying attitudes toward physically handicapped children. Perceptual and Motor Skills, 56, 1002.

0035
Marquart, Jerry E. (1984). The effects of cognitive and affective training on the beliefs, feelings, behavioral intentions and behaviors of elementary students toward their disabled peers. Unpublished doctoral dissertation, University of San Francisco.

0223
Martinez, Manuel. (1977). An examination of the effects of three conditions on components of attitudes of student teachers toward selected types of disability. Unpublished doctoral dissertation, University of Nebraska.

0093
Mathey, John Paul. (1977). The effects of an inservice teacher training workshop for regular classroom teachers on their attitudes toward handicapped children. Unpublished doctoral dissertation, University of Michigan.

0229
Matkin, Ralph E.; Hafer, Marilyn; Wright, W. Russell; & Lutzker, John R. (1983). Pretesting artifacts: A study of attitudes toward disabilities. Rehabilitation Counseling Bulletin, 26(5), 342-348.

0215A (McDaniel, 1982)
McDaniel, Lindy C. (1980). Effectiveness of preservice and inservice programs in changing attitudes of vocational teachers working with mainstreamed handicapped students. Unpublished doctoral dissertation, Texas A&M University.

0215B (McDaniel, 1980)
McDaniel, Lindy. (1982). Changing vocational teachers' attitudes toward the handicapped. Exceptional Children, 48, 377-378.

0189*
McHale, Susan M., & Simeonsson, Rune J. (1980). Effects of interaction on nonhandicapped children's attitudes toward autistic children. American Journal of Mental Deficiency, 85(1), 18-24.

0120
McKerracher, William S. (1982). Changing children's attitudes toward the physically disabled. (Report No. ISBN-0-7241-4323-9). Curriculum Services Unit Education Department of Victoria. (ERIC No. ED 225 192)

0192*
Meehan, Michael John. (1980). The use of films in changing attitudes of sixth graders in a bilingual setting toward the disabled. Unpublished doctoral dissertation, University of San Francisco.

0216
Meltzer, Malcolm L., & Grigorian, Haikaz, M. (1972). Effect of psychiatric education on attitudes of medical students to mental illness. Psychiatry, 35, 195-204.

0265*
Messinger-Revell, Judith. (1983). A comparative analysis of the relative efficacy of two methods of increasing the attitude of acceptance in parents toward young special needs children. Unpublished doctoral dissertation, Boston University.

0181
Meyer, Peter. (1963). Effect of group counseling upon certain educative and emotional factors of first year students in an associate degree program in nursing. Unpublished doctoral dissertation, New York University.

0171*
Miller, Maurice; Armstrong, Susan; & Hagan, Michael. (1981). Effects of teaching on elementary students' attitudes toward handicaps. *Education and Training of the Mentally Retarded*, *16*, 110-113.

0154
Moosbruker, Jane B., & Giddon, Donald B. (1966). Effect of experience with aged, chronically ill, and handicapped patients on student attitudes. *Journal of Dental Education*, *31*, 278-286.

0071
Morris, Katherine D. (1964). A concomitant of attitude change in nursing students. *Nursing Research*, *13*(2), 132-138.

0100
Morrison, James [illegible].; Cocozza, Joseph J.; & Vanderwyst, Donna. (1978). Changing students' constructs of mental patients by means of educative seminars. *Journal of Clinical Psychology*, *34*, 482-483.

0266
Morton, Linda Ann. (1977). *Effecting attitudinal change toward physically disabled students in higher education.* Unpublished doctoral dissertation, Indiana University.

0135
Moseley, Mac Ray. (1973). *Attitude difference toward disabled persons as a function of educational integration.* Unpublished doctoral dissertation, University of Houston.

0088
Mulkey, Sidney Wayne. (1980). *Standard stimulus effects on attitudes toward disabled persons: An experimental study.* Unpublished doctoral dissertation, Florida State University.

0177(*)
Naor, Margalit, & Milgram, Roberta M. (1980). Two preservice strategies for preparing regular class teachers for mainstreaming. *Exceptional Children*, *47*, 126-129.

0053
Nelson, Otis Nicholas. (1970). *Positive attitude changes toward handicapped children and special education in a beginning educational psychology class.* Unpublished doctoral dissertation, University of Michigan.

0267*
Nemon, Alice. (1980). *The effects of a continuing education class on knowledge of, and attitudes and behavior toward, deaf persons.* Unpublished doctoral dissertation, University of California, Berkeley.

0232
Newcomer, Barbara L. (1975). Impact of education on nursing students' emotions, attitudes, and approaches to mental retardation. *Washington State Journal of Nursing*, *47*(2), 14-16.

0199A (Newman & Simpson, 1983)
Newman, Rebecca Kay. (1978). The effects of informational and experimental activities on the attitudes of regular classroom students toward severely handicapped children and youth. Unpublished doctoral dissertation, University of Kansas.

0199B (Newman, 1978)
Newman, Rebecca K. & Simpson, Richard L. (1983). Modifying the least restrictive environment to facilitate the integration of severely emotionally disturbed children and youth. Behavioral Disorders, 8(2), 103-112.

0140
Noe, Eva Jane. (1982). The effects of three instructional methods on knowledge and attitudes of undergraduate elementary and secondary education majors toward handicapped children. Unpublished doctoral dissertation, University of Missouri-Columbia.

0230
Nolan, Kathleen. (1982). Effects of the physical disabilities awareness curriculum on attitudes toward physically disabled persons. Unpublished doctoral dissertation, Saint Louis University.

0200(*)
Oberle, Judson B. (1975). The effect of personalization and quality of contact on changing expressed attitudes and hiring preferences toward disabled persons. Unpublished doctoral dissertation, Syracuse University.

0127
Olsen, Donald H. (1978). The effects of cross-age tutoring program on the reading achievement of mildly retarded student tutees and on the attitude of "normal" fifth and sixth grade elementary school student tutors toward retarded children. Unpublished doctoral dissertation, Case Western Reserve University.

0176
Owen, Sandra M. (1984). The enduring effects of active participation on positive attitude change toward handicapped students by sixth grade nonhandicapped students. Unpublished doctoral dissertation, University of Idaho.

0233
Ozyurek, Mehmet. (1977). Effects of live, audio, and print presentations of a discussion about physical disabilities on attitude modification toward disabled persons in Turkey. Unpublished doctoral dissertation, University of Northern Colorado.

0080
Pando, Joseph L. (1984). The use of short informational videotapes to modify attitudes toward disabled persons: An analysis of the effects of source credibility and message repetition. Unpublished doctoral dissertation, Hofstra University.

0211*
Parish, Thomas S.; Eads, Gerald M.; Reece, Nancy H.; & Piscitello, Mary A. (1977). Assessment and attempted modification of future teachers'

attitudes toward handicapped children. Perceptual and Motor Skills, 44, 540-542.

0180
Parish, Thomas S.; Nunn, Gerald D.; & Hattrup, Darlene. (1982). An attempt to reduce negative attitudes of future teachers toward exceptional children. College Student Journal, 16(3), 254-257.

0270
Parrish, Nancy Buchele. (1974). The effect of informational and integrational activities on the attitudes of regular classroom students toward children in special education. Unpublished master's thesis, University of Kansas.

0090
Paxton, Karen R. (1983). A comparison of the effectivness of three short-term instructional programs on the attitudes of fifth and sixth grade students toward the handicapped. Unpublished doctoral dissertation, University of Alabama.

0085
Perez, Katherine D. (1983). Changes in children's attitude toward handicapped peers as a result of an activity-based program. Unpublished doctoral dissertation, Brigham Young University.

0106
Perry, David Courtney. (1983). An attempt to modify attitudes of business leaders toward disabled persons. Unpublished doctoral dissertation, University of North Dakota.

0136
Peterson, Jo-Alyce Karen. (1977). Simulation activities as a means of changing attitudes and transmitting knowledge of adaptations and techniques for handicapped individuals to teacher trainees. Unpublished doctoral dissertation, Western Michigan University.

0134
Petrangelo, George Joseph. (1976). Attitudes of non-disabled college students toward their disabled classmates as a function of educational contact. Unpublished doctoral dissertation, University of Northern Colorado.

0003
Pomerantz, Robert M. (1983). The effectiveness of training modules designed to improve the attitudes of college faculty toward students with disabilities: An evaluation. Unpublished doctoral dissertation, Temple University.

0205*
Prothero, Jon C., & Ehlers, Walter H. (1974). Social work students' attitude and knowledge changes following study of programmed materials. American Journal of Mental Deficiency, 79, 83-86.

0029A (Distefano & Pryer, 1979)
Pryer, Margaret W., & Distefano, M. K., Jr. (1978). Effects of training on

the mental health attitudes of black and white psychiatric aides. Psychological Reports, 42, 1017-1018.

0091
Pryer, Margaret W.; Distefano, M. K., Jr.; & Marr, Larry W. (1969). Attitude changes in psychiatric attendants following experience and training. Mental Hygiene, 53, 253-257.

0065
Putnam, JoAnne W. (1983). Social integration of moderately handicapped students through cooperative goal structuring: Influence of teacher instruction on cooperation. Unpublished doctoral dissertation, University of Minnesota.

0114
Quay, Lorene C.; Bartlett, C. J.; Wrightsman, L. S., Jr.; & Catron, D. (1961). Attitude change in attendant employees. The Journal of Social Psychology, 55, 27-31.

0108
Rae, Rosamond. (1983). Attitudes of non-disabled children toward mentally retarded persons: The influence of personal interaction. Unpublished doctoral dissertation, Columbia University Teachers College.

0166
Rapier, Jacqueline; Adelson, Ruth; Carey, Richard; & Croke, Katherine. (1972). Changes in children's attitudes toward the physically handicapped. Exceptional Children, 39, 219-223.

0259
Reinhardtsen, Janet Marie. (1980). The effects of a structured informational program on the attitudes of regular high school students toward the handicapped. Unpublished doctoral dissertation, University of Kansas.

0164*
Richardson, H. Burtt, & Guralnick, Michael J. (1978). Pediatric residents and young handicapped children: Curriculum evaluation. Journal of Medical Education, 53(6), 487-492.

0109
Robertson, Leon Wayne. (1984). Evaluation of an inservice component to facilitate a productive staff attitude toward the disabled. Unpublished doctoral dissertation, Boston University.

0072
Robinson, Jean E. (1980). The effect of empathy training on the modification of attitudes and interpersonal effectiveness of learning disabled and non-handicapped persons. Unpublished doctoral dissertation, George Washington University.

0097
Rodriguez, Mary F. G. (1978). The effects of short-term contact and persuasive speech on special education students' attitudes toward handicappers. Unpublished doctoral dissertation, Michigan State University.

0209
Rosswurm, Mary Ann. (1980). Changing nursing students' attitudes toward persons with physical disabilities. Association of Rehabilitation Nurses Journal, 5(1), 12-14.

0139
Rothschild, Ilene Nancy. (1978). A comparison of cognitively and affectively oriented in-service training programs in changing teacher attitudes toward the handicapped. Unpublished doctoral dissertation, Teachers College, Columbia University.

0121
Rowe, Joanne & Stutts, Rose Marie. (In press). Impact of experience with disabled individuals on teachers' attitudes. Adapted Physical Activity Quarterly.

0006
Royster, Julia D. (1981). Changing employers' attitudes and behaviors toward disabled persons through participative workshops. Unpublished doctoral dissertation, North Carolina State University.

0248*
Rusalem, Herbert. (1967). Engineering changes in public attitudes toward a severely disabled group. Journal of Rehabilitation, 33(3), 26-27.

0212
Rynders, John E.; Johnson, Roger T.; Johnson, David W.; & Schmidt, Bernie. (1980). Producing positive interaction among down syndrome and nonhandicapped teenagers through cooperative goal structuring. American Journal of Mental Deficiency, 85, 268-273.

0234*
Sadlick, Marie, & Penta, Frank B. (1975). Changing nurse attitudes toward quadriplegics through use of television. Rehabilitation Literature, 36, 274-278, 288.

0043
Sager, Wilfred Gus. (1973). A study of changes in attitudes, values, and self concepts of senior high youth while working as full-time volunteers with institutionalized mentally retarded people. Unpublished doctoral dissertation, University of South Dakota.

0250
Salend, Spencer J., & Knops, Betsy. (1984). Hypothetical examples: A cognitive approach to changing attitudes toward the handicapped. The Elementary School Journal, 85(2), 228-235.

0182
Salend, Spencer J., & Moe, Laura. (1983). Modifying nonhandicapped students' attitudes toward their handicapped peers through children's literature. Journal for Special Educators, 19(3), 22-28.

0008
Sandberg, Leslee Diane Hoenscheid. (1980). Attitudes of nonhandicapped elementary school students towards school-aged trainable mentally retarded

students: An empirical investigation. Unpublished doctoral dissertation, The University of Iowa.

0207
Sanders, Johnny, Jr. (1978). The differential effects of selected methods of imparting information on teacher trainees' attitudes toward disabled persons. Unpublished doctoral dissertation, University of Georgia.

0218
Sartin, James L. (1964). A study of the modification of students' attitudes toward mentally retarded children. Unpublished doctoral dissertation, Syracuse University.

0193
Sasso, Gary M. (1983). The effects of multiple procedures on the behavior and attitudes of nonhandicapped students toward their severely emotionally disturbed peers. Unpublished doctoral dissertation, University of Kansas.

0249*
Sasso, Gary M.; Simpson, Richard L.; & Novak, Christine G. (1985). Procedures for facilitating integration of autistic children in public school settings. Analysis and Intervention in Developmental Disabilities, 5, 233-246.

0148
Saunders, Franklin F. (1969). Attitudes toward handicapped persons: A study of the differential effects of five variables. Unpublished doctoral dissertation, Florida State University.

0153
Sawyer, Horace W., & Clark, W. Dave. (1980). Disability simulation as a strategy for attitude change. Journal of Applied Rehabilitation Counseling, 11, 132-135.

0123
Scheibe, Karl E. (1965). College students spend eight weeks in mental hospital: A case report. Psychotherapy: Theory, Research and Practice, 2, 117-120.

0145
Schildhaus, Andrew Frank. (1971). Student nurse attitudes toward mental illness following a psychiatric training program. Unpublished doctoral dissertation, Columbia University.

0251*
Schroeder, Carolyn S. (1978, August). The psychologist's role in P. L. 94-142: Consultation strategies with peer groups of handicapped children. Paper presented as part of the Symposium "Pediatric Psychology: School as a Mental Health Provider" Eighty-sixth Annual American Psychological Association Convention, Toronto, Ontario, Canada. (ERIC No. ED 174 902)

0191
Schwartzwald, Frances. (1981). A study of the effect of cognitive and affective intervention on attitudes of upper elementary children toward physically disabled children. Unpublished doctoral dissertation, Texas Woman's University.

0228*
Shaver, Phillip R., & Scheibe, Karl E. (1967). Transformation of social identity: A study of chronic mental patients and college volunteer' in a summer camp setting. The Journal of Psychology, 66, 19-37.

0194
Shein, Eileen Wolpe. (1978). The effect of lectures and sign language instruction on changing attitudes toward the hearing impaired in an elementary school population. Unpublished doctoral dissertation, Hofstra University.

0240
Shortridge, Susan Denegan. (1982). Facilitating attitude change toward the handicapped. The American Journal of Occupational Therapy, 36(7), 456-460.

0252*
Shotel, Jay R.; Iano, Richard P.; & McGettigan, James F. (1972). Teacher attitudes associated with the integration of handicapped children. Exceptional Children, 677-683.

0169
Sigler, Gary; Mabee, Walter; & Lazar, Al. (1978). Children's attitudes toward handicapped students as a result of mainstreamed settings. Mental Retardation Bulletin, 6-7, 92-97.

0046
Simon, Gottlieb Chayim. (1970). Construction of a univocal measure of positive attitude toward people and its relation to change in attitude toward amputees. Unpublished doctoral dissertation, New York University.

0002(*)
Simpson, Richard L.; Parrish, Nancy E.; & Cook, Janis J. (1976). Modification of attitudes of regular class children towards the handicapped for the purpose of achieving integration. Contemporary Educational Psychology, 1, 46-51.

0021
Siperstein, Gary N., & Bak, John J. (1980). Improving children's attitudes toward blind peers. Journal of Visual Impairment and Blindness, 74, 132-135.

0012A (Lazar, Sippel, & D'Alonzo, 1977)
Sippel, Harry; Lazar, Alfred; & D'Alonzo, Bruno J. (1976). Changing attitudes toward TMR through an integrated recreation program. (Report to U. S. Office of Education, Bureau of Education for Handicapped, Report No. 451AH60850). Long Beach, CA: California State University, Department of Educational Psychology. (ERIC No. ED 123 860)

0049
Skrtic, Thomas M. (1976). The influence of inservice on the attitudes and behaviors of regular elementary classroom teachers toward mainstreamed learning disabled students. Unpublished doctoral dissertation, University of Iowa.

0132
Skrtic, Thomas M.; Clark, Frances L.; & White, Warren J. (1982). Modification of attitudes of regular education preservice teachers towar visually im-ired students. Visual Impairment and Blir-ness, 76(2), 49-52.

0168
Smith, Douglas B., & Larson, Patricia A. (1980). Adolescent attitudes toward disabled persons in an integrated and nonintegrated school setting. Education, 100, 390-394.

0133
Smith, Jean J. (1969). Psychiatric hospital experience and attitudes toward "mental illness". Journal of Consulting and Clinical Psychology, 33(3), 302-306.

0] 3
Sm-ch, Risa Cardwell. (1980). The efficacy of in-service to prepare regular classroom teachers for mainstreaming. Unpublished doctoral dissertation, The State University of New Jersey.

0070
Spiegel, Donald R.; Keith-Spiegel, Patricia; & Wine, David B. (1968, April). Evaluating the use of student volunteers in psychiatric settings. Paper presented at the Thirteenth Annual Conference of VA Cooperative Studies in Psychiatry, Denver, CO.

0241*
Spiegel, Don; Keith-Spiegel, Patricia; Zirgulis, Justine; and Wine, David B. (1973). Effects of student visits on social behavior of regressed schizophrenic patients. Journal of Clinical Psychology, 27, 396-400.

0075
Sponsler, Ruth E. (1965). An investigation of the effect produced by an educational film on attitudes toward the blind held by open and closed-minded students. Unpublished master's thesis, Sout' ern Illinois University.

0027*
Spreen, Otfried. (1977). Attitudes towards mental -ardation and attitude change: An experimental study. Zeitschrift fur experimentelle und angewandte Psychologie, 24, 303-323.

0039
Steen, Marcia Ruth. (1980). Simulation activities as a means c odifying teacher attitudes toward handicapped individuals. Unpublish octoral dissertation, Texas Tech University.

0273
Stemplinger, Robert G. (1985). The effects of multimodal instruction on the attitudes and behavioral intenti ns of high school students toward learning disabled peers. Unpublished doctoral dissertation, Rutgers University.

0059
Stensrud, Carol Jean. (1983). Meeting the needs of college students with disabilities: The effects of the sensitivity and special populations

mediated inservice training. Unpublished doctoral dissertation, University of Oregon.

0256
Stockton, Michael D. (1975). Attitudinal and behavioral change asociated with psychiatric attendant training. Unpublished doctoral dissertation, Florida State University.

0081
Storey, Kim Susan. (1979). The effects of the television series 'Feeling Free' on children's attitudes toward handicapped people. Unpublished doctoral dissertation, Harvard University.

0170
Strauch, James D. (1970). Social contact as a variable in the expressed attitudes of normal adolescents toward EMR pupils. Exceptional Children, 36, 495-500.

0138
Strauch, James D.; Chester, Peter N.; & Rucker, Chauncy N. (1970). Teacher aide attitudes toward the mentally retarded. Training School Bulletin, 67(1), 15-19.

0254*
Terrell, Laura W. (1981). Children's attitudes toward visual impairment: The effects of a teaching unit. Education of the Visually Handicapped, 13(3), 68-76.

0143
Thams, Gretchen Markle. (1975). The effects of a professional development seminar on the attitudes, knowledge, and behavioral strategies of counselors toward handicapped students. Unpublished doctoral dissertation, Wayne State University.

0235
Thornburg, Carol Ann. (1983). The effects of role playing puppets upon the attitudes of teachers toward handicapped children. Unpublished doctoral dissertation, George Peabody College for Teachers of Vanderbilt University.

0219*
Threlkeld, Robert M., & DeJong, William. (1983). Hiring the disabled: An attempt to influence behavioral intentions through a media presentation. Rehabilitation Psychology, 28, 105-114.

0103
Towner, Arthurlene G. (1981). Role playing and the modification of preservice teacher attitudes toward hearing impaired individuals: An experimental analysis. Unpublished doctoral dissertation, University of California.

0155
Tweto-Johnson, Linda. (1984). Study of changes in paraprofessionals' knowledge and attitudes after inservice training. Unpublished doctoral dissertation, Northern Arizona University.

0202
Urie, Robert M., & Smith, Alvin H. (1970-71). The effects of peer contact on attitudes toward disabled college students. Journal of Applied Rehabilitation Counseling, 1(4), 24-32.

0028
Vinish, Mary T. (1974). Modification of children's attitudes toward the handicapped: An assessment of the effectiveness of a specific unit of study. Unpublished master's thesis, University of Saskatchewan.

0187
Voeltz, Luanna M. (1980). Children's attitudes toward handicapped peers. American Journal of Mental Deficiency, 84, 455-464.

0271
Voeltz, Luanna M. (1982). Effects of structured interactions with severely handicapped peers on children's attitudes. American Journal of Mental Deficiency, 86, 380-390.

0236
Walker, Sylvia. (1978). The disabled in Ghana: Status and change in information and attitude. Unpublished doctoral dissertation, Columbia University Teachers College.

0149*
Wallace, Mildred G. (1980). Attitudes toward mainstreaming: Impact of an inservice training program on participants. Unpublished doctoral dissertation, Vanderbilt University.

0150(*)
Wallston, Barbara S.; Blanton, Richard L.; Robinson, Janice S.; & Polichik, Linda L. (1972). Community Human resources development in the rehabilitation of the handicapped. (Report to the Social and Rehabilitation Service, Department of Health, Education, and Welfare, Grant No. RD-12-P-5516714-01). Nashville, TN: Outlook Nashville, Inc. (ERIC No. ED 078 239)

0099
Walsh, Joan E. (1970). A study of the relationship between instruction in psychiatric nursing, level of anxiety, and direction of attitudes toward the mentally ill. Unpublished doctoral dissertation, New York University.

0161
Weaver, Roberta Brunner. (1982). Effect of school experiences with severely handicapped children on the attitudes of faculty and children. Unpublished doctoral dissertation, University of Cincinnati.

0144A (Westervelt & McKinney, 1980)
Westervelt, Van D. (1978). The influence of a filmed intervention on the attraction of non-handicapped children toward physically handicapped children. Unpublished master's thesis, University of North Carolina at Chapel Hill.

0144B (Westervelt, 1978)
Westervelt, Van D., & McKinney, Jam D. (1980). Effects of a film on nonhandicapped children's attitudes toward handicapped children. Exceptional Children, 46, 294-296.

0141
White, Susan Menke. (1981). The effects of an informational sensitization program on the attitudes of regular seventh grade students toward special education students. Unpublished doctoral dissertation, University of Wisconsin-Milwaukee.

0011
White, Walter A. (1973). Attitude and attitude change toward mental retardation of certain worker and student groups as a function of related education and experience. Unpublished doctoral dissertation, Boston University.

0237
Williams, Martha S. (1977). Change in teacher attitudes toward exceptional children as a result of an in-service education workshop. Unpublished doctoral dissertation, George Peabody College for Teachers.

0258*
Wilson, Earl D. (1971). A comparison of the effects of deafness simulation and observation upon attitudes, anxiety, and behavior manifested toward the deaf. The Journal of Special Education, 5(4), 343-349.

0146
Wilson, Earl D., & Alcorn, Dewaine. (1969). Disability simulation and development of attitudes toward the exceptional. The Journal of Special Education, 3(3), 303-307.

0238
Wilson, Rose Confer. (1979). An experimental comparison of instructional programs and their effects on changing attitudes toward the disabled. Unpublished doctoral dissertation, McNeese State University.

0054
Winkler, Sharon H. (1981). A comparative study of teacher attitudes toward learning handicapped students. Unpublished doctoral dissertation, United States International University.

0225
Woods, Paul B., Jr. (1985). An examination of the perceptions and preferences of nonhandicapped kindergarten children toward physically handicapped peers: Effects of a brief film and teacher-led discussion. Unpublished doctoral dissertation, University of North Carolina--Chapel Hill.

0239
Wurzel, Susan Tabbat. (1980). The reduction of prejudice and stereotyping: A treatment to improve children's attitudes towards people with disabilities. Unpublished doctoral dissertation, Boston University School of Education.

0160*
Wyrick, Joane M. (1967). Lack of effect on attitude change of two films dealing with cerebral palsy. American Journal of Occupational Therapy, 21, 76-78.

0129
Wyrick, Joane M. (1968). The effect of the lecture method on attitudes toward physical disability. Unpublished master's thesis, University of Kansas.

0158
Yerxa, Elizabeth June. (1971). The effects of a dyadic, self-administered instructional program in changing the attitudes of female college students toward physically disabled persons. Unpublished doctoral dissertation, Boston University.

0242
Zahiri Tousi, Javad. (1980). The immediate effects of informative books and films upon the attitudes of fourth grade students toward handicapped children. Unpublished doctoral dissertation, University of Missouri - Columbia.

APPENDIX F

DESCRIPTIVE DATA FOR STUDIES CODED

(1) Treatment versus Control, Treatment versus Placebo, Single-group, Pre-posttest Effect Sizes

(2) Treatment A versus B Effect Sizes

(3) Mainstreaming Effect Sizes

(4) Missing Information Results

TREATMENT VERSUS CONTROL, TREATMENT VERSUS PLACEBO, SINGLE-GROUP, PRE-POSTTEST EFFECT SIZES

KEY

Column	Codes
Author/Year	Reports are identified by the last name or names of authors and by year. Full citations can be found in Appendix E.
ID#	Arbitrary number for recordkeeping purposes.
ESID#	Numbers arbitrarily assigned to effect sizes during coding.
ATTOWAD	The disability or disabilities attitude target toward which the attitude change treatment was directed:

1=disabled in general
2=physically disabled in general
3=retarded, general
4=moderately retarded
5=severely retarded
6=mentally ill
7=emotionally disturbed
8=visually impaired
9=hearing impaired/deaf
10=learning disabled
11=speech/language impaired
12=cerebral palsied
13=epileptic
14=para/quadriplegic
15=other physically impaired
16=autistic
17=health impaired
18=multiply handicapped
19=other
20=combination

Column	Codes
TOTN	The total N for the comparison (treatment plus comparison group, or the N of the treatment group for single-group, pre-posttest comparisons).
SETTINGE	The specific type of place where the experimental treatment was conducted:

0=can't tell
1=regular classroom
2=special classroom
3=home
4=institution
5=group home
6=hospital
7=dormitory
8=playground
9=camp
10=recreation facility
11=laboratory
12=individual/small group
13=normal life
14=other
15=combination

Column	Codes
EDLEVEXP	The education level of the experimental Ss:

0=can't tell
1=preschool
2=primary
3=intermediate
4=middle school
5=junior high school
6=senior high school
7=combination
8=undergraduates
9=graduate students
10=post-professional
11=adults not in school
12=other

Column	Codes

UNIVMAJE

The major(s) of the experimental Ss, if college or university students:

0=not applicable
1=can't tell
2=elementary education
3=secondary education
4=elementary & secondary education
5=education, unspecified
6=nursing
7=occupational & physical therapy
8=philosophy
9=psychology
10=rehabilitation counseling
11=social work
12=sociology
13=special education
14=communicative disorders
15=medicine
16=other

OCCUPEXP

The occupation of experimental Ss who were not students:

0=not applicable
1=can't tell
2=child care workers
3=community recreation workers
4=employees in institutions
5=regular class teachers, elementary
6=middle or junior high
7=high school
8=school administrators
9=special class teachers
10=vocational rehabilitation counselors
11=parents
12=police
13=medical
14=general public
15=other

COMINSTR

Whether an instrument identified as commonly used in research on attitudes toward persons with disabilities was the dependent measure:

0=no
1=Attitudes Towards Disabled Persons Scale (ATDP)
2=Opinions about Mental Illness Scale (OMI)
3=Rucker-Gable Educational Programming Scale (RGEPS)
4=Attitudes Towards Handicapped Individuals Scale (ATHI)
5=Attitude to Blindness Scale (ATBS)
6=Mental Retardation Attitude Inventory (MRAI)
7=Minnesota Teacher Attitude Inventory (MTAI), revised
8=combination

ASSESTYP

The type of instrumentation used for the dependent measure:

1=interview, structured
2=interview, nonstructured
3=attitude questionnaire
4=sociometric measure
5=peer assessment
6=social distance scale
7=informal observation
8=systematic observation
9=semantic differential
10=telephone or mail survey
11=telephone or mail request
12=Q-sort
13=projective test, pictures
14=sentence completion
15=adjective checklist
16=rankings
17=other

Column	Codes
TRTECHE	The attitude modification technique used as the treatment: 1=information 2=direct contact 3=vicarious experience 4=positive reinforcement 5=persuasive message 6=persuasive messages, contrast 7=1&2 8=1&3 9=other 10=systematic desensitization
TOTHRSE	The total hours of the experimental treatment. If no number is entered, the information was not available.
ES	The effect size, defined as $D = \frac{\bar{X}_T - \bar{X}_C}{SD_C}$, or an estimate thereof.

558

TREATMENT VERSUS CONTROL, TREATMENT VERSUS PLACEBO, SINGLE-GROUP, PRE-POSTTEST EFFECT SIZES

AUTHOR	YEAR	ID#	ESID#	ATTOWARD	TOTN	SETTINGE	EDLEVEXP	UNIVMAJE	OCCUPEXP	COMINSTR	ASSESTYP	TRTECHE	TOTHRSE	ES
Quay, et al.	1961	114	1	3	34	4	0	0	4	0	3	1	1 0	43
		114	2	3	34	4	0	0	4	0	3	1		22
		114	3	3	34	4	0	0	4	0	3	1	1 0	- 06
Butler	1981	63	1	1	48	1	2	0	0	0	3	1	4 0	17
		63	2	1	51	1	2	0	0	0	3	1	4 0	- 07
Bucich-Naylor	1978	195	2	1	52	1	2	0	0	0	3	1	3 7	39
		195	8	1	52	1	2	0	0	0	6	1	3 7	03
Woods	1985	225	1	1	89	1	2	0	0	0	6	1	6	07
Lory	1983	227	1	1	53	1	2	0	0	0	16	1	4 0	23
		227	3	1	53	1	2	0	0	0	17	1	4 0	77
		227	5	1	53	1	2	0	0	0	17	1	4 0	59
		227	7	1	53	1	2	0	0	0	17	1	4 0	29
		227	8	1	53	1	2	0	0	0	17	1	4 0	00
Glauberman	1980	44	1	2	40	12	2	0	0	0	3	1	2	41
Rae	1983	108	2	5	30	1	2	0	0	0	3	1	7	- 18
		108	6	5	30	14	2	0	0	0	3	1	7	60
Sasso	1983	193	2	7		1	2	0	0	0	3	1	8 0	1.21
Lipsky	1978	10	4	1	60	1	3	0	0	0	3	1	5	23
Vinish	1974	28	2	1	58	1	3	0	0	0	3	1	20 0	32
Dresang	1981	52	2	1	51	1	3	0	0	0	4	1	7	- 08
		52	3	1	48	1	3	0	0	0	4	1	7	- 02
		52	5	1	73	1	3	0	0	0	9	1	7	15
		52	6	1	76	1	3	0	0	0	9	1	7	00
Paxton	1983	90	1	1	93	1	3	0	0	0	15	1	8	41
		90	3	1	103	1	3	0	0	0	15	1	8	41
		90	4	1	93	1	3	0	0	0	15	1	8	25
		90	6	1	103	1	3	0	0	0	15	1	8	- 15
Salend, & Moe	1983	182	1	1	160	1	3	0	0	0	15	1		23
Zahiri Tousi	1980	242	1	1	36	1	3	0	0	0	6	1	1 0	00
		242	2	1	37	1	3	0	0	0	6	1	1 0	- 02
		242	3	1	42	1	3	0	0	0	6	1	2 0	- 01
Salend, & Knops	1984	250	1	1	32	1	3	0	0	0	15	1	1 0	49
		250	2	1	37	1	3	0	0	0	15	1	1 0	70
		250	3	1	38	1	[illegible]	0	0	0	15	1	1 0	87
		250	4	1	32	1	3	0	0	0	15	1	1 0	.55
		250	5	1	37	1	3	0	0	0	15	1	1 0	77
		250	6	1	38	1	3	0	0	0	15	1	1 0	86
Marquart	1984	35	1	1	47	2	3	0	0	0	3	1	4 2	14
		35	5	1	47	2	3	0	0	0	15	1	4 2	- 11
		35	9	1	47	2	3	0	0	0	6	1	4 2	- 29
Perez	1983	85	1	1	619	2	3	0	0	0	3	1	15 0	- 09
Lipsky	1978	10	2	1	60	1	3	0	0	1	3	1	5	- 96
Mulkey	1980	88	4	1	32	1	3	0	0	1	3	1	3	- 82
Shein	1978	194	2	1	40	1	3	0	0	1	3	1	3 0	32
		194	5	1	40	1	3	0	0	1	3	1	3 0	01
Parrish	1974	270	3	1	28	1	3	0	0	1	3	1	4 0	85
		270	4	1	28	1	3	0	0	1	3	1	4 0	25
Schwartzwald	1981	191	1	2	156	1	3	0	0	1	3	1	7	26
Nolan	1982	230	1	2	40	1	3	0	0	1	3	1	3 0	69
Siperstein, & Bak	1980	21	1	8	109	1	3	0	0	0	15	1	1 0	59
		21	2	8	109	1	3	0	0	0	5	1	1 0	- 53

AUTHOR	YEAR	ID#	ESID#	ATTOWARD	TOTN	SETTINGE	EDLEVEXP	UNIVMAJE	OCCUPEXP	COMINSTR	ASSESTYP	TRTECHE	TOTHRSE	ES
Shein	1978	194	8	9	40	1	3	0	0	0	3	1	3 0	76
		194	11	9	40	1	3	0	0	0	3	1	3 0	31
Elliott, & Byrd	1983	201	1	1	62	1	4	0	0	1	3	1	3	28
		201	2	1	66	1	4	0	0	1	3	1	3	1 27
Bond	1979	112	1	2	176	3	4	0	0	1	3	1		30
		112	2	2	176	3	4	0	0	1	3	1		25
Rae	1983	108	4	5	30	1	4	0	0	0	3	1	7	- 37
		108	8	5	30	14	4	0	0	0	3	1	7	- 03
Owen	1984	176	2	20	24	1	4	0	0	0	3	1	2 7	- 09
		176	4	20	24	1	4	0	0	0	3	1	2. 7	- 01
		176	6	20	24	1	4	0	0	0	3	1	2 7	23
		176	8	20	24	1	4	0	0	0	3	1	2 7	17
Mulkey	1980	88	6	1	55	1	5	0	0	1	3	1	3	23
McKerracher	1982	120	6	2	93	1	5	0	0	1	3	1	2 7	29
White	1981	141	1	20	114	1	5	0	0	1	3	1		22
		141	2	20	120	1	5	0	0	1	3	1		57
		141	3	20	118	1	5	0	0	1	3	1		14
		141	4	20	123	1	5	0	0	1	3	1		57
Reinhardtsen	1980	259	2	1	335	1	6	0	0	0	3	1	4 5	35
Robinson	1980	72	1	1	16	0	6	0	0	1	3	1	24 0	- 97
		72	2	1	32	0	6	0	0	1	3	1	24 0	- 57
Mulkey	1980	88	8	1	46	1	6	0	0	1	3	1	3	1 32
Reinhardtsen	1980	259	1	1	335	1	6	0	0	4	3	1	4 6	23
Cerreto	1976	190	3	1	84	1	7	0	0	0	17	1	7	32
		190	4	1	84	1	7	0	0	0	17	1	7	- 23
Farrell	1985	18	4	1	40	14	7	0	0	0	4	1		76
Baker, et al.		203	2	1	1094	1	7	0	0	1	3	1	6	57
		203	4	1	1313	1	7	0	0	1	3	1	6	39
Drake	1977	185	1	1	31	1	7	1	0	4	3	1	37 5	41
		185	2	1	23	1	7	1	0	4	3	1	37 5	78
		185	3	1	32	1	7	1	0	4	3	1	37 5	13
Dubin	1981	210	1	7	472	1	7	0	0	1	3	1	8 0	70
Krieger	1978	116	3	8	40	1	7	16	0	5	3	1	2	73
Storey	1979	81	1	20	203	1	7	0	0	0	3	1	1 5	30
		81	2	20	203	1	7	0	0	0	3	1	1 5	42
Martinez	1977	223	6	1	30	0	8	5	0	0	9	1	2	- 83
Walker	1978	236	1	1	59	1	8	5	0	0	3	1	28 0	16
		236	2	1	59	1	8	5	0	0	3	1	28 0	16
		236	3	1	59	1	8	5	0	0	3	1	28 0	39
		236	4	1	59	1	8	5	0	0	3	1	28 0	26
Ibrahim	1979	204	6	1	30	1	8	14	0	0	3	1	5 0	55
		204	8	1	30	1	8	14	0	0	3	1	5 0	1. 20
		204	10	1	30	1	8	14	0	0	11	1	5 0	33
Elliott, & Byrd	1984	66	1	1	30	0	8	1	0	1	3	1	3	1 10
		66	2	1	30	0	8	1	0	1	3	1	3	67
Martinez	1977	223	3	1	30	0	8	5	0	1	3	1	2	- 49
Hamilton, & Anderson	1983	16	2	1	123	1	8	1	0	1	3	1	2 5	40
Abernathy	1980	50	1	1	104	1	8	1	0	1	3	1	7	37
Hafer, & Marcus	1979	51	1	1	36	1	8	1	0	1	3	1	6	- 72
		51	2	1	36	1	8	1	0	1	3	1	6	- 09
Wilson	1979	238	1	1	210	1	8	1	0	1	3	1	7	- 48

AUTHOR	YEAR	ID#	ESID#	ATTOWARD	TOTN	SETTINGE	EDLEVEXP	UNIVMAJE	OCCUPEXP	COMINSTR	ASSESTYP	TRTECHE	TOTHRSE	ES
Wilson	1979	238	2	1	146	1	8	1	0	1	3	1		- 18
Saunders	1969	148	1	1	17	1	8	2	0	1	3	1	22 5	28
		148	2	1	15	1	8	2	0	1	3	1	22 5	- 56
		148	3	1	70	1	8	2	0	1	3	1	22 5	-1 61
Peterson	1977	136	2	1	35	1	8	13	0	1	3	1	2 0	- 2
Ibrahim	19[illegible]	204	2	1	30	1	8	14	0	1	3	1	5.0	51
		204	4	1	30	1	8	14	0	1	3	1	5 0	1.35
McDaniel	1980	215	3	1	106	1	8	16	0	1	3	1		54
		215	4	1	139	1	8	16	0	1	3	1		29
Wilson	1979	238	3	1	144	13	8	1	0	1	3	1	3 0	02
Lazar, et al.	1975	255	1	1	34	1	8	1	0	4	3	1		.05
		255	2	1	34	1	8	1	0	4	3	1		.21
		255	3	1	34	1	8	1	0	4	3	1		- 26
Cohen	1978	220	6	2	36	0	8	0	0	0	6	1	1.3	36
		220	9	2	36	0	8	0	0	0	9	1	1 3	30
		220	12	2	36	0	8	0	0	0	9	1	1 3	39
		220	15	2	36	0	8	0	0	0	9	1	1 3	13
Yerxa	1971	158	1	2	76	12	8	16	0	0	3	1	8 0	47
		158	2	2	76	12	8	16	0	0	3	1	8 0	- 07
		158	3	2	76	12	8	16	0	0	3	1	8 0	43
		158	4	2	76	12	8	16	0	0	3	1	8.0	35
		158	5	2	76	12	8	16	0	0	3	1	8 0	15
		158	6	2	76	12	8	16	0	0	3	1	8 0	31
Cohen	1978	220	3	2	36	0	8	0	0	1	3	1	1 3	2 13
Ozyurek	1977	233	2	2	44	0	8	2	0	1	3	1		63
		233	3	2	47	0	8	2	0	1	3	1		52
Aukayanagul	1979	137	9	2	20	0	8	5	0	1	3	1	1 3	80
		137	10	2	20	0	8	5	0	1	3	1	1 3	39
		137	11	2	20	0	8	5	0	1	3	1	1 3	1.04
		137	12	2	20	0	8	5	0	1	3	1	1 3	1 01
		137	2	2	20	0	8	15	0	1	3	1	1.3	12
		137	3	2	20	0	8	15	0	1	3	1	1 3	22
		137	4	2	20	0	8	15	0	1	3	1	1 3	53
		137	6	2	20	0	8	16	0	1	3	1	1 3	63
		137	7	2	20	0	8	16	0	1	3	1	1 3	-.06
		137	8	2	20	0	8	16	0	1	3	1	1 3	33
		137	14	2	20	0	8	16	0	1	3	1	1 3	1 26
		137	15	2	20	0	8	16	0	1	3	1	1 3	85
		137	16	2	20	0	8	16	0	1	3	1	1 3	1 16
Wyrick	1968	129	2	2	51	1	8	1	0	1	3	1	33 7	30
		129	1	2	61	1	8	7	0	1	3	1	33 7	14
Yerxa	1971	158	7	2	76	12	8	16	0	1	3	1	8 0	46
Sartin	1964	218	1	3	100	1	8	16	0	0	9	1	3 0	41
		218	2	3	100	1	8	16	0	0	9	1	3 0	- 24
Boehm	1978	32	2	5	41	15	8	13	0	0	9	1	75 0	35
		32	4	5	32	15	8	13	0	0	9	1	75 0	64
Frazier	1975	62	3	6	27	1	8	1	0	0	9	1		37
		62	6	6	27	1	8	1	0	0	9	1		15
Costin, & Kerr	1962	25	1	6	197	1	8	1	0	2	3	1	33 7	29
		25	2	6	197	1	8	1	0	2	3	1	33 7	1 00
Frazier	1975	62	9	6	27	1	8	1	0	2	3	1		- 22

AUTHOR	YEAR	ID#	ESID#	ATTWARD	TOTN	SETTINGE	EDLEVC(P	UNIVMAJE	OCCUPE(P	COMINSTR	ASSESTYP	TRTECHE	TOTHRSE	CS
Frazier	1975	62	12	6	27	1	8	1	0	2	3	1		- 10
Bateman	1964	12	1	8	62	1	8	1	0	0	17	1	1 0	1 31
Skrtic, et al.	1982	132	1	8	58	1	8	5	0	5	3	1		- 19
		132	3	8	46	1	8	5	0	5	3	1		- 15
Sponsler	1965	75	1	8	160	7	8	1	0	5	3	1	4	20
		75	2	8	160	7	8	1	0	5	3	1	4	08
Dampier	1982	246	3	9	28	1	8	1	0	0	9	1	4	1 73
		246	4	9	28	1	8	1	0	0	9	1	4	90
Guldager	1973	131	1	19	390	1	8	1	0	0	3	1	1	30
		131	2	19	390	1	8	1	0	0	3	1	1	55
Naor, & Milgram	1980	177	2	20	54	1	8	2	0	0	3	1	22 5	[illegible]
		177	4	20	54	1	8	2	0	0	17	1	22 5	[illegible]
Elliott, & Byrd	1984	37	2	1	29	1	9	10	0	1	3	1	7	93
		37	4	1	24	1	9	10	0	1	3	1	7	20
Morton	1977	266	2	2	30	1	9	16	0	1	3	1	1 0	08
		266	1	2	30	13	9	16	0	1	3	1	1 0	- [illegible]
Mathey	1977	93	2	1	60	0	10	0	0	0	9	1	12 0	47
		93	3	1	60	0	10	0	0	0	9	1	12 0	59
		93	4	1	60	0	10	0	0	0	9	1	12 0	36
Stensrud	1983	59	1	1	185	14	10	0	15	0	3	1	5.0	79
Haring	1956	179	1	1	117	14	10	0	15	0	13	1	30 0	22
Steen	1980	89	2	1	28	0	10	0	15	1	3	1	4 5	- 08
McDaniel	1980	215	1	1	142	0	10	0	15	1	3	1	6 0	86
Williams	1977	237	1	1	103	0	10	0	5	1	3	1	40 0	- 10
Rothschild	1978	139	2	1	34	1	10	0	5	1	3	1	20.0	- 02
Thams	1975	143	1	1	26	1	10	0	15	1	3	1	18 0	73
McDaniel	1980	215	2	1	138	1	10	0	15	1	3	1		40
Perez	1983	85	2	1	85	2	10	0	5	1	3	1	15 0	20
Mathey	1977	93	1	1	60	0	10	0	0	3	6	1	12 0	55
Jacobs	1977	92	1	1	440	1	10	0	15	3	6	1	36 0	1 26
Finn	1979	107	1	1	81	0	10	0	0	4	3	1	10 7	-.31
Cobun	1972	156	1	2	62	1	10	0	10	0	3	1	138 7	- 10
		156	2	2	62	1	10	0	10	0	3	1	138 7	- 13
		156	3	2	62	1	10		10	0	3	1	138 7	- 52
		156	4	2	62	1	10		10	0	3	1	138 7	- 43
		156	5	2	62	1	10	0	10	0	3	1	138 7	- 37
		156	6	2	62	1	10	0	10	0	3	1	138 7	- 73
Kauffman	1976	157	1	4	27	14	10	0	5	0	3	1	8	21
		157	4	4	27	14	10	0	5	0	3	1	8	11
		157	7	4	27	14	10	0	5	0	3	1	8	10
		157	10	4	27	14	10	0	5	0	3	1	8	46
Skrtic	1976	49	1	10	31	1	10	0	5	4	3	1	9 0	- 03
		49	2	10	31	1	10	0	5	4	3	1	9 0	40
Tweto-Johnson	1984	155	1	1	30	1	11	0	15	4	3	1	16 0	28
		155	2	1	30	1	11	0	15	4	3	1	16 0	07
Alese	1973	1	1	3	363	13	11	0	1	0	3	1	7	39
Bitter	1963	17	1	5	16	1	11	0	11	0	9	1	14 0	- 19
Levinson, & Distefano	1979	78	1	6	24	15	11	0	12	2	3	1	14 0	14
		78	2	6	24	15	11	0	12	2	3	1	14 0	- 25
Hogan	1983	214	1	20	20	1	11	0	11	0	3	1	20 0	- 17
Sippel, et al.	1976	12	1	5	26	10	0	0	1	4	3	2		33
Emerton & Rothman	1975	151	1	9	30	7	8	1	0	0	17	2	.	-.52

AUTHOR	YEAR	ID#	ESID#	ATTOWARD	TOTN	SETTINGE	EDLEVEXP	UNIVMAJE	OCCUPEXP	COMINSTR	ASSESTYP	TRTECHE	TOTHRSE	ES
Hus	1979	243	1	9	19	9	0	0	3	0	3	2	120 0	75
Bucich-Naylor	1978	195	1	1	43	1	2	0	0	0	3	2	3 7	17
		195	7	1	43	1	2		0	0	0	2	3 7	- 01
Hawisher	1977	196	2	15	73	1	2	0	0	0	3	2	7	83
Putnam	1983	65	1	1	40	1	3	0	0	0	3	2		25
		65	2	1	40	1	3	0	0	0	3	2	7 5	- 10
Friedman	1975	115	1	1	55	15	3	0	0	1	3	2	7 5	52
Schwartzwald	1981	191	2	2	163	14	3	0	0	1	3	2	45 0	26
Lowther	1978	55	1	4	250	1	3	0	0	0	4	2	1.5	3 67
		55	2	4	300	1	3	0	0	0	4	2	1.5	04
		55	3	4	275	1	3	0	0	0	4	2	1.5	4 40
Olsen	1978	127	2	4	54	2	3	0	0	0	3	2	25 0	17
		127	4	4	47	2	3	0	0	0	17	2	25 0	1.05
Sandberg	1980	8	1	5	400	14	3	0	0	0	6	2		13
Weaver	1982	161	2	19	147	1	3	0	0	0	3	2	900 0	32
		161	3	19	139	1	3	0	0	0	3	2	900 0	30
		161	4	19	147	1	3	0	0	0	3	2	900 0	48
		161	5	19	159	1	3	0	0	0	3	2	900 0	- 04
Fenrick, & Petersen	1984	213	1	20	63	2	3	0	0	0	9	2	10 5	75
		213	2	20	63	2	3	0	0	0	6	2	10 5	34
McKerracher	1982	120	7	2	90	1	5	0	0	1	3	2	1 0	02
Sager	1973	43	1	5	22	4	5	0	0	0	9	2	360 0	1 03
Farrell	1985	18	3	1	40	14	7	0	0	0	4	2		42
Wallston, et al.	1972	150	1	1	43	14	7	0	0	0	6	2		- 47
Anthony	1969	5	1	1	26	9	7	0	3	1	3	2		53
		5	2	1	16	9	7	0	3	1	3	2		49
Allen, & Fosnee	1966	269	1	3	871	15	7	0	0	0	9	2		- 01
		269	2	3	871	15	7	0	0	0	9	2		05
Allen, & Allen	1968	272	1	3	367	15	7	1	0	0	3	2		.55
Freeman	1970	15	1	7	13	6	7	13	0	0	9	2		13
		15	2	7	12	6	7	13	0	0	9	2		-.10
		15	3	7	13	6	7	13	0	0	9	2		43
		15	4	7	12	6	7	13	0	0	9	2		49
Colasuonno	1981	183	2	9	32	0	7	16	0	0	3	2	1 2	- 12
Fender	1979	226	1	1	116	1	8	16	0	0	9	2	4 0	33
		226	2	1	116	1	8	16	0	0	9	2	4 0	52
		226	3	1	116	1	8	16	0	0	9	2	4 0	74
		226	4	1	116	1	8	16	0	0	9	2	4 0	34
		226	5	1	116	1	8	16	0	0	9	2	4 0	10
		226	6	1	116	1	8	16	0	0	9	2	4 0	- 08
		226	7	1	116	1	8	16	0	0	9	2	4 0	23
		226	8	1	116	1	8	16	0	0	6	2	4 0	14
		226	9	1	116	1	8	16	0	0	6	2	4 0	12
		226	10	1	116	1	8	16	0	0	6	2	4 0	44
		226	11	1	116	1	8	16	0	0	6	2	4 0	23
		226	12	1	116	1	8	16	0	0	6	2	4 0	- 16
		226	13	1	116	1	8	16	0	0	6	2	4 0	- 17
		226	14	1	116	1	8	16	0	0	6	2	4 0	27
Urie, & Smith	1970-71	202	2	1	254	13	8	1	0	0	3	2		- 08
Kisabeth, & Richardson	1985	9	1	1	41	1	8	1	0	1	3	2	32 5	04
Petrangelo	1976	134	1	1	57	1	8	16	0	1	3	2		58

AUTHOR	YEAR	ID#	ESID#	ATTOWARD	TOTN	SETTINGE	EDLEVEXP	UNIVMAJE	OCCUPEXP	COMINSTR	ASSESTYP	TRTECHE	TOTHRSE	ES
Petrangelo	1976	134	2	1	:55	1	8	16	0	1	[illegible]	2		26
Rowe, & Stutts	in press	121	3	1	40	2	8	16	0	1	3	2	24 0	1 81
		121	•	1	40	11	8	16	0	1	3	2	24 0	2 29
		121	2	1	35	11	8	16	0	1	3	2	24 0	1 88
Urie, & Smith	1970-71	202	1	1	254	13	8	1	0	1	3	2		15
Hamilton, & Anderson	1983	16	1	1	120	14	8	1	0	1	3	2	2 5	30
Rowe, & Stutts	in press	121	4	1	60	14	8	16	0	1	3	2	24 0	1 00
Evans	1976	36	2	2	40	12	8	1	0	1	3	2	7	13
White	1973	11	1	3	64	2	8	13	0	0	3	2		00
		11	2	3	56	2	8	13	0	0	3	2		- 09
		11	3	3	55	2	8	13	0	0	3	2		00
Landis	1981	159	1	3	94	2	8	13	0	0	9	2	120 0	55
Strauch, et al.	1970	138	1	3	10	4	8	1	0	0	9	2	180 0	26
		138	2	3	10	4	8	1	0	0	9	2	180 0	31
White	1973	11	4	3	66	4	8	13	0	0	3	2		71
Holzberg, & Gewirtz	1963	58	1	6	59	6	8	1	0	0	3	2	60 0	1 24
Kulik, et al.	1969	74	2	6	318	6	8	1	0	0	15	2	360 0	23
		74	3	6	318	6	8	1	0	0	15	2	360 0	13
Scheibe	1965	123	1	6	99	15	8	1	0	0	15	2	320 0	32
		123	2	6	99	15	8	1	0	0	15	2	320 0	35
Spiegel, et al.	1968	70	1	6	40	6	8	1	0	2	3	2		57
		70	2	6	40	6	8	1	0	2	3	2		64
Kulik, et al.	1969	74	1	6	318	6	8	1	0	2	3	2	360 0	30
Smith	1969	133	1	6	136	6	8	0	0	2	3	2		12
		133	2	6	136	6	8	6	0	2	3	2		11
Evans	1976	36	1	2	40	12	8	1	0	1	3	2	.7	.63
Kauffman	1976	157	3	4	24	2	10	0	5	0	3	2	7	- 52
		157	6	4	24	2	10	0	5	0	3	2	7	- 33
		157	9	4	24	2	10	0	5	0	3	2	7	- 44
		157	12	4	24	2	10	0	5	0	3	2	7	15
Colasuonno	1981	183	1	9	32	0	10	0	9	0	3	2	1 2	17
Weaver	1982	161	1	19	48	1	10	0	5	0	3	2	900 0	74
Klein	1969	79	1	3	48	14	11	0	11	0	3	2		-.15
Smith	1969	133	3	6	134	6	11	0	4	2	3	2	999 9	23
		133	4	6	134	6	11	0	4	2	3	2	999 9	28
Granofsky	1956	104	1	19	135	6	11	0	1	0	13	2	8 0	- 16
		104	2	19	135	6	11	0	1	0	14	2	8 0	40
Hicks, & Spaner	1962	57	1	6	78	6	12	0	0	0	3	2	420 0	2 04
		57	2	6	354	6	12	0	0	0	3	2	420 0	91
		57	3	6	354	6	12	0	0	0	3	2	420 0	90
Beardsley	1979	40	1	1	110	1	2	0	0	0	3	3	2 9	00
		40	2	1	130	1	2	0	0	0	3	3	2 9	10
Lipsky	1978	10	3	1	60	1	3	0	0	0	3	3	5	78
Dahl, et al.	1978	26	6	1	8[illegible]	1	3	0	0	0	14	3	7	- 24
Paxton	1983	90	2	1	97	1	3	0	0	0	15	3	8	60
		90	5	1	97	1	3	0	0	0	15	3	8	- 14
Marcus	1979	264	1	1	396	1	3	0	0	0	4	3	25 0	- 09
Marquart	1984	35	3	1	47	2	3	0	0	0	3	3	4 2	54
		35	7	1	47	2	3	0	0	0	15	3	4 2	- 10
		35	11	1	47	2	3	0	0	0	6	3	4 2	- 24

AUTHOR	YEAR	ID#	ESID#	ATTOWARD	TOTN	SETTINGE	EDLE'_P	UNIVMAJE	OCCUPEXP	COMINSTR	ASSESTYP	TRTECHE	TOTHRSE	ES
Lipsky	1978	10	1	1	60	1	3	0	0	1	3	3	5	1 26
Dahl, et al.	1978	26	5	1	89	1	3	0	0	1	3	3	7	09
Margo	'983	206	1	2	44	0	3	0	0	0	15	3	1 0	05
Dahl, et al.	1978	26	3	2	89	1	3	0	0	0	o	3	7	42
Westerveit	1978	144	1	2	46	2	3	0	0	0	6	3	2	47
		144	2	2	4ل	2	3	0	0	0	6	3	2	06
Dahl, et al.	1978	26	4	3	89	1	3	0	0	0	6	3	7	30
		26	1	8	89	1	3	0	0	0	o	3	7	28
		26	2	9	89	1	3	0	0	0	6	3	7	23
Cerreto	1976	190	1	1	84	1	7	0	0	0	17	3	7	93
		190	2	1	84	1	7	0	0	0	17	3	7	- 35
Martinez	1977	223	4	1	30	0	8	5	0	0	9	3	2	-1. 42
		223	5	1	30	0	8	5	0	0	9	3	2	-1. 28
Ibrahim	1979	204	5	1	30	13	8	14	0	0	3	3	5 0	1. 39
		204	7	'	30	13	8	14	0	0	3	3	5 0	1. 97
		204	9	1	30	13	8	14	0	0	11	3	5 0	67
Martinez	1977	223	1	1	30	0	8	5	0	1	3	3	2	- 42
		223	2	1	30	0	8	5	0	1	3	3	2	- 63
Peterson	1977	136	3	1	43	1	8	13	0	1	3	3	2 0	- 13
Wilson, & Alcor	1969	146	1	1	80	13	8	5	0	1	3	3	8 0	01
Ibrahim	1979	204	1	1	30	13	8	14	0	1	3	3	5 0	91
		204	3	1	30	13	8	14	0	1	3	3	5 0	1 66
Avery, & Davis	1983	222	1	2	'2	1	9	16	0	1	3	3	1 7	84
		222	2	2	32	1	9	16	0	1	3	3	1. 7	1 02
		222	3	2	16	1	8	16	0	1	3	3	1' 7	1. 40
Brady	1966	22	1	8	70	7	8	1	0	5	3	3	1. 0	81
Levinson	1973	64	1	8	18	14	8	16	0	5	3	3	2	48
Dampier	1982	246	1	9	30	1	8	1	0	0	9	3	4	3 11
		246	2	9	20	1	8	1	0	0	9	3	4	1 07
Simon	1970	46	1	15	211	1	8	1	0	0	3	3	3	- 01
		46	2	15	211	1	8	1	0	0	3	3	3	01
		46	3	15	211	1	8	1	0	0	3	3	3	- 24
		46	4	15	211	1	8	1	0	0	3	3	3	- 05
		46	5	15	211	1	8	1	0	0	3	3	3	05
		46	6	15	211	1	8	1	0	0	3	3	3	- 12
		46	7	15	211	1	8	1	0	0	3	3	3	32
Dye	1978	34	2	6	21	12	9	6	0	2	3	3	24 0	54
		34	4	6	21	12	9	6	0	2	3	3	24 0	- 02
		34	6	6	21	12	9	6	0	2	3	3	24 0	44
		34	8	6	21	12	9	6	0	2	3	3	24 0	67
Sawyer, & Clark	1980	153	1	20	38	15	9	16	0	0	9	3	25 5	12
		153	2	20	38	15	9	16	0	0	9	3	25 5	1 31
		153	3	20	38	15	9	16	0	0	9	3	25 5	71
		153	4	20	38	15	9	16	0	0	9	3	25 5	1. 03
Th...nburg	1983	235	1	1	40	0	10	0	5	0	11	3	1 0	1 23
		235	2	1	40	0	10	0	5	0	11	3	1 0	1 23
Steen	1980	89	1	1	30	0	10	0	15	1	3	3	4 5	19
Rothschild	1978	139	1	1	34	1	10	0	5	1	3	3	20 0	- 70
Euse	1975	124	1	2	20	0	8	16	0	1	3	4	6 0	1 75
		124	2	2	20	0	8	16	0	1	3	4	6 0	1 73
Forader	1969	94	1	2	68	1	6	0	0	1	3	5	3	40

AUTHOR	YEAR	ID#	ESID#	ATTOWARD	TOTN	SETTINGE	EDLEVEXP	UNIVMAJE	OCCUPEXP	COMINSTR	ASSESTYP	TRTECHE	TOTHRSE	ES
Foraker	1969	94	[illegible]	2	72	1	6	0	0	1	3	5	3	43
		94	3	2	72	1	6	0	0	1	3	5	3	34
		94	4	2	63	1	6	0	0	1	3	5	3	25
		94	5	2	64	1	6	0	0	1	3	5	3	47
		94	6	2	62	1	6	0	0	1	3	5	3	19
Gottlieb	1930	47	1	3	26	1	7	0	0	0	15	5	2	2 42
		47	2	3	26	1	7	0	0	0	15	5	2	1 78
		47	3	3	26	1	7	0	0	0	15	5	2	93
Feldman, & Feldman	1985	14	1	1	60	13	8	5	0	1	3	5	2 5	65
		14	2	1	60	13	8	5	0	1	3	5	2 5	71
		14	3	1	60	13	8	5	0	1	3	5	2 5	75
Donaldson	1974	31	1	2	49	1	8	1	0	1	3	5	8	1 32
		31	2	2	48	1	8	1	0	1	3	5	8	66
		31	3	2	55	1	8	1	0	1	3	5	8	39
Morrison, et al.	1978	100	1	6	38	1	8	1	0	0	9	5	2 0	37
		100	2	6	38	1	8	1	0	0	9	5	2 0	45
Dye	1978	34	1	6	20	12	9	6	0	2	3	5	24 0	70
		34	3	6	20	12	9	6	0	2	3	5	24 0	30
		34	5	6	20	12	9	6	0	2	3	5	24 0	- 36
		34	7	6	20	12	9	6	0	2	3	5	24 0	74
Grove	1978	173	1	6	88	15	11	0	1	0	3	5	4 0	69
		173	2	6	88	15	11	0	1	0	6	5	4 0	94
Gerstein	1976	260	1	1	90	1	7	5	0	1	3	6	1	- 08
Sanders	1978	207	1	1	40	1	8	5	0		3	6	1 2	03
		207	2	1	39	1	8	5	0	1	3	6	1 2	62
		207	3	1	33	1	8	5	0	1	3	6	5	26
		207	4	1	40	1	8	5	0	1	3	6	7	11
		207	5	1	38	1	8	5	0	1	3	6	7	- 07
		207	6	1	40	1	8	5	0	1	3	6	1 2	- 15
		207	7	1	39	1	8	5	0	1	3	6	1 2	82
		207	8	1	33	1	8	5	0	1	3	6	5	19
		207	9	1	40	1	8	5	0	1	3	6	7	00
		207	10	1	38	1	8	5	0	1	3	6	7	- 28
Simpson, et al.	1976	2	1	1	38	1	2	0	0	1	3	7	20 0	55
Rae	1983	108	1	5	30	1	2	0	0	0	3	7	1 3	1. 20
		108	5	5	30	14	2	0	0	0	3	7	1 3	72
Cronk	1978	186	2	5	39	15	2	0	0	0	3	7		33
		186	3	5	40	15	2	0	0	0	3	7		51
Sasso	1983	193	1	7		15	2	0	0	0	3	7	11 0	1 38
Vinish	1974	28	1	1	60	1	3	0	0	0	3	7	20 0	30
Wurzel	1980	239	1	1	47	1	3	0	0	0	6	7	6 0	1 06
		239	2	1	47	1	3	0	0	0	3	7	6 0	56
		239	7	1	24	1	3	0	0	0	6	-	6 0	1 18
		239	8	1	24	1	3	0	0	0	3	7	6 0	91
Marquart	1984	35	2	1	44	2	3	0	0	0	3	7	4 2	37
		35	6	1	45	2	3	0	0	0	15	7	4 2	- 48
		35	10	1	44	2	3	0	0	0	6	7	4 2	- 55
Mulkey	1980	88	5	1	34	1	3	0	0	1	3	7	3	22
Parrish	1974	270	1	1	28	1	3	0	0	1	3	7	8 5	49
		270	2	1	28	1	3	0	0	1	3	7	8 5	- 11
Lazar, et al.	1971	76	1	1	44	14	3	0	0	1	3	7		91

AUTHOR	YEAR	ID#	ESID#	ATTOWARD	TOTN	SETTINGE	EDLEVEXP	UNIVMAJE	OCCUPEXP	COMINSTR	ASSESTYP	TRTECHE	TOTHRSE	ES
Olsen	1978	127	1	4	55	2	3	0	0	0	3	7	25 0	83
		127	3	4	49	2	3	0	0	0	17	7	25 0	2 75
Cronk	1978	186	4	5	48	15	3	0	0	0	3	7		38
Dickson	1974	244	1	1	46	15	4	0	0	0	3	7		1 10
		244	2	1	40	15	4	0	0	1	3	7		61
Rae	1983	109	3	5	30	1	4	0	0	0	3	7	1 3	47
		[illegible]	7	5	30	14	4	0	0	0	3	7	1 3	59
Mulkey	1980	88	7	1	46	1	5	0	0	1	3	7	3	33
McKerracher	1982	120	5	2	94	1	5	0	0	1	3	7	3 7	19
Mulkey	1980	88	9	1	47	1	6	0	0	1	3	7	3	68
Farrell	1985	18	2	1	40	14	7	0	0	0	4	7		87
Gronberg	1982	96	1	1	293	1	7	0	0	1	3	7	42 0	64
Baker, et al.	1981	203	1	1	1192	1	7	0	0	1	3	7	6	35
		203	3	1	1631	1	7	0	0	1	3	7	6	37
Kiernan	1974	84	2	1	97	1	8	2	0	1	3	7	2 5	60
		84	4	1	97	1	8	2	0	1	3	7	2 5	31
		84	1	1	76	1	8	13	0	1	3	7	2 5	99
		84	3	1	76	1	8	13	0	1	3	7	2 5	88
Leyser, & Abrams	1983	152	2	1	142	15	8	2	0	1	3	7		37
		152	1	1	254	15	8	13	0	1	3	7		43
Felton	1975	38	1	1	7	15	8	16	0	1	3	7	999 9	1 27
Nelson	1970	53	3	2	44	1	8	4	0	0	9	7		1 31
Ozyurek	1977	233	1	2	46	0	8	2	0	1	3	7		55
Aukayanagul	1979	137	1	2	20	0	8	15	0	1	3	7	1 3	57
		137	5	2	20	0	8	16	0	1	3	7	1 3	43
		137	13	2	20	0	8	16	0	1	3	7	1 3	2 27
Rosswurm	1980	209	1	2	83	15	8	6	0	.	3	7		36
		209	2	2	67	15	8	6	0		3	7		- 16
Nelson	1970	53	2	3	44	1	8	4	0	[illegible]	9	7		89
Newcomer	1975	232	1	3	14	4	8	6	0	0	3	7	137 2	1.24
		232	2	3	11	4	8	6	0	0	3	7	137 2	1.54
		232	3	3	14	4	8	6	0	0	3	7	137 2	2 14
Hourcade	1981	20	1	3	27	9	8	13	0	0	3	7	120 0	04
Baker	1974	128	1	3	72	15	8	6	0	0	6	7		- 02
		128	2	3	72	15	8	6	0	0	17	7		- 38
		128	3	3	72	15	8	6	0	0	3	7		- 21
Boehm	1978	32	1	5	74	15	8	13	0	0	9	7	205 0	- 02
		32	3	5	65	15	8	13	0	0	9	7	205 0	27
Morris	1964	71	1	6	56	6	8	6	0	2	3	7		1 20
		71	2	6	56	6	8	6	0	2	3	7		04
Walsh	1970	99	1	6	158	6	8	6	0	2	3	7	60 0	33
		99	2	6	158	6	8	6	0	2	3	7	60 0	- 42
Schildhaus	1971	145	1	6	85	6	8	6	0	2	3	7		- 04
Nelson	1970	53	1	7	44	1	8	4	0	0	9	7		1 02
Skrtic, et al.	1982	132	2	8	63	15	8	5	0	5	3	7		55
		132	4	8	51	15	8	5	0	5	3	7		87
Fox, et al.	1984	30	1	19	33	15	8	16	0	0	9	7		- 13
		30	2	19	33	15	8	16	0	0	3	7		1.25
Naor, & Milgram	1980	177	1	20	53	15	8	2	0	0	3	7	22 5	3 00
		177	3	20	53	15	8	2	0	0	17	7	22 5	1 34

AUTHOR	YEAR	ID#	ESID#	ATTOWARD	TOTN	SETTINGE	EDLEVE(P	UNIVMAJE	OCCUPE(P	COMINSTR	ASSESTYP	TRTECHE	TOTHRSE	ES
Anthony	1968	7	1	2	16	14	9	10	0	1	3	7		45
Gelfand, & Ullmann	1961	41	1	6	55	6	9	15	0	2	3	7		15
		41	2	6	55	6	9	15	0	2	3	7		09
Meltzer, & Grigorian	1972	216	1	6	90	6	9	15	0	2	3	7	160 0	66
		216	2	6	90	6	9	15	0	2	3	7	160 0	- 07
Crow, et al.	1970	217	1	6	90	15	9	15	0	2	3	7		06
Moosbruker, & Giddon	1966	154	2	20	153	14	9	16	0	0	17	7	24 0	- 10
		154	3	20	153	14	9	16	0	0	17	7	24 0	42
		154	4	2c	153	14	9	16	0	0	17	7	24 0	- 16
		154	5	20	150	14	9	16	0	0	17	7	24 0	- 39
		154	6	20	150	14	9	16	0	0	17	7	24 0	- 18
		154	1	20	185	14	9	16	0	1	3	7	24 0	.21
Hagton	1980	61	1	1	93	1	10	0	15	3	6	7	4 0	- 14
		61	2	1	93	1	10	0	15	4	3	7	4 0	03
Finell, et al.	1980	117	1	1	25	15	10	0	15	4	3	7	44 0	1 35
Kauffman	1976	157	2	4	24	15	10	0	5	0	3	7	1 5	96
		157	5	4	24	15	10	0	5	0	3	7	1 5	44
		157	8	4	24	15	10	0	5	0	3	7	1 5	03
		157	11	4	24	15	10	0	5	0	3	7	1 5	46
Cronk	1978	186	1	5	24	14	11	0	15	0	3	7		52
Pryer, & Distefano	1978	29	1	6	71	4	11	0	4	2	3	7	200 0	23
		29	2	6	71	4	11	0	4	2	3	7	200 0	71
		29	3	6	28	4	11	0	4	2	3	7	200 0	53
		29	4	6	28	4	11	0	4	2	3	7	200 0	31
Stockton	1978	256	1	6	58	4	11	0	4	2	3	7	320 0	47
		256	2	6	58	4	11	0	4	2	3	7	320 0	- 89
		256	3	6	58	4	11	0	4	2	3	7	320 0	92
		256	4	6	58	4	11	0	4	2	3	7	320 0	-1.07
Pryer, et al.	1969	91	1	6	39	15	11	0	15	2	3	7	200 0	.11
		91	2	6	39	15	11	0	15	2	3	7	200 0	30
		91	3	6	25	15	11	0	15	2	3	7	200 0	14
		91	4	6	25	15	11	0	15	2	3	7	200 0	57
Horan	1981	175	1	2	50	1	1	0	0	0	6	8	5 0	15
		175	2	2	50	1	1	0	0	0	6	8	5 0	- 12
Leung	1979	257	1	1	70	1	2	0	0	0	3	8	2 5	63
Cray-Andrews	1981	130	1	10	140	1	2	0	0	0	3	8	1 7	21
Salend, & Moe	1983	182	2	1	153	1	3	0	0	0	15	8		63
Maas	1983	263	1	1	109	1	3	0	0	0	6	8	8 0	18
		263	2	1	108	1	3	0	0	0	6	8	8 0	18
		263	3	1	108	1	3	0	0	0	6	8	8 0	05
		263	4	1	108	1	3	0	0	0	6	8	8 0	- 12
		263	5	1	109	1	3	0	0	0	6	8	8 0	15
		263	6	1	108	1	3	0	0	0	6	8	8 0	07
		263	7	1	109	1	3	0	0	0	6	8	8 0	- 19
		263	8	1	108	1	3	0	0	0	6	8	8 0	- 01
		263	9	1	109	1	3	0	0	0	6	8	8 0	22
		263	10	1	108	1	3	0	0	0	6	8	8 0	17
Kologinsky	1981	23	1	15	71	9	3	0	0	0	6	8	1 0	33
		23	2	15	71	9	3	0	0	0	6	8	1 0	- 13
Heitzman	1982	73	1	3	53	1	6	0	0	1	3	8	44 9	03
Stemplinger	1985	273	1	10	107	1	6	0	0	0	15	8	3 /	1 17

AUTHOR	YEAR	ID#	ESID#	ATTOWARD	TOTN	SETTINGE	EDLEVEXP	UNIVMAJE	OCCUPEXP	COMINSTR	ASSESTYP	TRTECHE	TOTHRSE	ES
Stemplinger	1985	273	2	10	107	1	6	0	0	0	17	8	3 7	78
Hazzard	1981	45	1	1	325	1	7	0	0	0	3	8	4 5	61
		45	2	1	325	1	7	0	0	0	3	8	4 5	33
		45	3	1	325	1	7	0	0	0	6	8	4 5	11
		45	4	1	325	1	7	0	0	0	6	8	4 5	03
		45	5	1	64	1	7	0	0	0	8	8	4 5	44
		45	15	1	189	1	7	0	0	0	17	8	4 5	-17
		45	16	1	190	1	7	0	0	0	17	8	4 5	00
Lundstrom	1979	118	1	1	31	1	7	16	0	0	9	8		20
		118	2	1	31	1	7	16	0	0	9	8		-07
		118	3	1	31	1	7	16	0	0	9	8		-.11
		118	4	1	31	1	7	16	0	0	9	8		-11
		118	5	1	31	1	7	16	0	0	9	8		.23
		118	6	1	31	1	7	16	0	0	9	8		25
		118	7	1	31	1	7	16	0	0	9	8		37
		118	8	1	30	1	7	16	0	0	9	8		35
		118	9	1	31	1	7	16	0	0	17	8		28
Shortridge	1982	240	1	2	424	2	7	0	0	0	17	8	7	19
		240	2	2	424	2	7	0	0	0	17	8	7	12
		240	3	2	424	2	7	0	0	0	17	8	7	46
		240	4	2	424	2	7	0	0	0	17	8	7	14
Krieger	1976	116	1	8	39	1	7	16	0	5	3	8	2 0	69
		116	2	8	39	1	7	16	0	5	3	8	1 8	40
Holmquist	1981	167	1	1	15	1	8	1	0	0	3	8	8 0	67
Elliott, & Byrd	1984	37	1	1	30	1	8	1	0	1	3	8	7	70
		37	3	1	23	1	8	1	0	1	3	8	7	47
Peterson	1977	136	1	1	32	1	8	13	0	1	3	8	3. 0	-31
Parish, et al.	1982	180	1	2	48	1	8	3	0	0	15	8	2 5	.35
		180	3	3	51	1	8	5	0	0	15	8	2. 5	31
		180	2	7	53	1	8	5	0	0	15	8	2 5	87
		180	4	10	53	1	8	5	0	0	15	8	2 5	20
Pomerantz	1983	3	3	1	57	12	10	0	15	0	3	8	2 0	67
		3	1	1	71	12	10	0	15	1	3	8	2 0	-51
		3	2	1	60	12	10	0	15	1	3	8	2 0	-41
Royster	1981	6	4	1	66	0	11	0	15	0	17	8	7 5	28
		6	6	1	70	0	11	0	15	0	17	8	7 5	60
		6	8	1	73	0	11	0	15	0	17	8	7 5	08
		6	10	1	70	0	11	0	15	0	17	8	7 5	.40
Robertson	1984	109	2	1	39	4	11	0	4	0	3	8	1 0	-14
		109	4	1	27	4	11	0	4	0	3	8	1 0	-28
Royster	1981	6	1	1	74	0	11	0	15	1	3	8	7 5	67
Robertson	1984	109	1	1	39	4	11	0	4	1	3	8	1 0	-76
		109	3	1	28	4	11	0	4	1	3	8	1 0	-33
Esposito, & Peach	1983	39	1	19	9	1	1	0	0	0	2	9	21 0	1 83
Jones, et al.	1981	67	1	1	74	1	2	0	0	0	3	9	5 0	1 69
Hawisher	1977	196	1	15	91	1	2	0	0	0	3	9	3 3	04
Esposito, & Peach	1983	39	2	19	9	1	2	0	0	0	2	9	21 0	1 50
Marquart	1984	35	4	1	43	2	3	0	0	0	3	9	4 2	21
		35	8	1	43	2	3	0	0	0	15	9	4 2	-16
		35	12	1	40	2	3	0	0	0	6	9	4 2	-79
Shein	1978	194	1	1	40	1	3	0	0	1	3	9	6 0	37

AUTHOR	YEAR	ID#	ESID#	ATTOWARD	TOTN	SETTINGE	EDLEVEXP	UNIVMAJE	OCCUPEXP	COMINSTR	ASSESTYP	TRTECHE	TOTHRSE	ES
Shein	1978	194	3	1	40	1	3	0	0	1	3	7	3 0	37
		194	4	1	40	1	3	0	0	1	3	9	6 0	63
		194	6	1	40	1	3	0	0	1	3	9	3 0	49
Schwartzwald	1981	191	3	2	156	15	3	0	0	1	3	9		16
		191	4	2	163	15	3	0	0	1	3	9		37
Shein	1978	194	7	9	40	1	3	0	0	0	3	9		- 10
		194	9	9	40	1	3	0	0	0	3	9	6.0	62
		194	10	9	40	1	3	0	0	0	3	9	3 0	- 02
		194	12	9	40	1	3	0	0	0	3	9	6 0	52
Owen	1984	176	1	1	24	1	4	0	0	0	3	9	3 0	72
		176	3	1	24	1	4	0	0	0	3	9	2 7	10
		176	5	1	24	1	4	0	0	0	3	9	2 7	49
		176	7	1	24	1	4	0	0	0	3	9	2 7	- 06
McKerracher	1982	120	4	2	94	14	5	0	0	1	3	9	2 7	62
		120	1	2	86	15	5	0	0	1	3	9	1 7	58
		120	2	2	89	15	5	0	0	1	3	9	5. 1	59
		120	3	2	88	15	5	0	0	[illegible]	3	9	4 1	97
Harrell	1982	60	1	20	140	1	6	0	0	0	6	9	2 7	76
		60	2	20	134	1	6	0	0	0	6	9	8. 3	74
Pando	1984	80	1	1	29	11	8	1	0	1	3	9	8 3	64
		80	2	1	28	11	8	1	0	1	3	9	1	1. 09
		80	3	1	30	11	8	1	0	1	3	9	1	. 64
		80	4	1	30	11	8	[illegible]	0	1	3	9	1	35
		80	5	1	30	11	8	1	0	1	3	9	1	41
		80	6	1	30	11	8	1	0	1	3	9	1	87
Cohen	1978	220	4	2	33	15	8	0	0	0	6	9	. 1	. 45
		220	7	2	33	15	8	0	0	0	9	9	5. 3	. 47
		220	10	2	33	15	8	0	0	0	9	9	5 3	71
		220	13	2	33	15	8	0	0	0	9	9	5 3	44
		220	1	2	33	15	8	0	0	1	3	9	5. 3	1 68
Jackson	1976	231	1	3	295	1	8	1	0	0	3	9	5 3	2 26
		231	2	3	291	1	8	1	0	0	3	9	6	2. 00
Frazier	1975	62	1	6	27	15	8	1	0	0	9	9	6	51
		62	4	6	27	15	8	1	0	0	9	9		- 13
		62	7	6	27	15	8	1	0	2	3	9		81
		62	10	6	27	15	8	1	0	2	3	9		17
Towner	1979	103	1	9	35	12	8	2	0	0	3	9	1. 0	- 44
		103	2	9	33	12	8	2	0	0	3	9	1 0	- 44
		103	3	9	33	12	8	2	0	0	3	9	1 0	- 55
		103	4	9	35	12	8	2	0	0	3	9	1 0	- 47
		103	5	9	33	12	8	2	0	0	3	9	1 0	- 44
		103	6	9	33	12	8	2	0	0	3	9	1 0	- 57
		103	7	9	16	12	8	2	0	0	3	9	1 0	- 65
		103	8	9	15	12	8	2	0	0	3	9	1 0	- 14
		103	9	9	13	12	8	2	0	0	3	9	1 0	-1 22
Lazar, et al.	1973	77	1	1	20	1	9	13	0	1	3	9	44 0	1 12
Lennington	1973	4	1	1	187	15	10	0	15	0	3	9		- 10
Smith	1980	178	1	1	117	0	10	0	15	1	3	9	7 0	23
Cortez	1983	33	1	1	41	1	10	0	15	1	3	9	5 7	86
Lapp	1974	19	1	1	27	12	10	0	5	3	6	9	360 0	76
Chow, et al.	1976	142	1	4	37	[illegible]	10	0	5	0	3	9	24 0	67

AUTHOR	YEAR	ID#	ESID#	ATTOWARD	TOTN	SETTINGE	EDLEVEXP	UNIVMAJE	OCCUPEXP	COMINSTR	ASSESTYP	TRTECHE	TOTHRSE	ES
Royster	1981	6	5	1	80	0	11	0	15	0	17	9	7 5	21
		6	7	1	80	0	11	0	15	0	17	9	7 5	14
		6	9	1	83	0	11	0	15	0	17	9	7 5	07
		6	11	1	81	0	11	0	15	0	17	9	7 5	- 12
		6	2	1	88	0	11	0	15	1	3	9	7 5	28
Perry	1983	106	1	1	48	15	11	0	15	1	3	9	5.3	37
		106	2	1	25	15	11	0	15	1	3	9	5.3	30
Brantlinger	1978	208	1	3	59	14	11	0	15	0	3	9	8.0	51
		208	3	3	59	14	11	0	15	0	3	9	8.0	20
		208	2	3	50	15	11	0	4	0	3	9	8 0	89
		208	4	3	50	15	11	0	4	0	3	9	8.0	38
Friedel	1980	188	1	20	154	15	12	3	0	3	6	9	6.0	- 39
Clark	1978	224	3	20	40	1	7	13	0	0	3	10	4 5	56
		224	5	20	40	1	7	13	0	0	9	10	4 5	1.07
		224	4	20	40	15	7	13	0	0	3	10	4 2	63
		224	6	20	40	15	7	13	0	0	9	10	4 2	58
		224	1	20	40	1	7	13	0	1	3	10	4	62
		224	2	20	40	15	7	13	0	1	3	10	4 .	80
Haddle	1973	69	1	1	87	11	8	1	0	1	3	10	3.3	13
Cohen	1978	220	5	2	37	11	8	0	0	0	6	10	4 0	11
		220	8	2	37	11	8	0	0	0	9	10	4 0	12
		220	11	2	37	11	8	0	0	0	9	10	4 0	17
		220	14	2	37	11	8	0	0	0	9	10	4 0	- 16
		220	2	2	37	11	8	0	0	1	3	10	4 0	1 26
Frazier	1975	62	2	6	26	11	8	1	0	0	9	10	3 0	- 06
		62	5	6	26	11	8	1	0	0	9	10	3.0	- 16
		62	8	6	26	11	8	1	0	2	3	10	3.0	51
		62	11	6	26	11	8	1	0	2	3	10	3.0	- 69
Haddle	1973	69	2	8	87	11	8	1	0	0	3	10	3.3	27
		69	3	8	87	11	8	1	0	0	3	10	3.3	32
		69	4	8	87	11	8	1	0	0	3	10	3.3	32
		69	5	8	87	11	8	1	0	0	3	10	3.3	30
		69	6	8	87	11	8	1	0	0	9	10	3.3	02

NUMBER OF CASES READ = 643 NUMBER OF CASES LISTED - 643

TREATMENT A VERSUS TREATMENT B EFFECT SIZES

KEY

The key for these effect sizes is the same as for the T vs. C, T vs. P, and Pre-post effect sizes, except:

TRTECHE is Treatment A, and
TRTECHC is Treatment B.

TREATMENT A VERSUS TREATMENT B EFFECT SIZES

AUTHOR	YEAR	ID#	ESID#	ATTOWARD	TOTN	SETTINGE	EDLEVEXP	UNIVMAJE	OCCUPEXP	COMINSTR	ASSESTYP	TRTECHE	TRTECHC	TOTHRSE	ES
Dresang	1981	52	1	1	76	1	3	0	0	0	4	1	2	7	05
		52	4	1	114	1	3	0	0	0	9	1	2	7	-12
Mulkey	1980	88	1	1	32	1	3	0	0	1	3	1	8	3	-29
		88	2	1	51	1	5	0	0	1	3	1	8	3	-46
		88	3	1	47	1	6	0	0	1	3	1	8	3	52
Oberle	1975	200	3	1	80	0	7	16	0	1	3	1	2	1	3[illegible]
January	1978	165	3	1	63	1	8	1	0	0	9	1	2	2	-10
		165	4	1	63	1	8	1	0	0	9	1	2	2	16
Dailey	1978	184	2	1	52	1	8	1	0	0	3	1	2	47 5	-06
January	1978	165	1	1	63	1	8	1	0	1	3	1	2	2	05
		165	2	1	63	1	8	1	0	1	3	1	2	2	10
Dailey	1978	184	1	1	52	1	8	1	0	1	3	1	2	47 5	1.12
Noe	1982	140	1	1	74	1	8	4	0	4	3	1	2	28 3	-20
		140	2	1	59	.	8	4	0	4	3	1	8	28 3	-55
		140	3	1	57	1	8	4	0	4	3	1	8	28 3	-35
		140	4	1	74	1	8	4	0	4	3	1	2	28 3	-32
		140	5	1	59	1	8	4	0	4	3	1	8	28 3	-95
		140	6	1	57	1	8	4	0	4	3	1	8	28 3	-64
		140	7	1	74	1	8	4	0	4	3	1	2	28 3	-25
		140	8	1	59	1	8	4	0	4	3	1	8	28 3	-58
		140	9	1	57	1	8	4	0	4	3	1	8	28 3	-33
Graham	1968	48	1	6	63	1	8	1	0	2	3	1	2	40 0	42
		48	2	6	63	1	8	1	0	2	3	1	2	40 0	41
January	1978	165	5	14	63	1	8	1	0	0	9	1	2	2	08
		165	6	14	63	1	8	1	0	0	9	1	2	2	27
Bryant	1985	268	1	19	102	1	9	16	0	0	3	1	2	1.0	23
Matkin et al.	1983	229	1	1	3[illegible]	0	12	16	15	1	3	1	2	.7	09
		229	2	1	2[illegible]	0	12	16	15	1	3	1	2	7	20
Emerton & Rothman	1975	151	2	9	52	7	8	1	0	0	1	2	2	.	.02
MacIntyre	1981	163	1	1	51	14	3	0	0	0	3	2	3	5	45
		163	2	1	47	14	3	0	0	0	3	2	3	5	69
		163	3	1	48	14	3	0	0	0	3	2	3	5	23
Rynders	1980	212	3	5	12	12	5	0	0	0	16	2	3	8 0	84
Oberle	1975	200	1	1	80	11	7	16	0	1	3	2	2		-43
		200	2	1	80	11	7	16	0	1	3	2	2		-04
Cautela	1971	221	1	3	42	15	8	1	0	0	3	4	11		1 26
Oberle	1975	200	4	1	120	15	7	16	0	1	3	7	8		-30
Fisher	1975	42	1	1	50	2	8	13	0	0	3	7	8		14
		42	2	1	50	2	8	13	0	0	3	7	8		-11
		42	3	1	50	2	8	13	0	0	3	7	8		-10
Dyer	1970	56	1	1	67	15	8	16	0	0	3	7	2		18
		56	2	1	67	15	8	16	0	0	6	7	2		44
		56	3	1	67	15	8	16	0	0	6	7	2		48
Royster	1981	6	3	1	55	0	11	0	15	0	17	8	10	7 5	29
		6	12	1	85	0	11	0	15	0	17	8	10	7 5	30
		6	13	1	80	0	11	0	15	0	17	8	10	7 5	34
Rynders	1980	212	1	5	12	12	5	0	0	0	16	9	3	8.0	2 71
		212	2	5	12	12	5	0	0	0	16	9	3	8 0	3 55
Hazzard	1981	45	6	1	150	1	7	0	0	0	3	9	9	5 0	-03
		45	7	1	150	1	7	0	0	0	3	9	9	5 0	00
		45	8	1	150	1	7	0	0	0	6	9	9	5 0	-21
		45	9	1	150	1	7	0	0	0	6	9	9	5 0	00

AUTHOR	YEAR	ID#	ESID#	ATTOWARD	TOTN	SETTINGE	EDLEVEXP	UNIVMAJE	OCCUPEXP	COMINSTR	ASSESTYP	TRTECHE	TRTECHC	TOTHRSE	ES
Newman	1978	199	1	7	205	15	7	0	0	0	8	9	9	4 5	-1 07
		199	2	7	205	15	7	0	0	0	8	9	9	4 5	1[illegible]
		199	4	7	205	15	7	0	0	0	9	9	9	4 5	2[illegible]
		199	3	7	205	15	7	0	0	1	3	9	9	4 5	- 02
Rodriguez	1978	97	2	1	71	1	8	16	0	0	6	9	7	2 0	- 05
		97	3	1	80	1	8	16	0	0	17	9	7	2 0	- 43
		97	1	1	71	1	8	16	0	1	3	9	7	2 0	- 13
Meyer	1963	181	1	1	34	15	8	6	0	1	3	9	2	10 0	4[illegible]
		181	2	1	24	15	8	6	0	1	3	9	2	15 0	27

NUMBER OF CASES READ = 61 NUMBER OF CASES LISTED = 61

100 mm

A5

1.0 1.1 1.25 1.4 1.6 1.8 2.0 2.2 2.5

ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz1234567890

ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz
1234567890

1.0 mm

1.5 mm

2.0 mm

MAINSTREAMING EFFECT SIZES

KEY

The key for these effect sizes is the same as for the T vs. C, T vs. P, and Pre-post effect sizes except:

Column	Codes
MSINSTRC	The type of instruction in the mainstreamed classrooms: 0=can't tell 1=standard group/class instruction 2=cooperative learning 3=individualized instruction 4=peer tutoring 5=combination 6=other
MSDISCHD	The disabilities of the mainstreamed children: 0=can't tell 1=mildly and moderately retarded 2=emotionally disturbed 3=visually impaired 4=blind 5=hearing impaired 6=deaf 7=communication disordered 8=physically and health impaired 9=learning disabled 10=combination 11=others
MSPEERS	Any special instruction for nondisabled peers in the mainstreamed classroom: 0=can't tell 1=none 2=information 3=vicarious experience 4=reinforcement 5=persuasive messages
MSMONTHS	The number of months the experimental (nondisabled) Ss participated in mainstreaming. If no number is entered, the information was not available.

MAINSTREAMING EFFECT SIZES

AUTHOR	YEAR	ID#	ESID#	TOTN	TRTCONXT	EDLEVEXP	MSINSTRC	MSDISCHD	MSPEERS	COMINSTR	ASSESTYP	MSMONTHS	ES
Sigler et al.	1978	169	1	180	1	3	0	0	0	4	3		10
Brighi	1978	24	2	96	1	3	0	8	0	1	3		.39
		24	3	83	1	3	0	8	0	1	3		.49
		24	4	69	1	3	0	8	0	1	3		.10
Voeltz	1982	271	1	520	1	3	0	10	2	0	3	17	02
		271	2	537	1	3	0	10	1	0	3	13	- 16
		271	7	520	1	3	0	10	2	0	3	17	19
		271	8	537	1	3	0	10	1	0	3	13	- 07
Rapier et al.	1972	166	1	148	1	3	1	8	1	0	9	9	31
Smith & Larson	1980	168	1	154	1	5	0	10	1	1	3	18	72
Strauch	1970	170	1	124	1	5	1	1	1	0	9		05
		170	2	124	1	5	1	1	1	0	9		24
Brighi	1978	24	1	542	1	7	0	8	0	[illegible]	3		43
Voeltz	1980	187	1	856	1	7	0	10	2	[illegible]	3	8	1 05
		187	2	877	1	7	0	10	1	0	3	6	.41
		187	3	856	1	7	0	10	2	0	3	8	.36
		187	4	877	1	7	0	10	1	0	3	6	- 01
Moseley	1973	135	1	80	1	7	1	0	1	1	3	9	.16
		135	2	80	1	7	1	0	1	1	3	9	-.81
Winkler	1981	54	1	114	1	10	0	11	1	0	11		.07

NUMBER OF CASES READ = 20 NUMBER OF CASES LISTED = 20

MISSING INFORMATION RESULTS
(No Treatment A vs. B results)

KEY

The key for these effect sizes is the same as for the T vs. C, T vs. P, and Pre-post effect sizes, except:

Instead of an ES column, there is an ESAVAIL (Effect Size Available) column, for which the codes are:

1=No, can't tell direction of result
2=No, negative result
3=No, positive result

MISSING INFORMATION RESULTS
(No Treatment A vs. B results)

AUTHOR	YEAR	ID#	ESID#	ATTOWARD	TOTN	SETTINGE	EDLEVEXP	UNIVMAJE	OCCUPEXP	COMINSTR	ASSESTYP	TRTECHE	TOTHRSE	ESAVAIL
Baran	1977	101	1	3	80	3	0	0	1	0	11	1	2 0	3
Miller et al.	1981	171	1	1	42	1	2	0	0	0	17	1	6 0	1
DuHoux	1979	197	5	1	52	1	2	0	0	0	3	1	2	3
Sasso et al.	1985	249	2	16	36	1	2	0	0	0	3	1	8 0	3
Miller et al.	1981	171	2	1	29	1	3	0	0	0	17	1	6 0	3
Simpson et al.	1976	2	2	1	28	1	3	0	0	1	3	1	4 0	1
		2	3	1	28	1	3	0	0	1	3	1	4 0	1
Meehan	1980	192	5	1	43	1	4	0	0	0	3	1	9	1
		192	6	1	43	1	4	0	0	0	3	1	9	1
		192	7	1	43	1	4	0	0	0	3	1	9	1
		192	8	1	43	1	4	0	0	0	3	1	9	1
		192	1	1	43	1	4	0	0	1	3	1	9	1
		192	2	1	43	1	4	0	0	1	3	1	9	1
		192	3	1	43	1	4	0	0	1	3	1	9	1
		192	4	1	43	1	4	0	0	1	3	1	9	1
Schroeder	1978	251	1	1	70	1	5	0	0	0	9	1	7	1
		251	2	1	70	1	5	0	0	0	9	1	7	1
		251	3	1	70	1	5	0	0	0	9	1	7	1
		251	4	1	70	1	5	0	0	0	9	1	7	1
Rusalem	1967	248	1	20	14	1	6	0	0	0	3	1	5 0	1
		248	2	20	14	1	6	0	0	0	3	1	5 0	3
Austin et al.	1980	68	2	1	.	1	8	16	0	4	3	1	2	3
Spreen	1977	27	1	3	35	1	8	16	0	0	9	1	103 2	3
		27	2	3	88	1	8	16	0	0	9	1	103 2	3
		27	3	3	54	1	8	16	0	0	9	1	103 2	3
Prothero & Ehlers	1974	205	1	4	38	1	8	11	0	0	9	1		3
Dixon	1967	98	1	6	19	1	8	6	0	2	3	1		2
Naor & Milgram	1980	177	6	20	54	1	8	2	0	0	17	1	22 5	3
Parish et al.	1977	211	1	20	45	1	8	16	0	0	15	1		1
		211	2	20	45	1	8	16	0	0	15	1		1
		211	3	20	45	1	8	16	0	0	15	1		1
Harasymiw & Horne	1976	105	1	1	352	0	10	0	15	0	6	1		3
Ingram	1976	172	1	1	118	1	10	0	15	0	3	1	155 0	3
		172	2	1	118	1	10	0	15	0	3	1	155 0	3
Aldridge	1978	245	1	1	114	15	10	0	0	0	3	1		1
Ashmore	1981	174	1	1	23	0	10	0	8	1	3	1	.	2
Frith & Lindsey	1981	162	1	1	76	1	10	0	15	1	3	1	50 0	3
Wallace	1980	149	1	1	75	1	10	0	15	3	6	1	24 0	2
D'Zamko	1979	247	1	1	60	1	10	0	15	3	6	1	10 0	1
Kromer	1976	261	1	20	.	1	10	0	5	3	6	1	8.0	2
		261	2	20	.	1	10	0	5	3	6	1		3
Messinger-Revell	1983	265	1	1	19	12	11	0	11	0	3	1	13 5	3
		265	2	1	18	12	11	0	11	0	3	1	13 5	3
		265	3	1	19	12	11	0	11	0	3	1	13.5	3
		265	4	1	18	12	11	0	11	0	3	1	13 5	1
		265	5	1	19	12	11	0	11	0	3	1	13 5	3
		265	6	1	18	12	11	0	11	0	3	1	13 5	1
		265	7	1	19	12	11	0	11	0	3	1	13 5	1
		265	8	1	18	12	11	0	11	0	3	1	13 5	3
Threlkeld & DeJong	1983	219	1	2	131	6	11	0	15	0	9	1	3	1
		219	2	2	131	6	11	0	15	0	9	1	3	1

AUTHOR	YEAR	ID#	ESID#	ATTOWARD	TOTN	SETTINGE	EDLEVEXP	UNIVMAJE	OCCUPEXP	COMINSTR	ASSESTYP	TRTECHE	TOTHRSE	ESAVAIL
Threlkeld & DeJong	1983	219	3	2	131	6	11	0	15	0	9	1	3	2
		219	4	2	131	6	11	0	15	0	9	1	3	2
		219	5	2	131	6	11	0	15	0	17	1	3	3
		219	6	2	131	6	11	0	15	0	17	1	3	3
		219	7	2	131	6	11	0	15	0	17	1	.3	1
		219	8	2	131	6	11	0	15	0	17	1	.3	1
		219	9	2	131	6	11	0	15	0	17	1	.3	1
		219	10	2	131	6	11	0	15	0	17	1	3	1
		219	11	2	70	14	11	0	10	0	9	1	3	1
		219	12	2	60	14	11	0	10	0	9	1	3	1
		219	13	2	70	14	11	0	10	0	9	1	.3	1
		219	14	2	60	14	11	0	10	0	9	1	.3	1
		219	15	2	70	14	11	0	10	0	17	1	.3	3
		219	16	2	60	14	11	0	10	0	17	1	3	2
		219	17	2	70	14	11	0	10	0	17	1	3	1
		219	18	2	60	14	11	0	10	0	17	1	3	1
		219	19	2	70	14	11	0	10	0	17	1	3	1
		219	20	2	60	14	11	0	10	0	17	1	3	1
Giannotti & Doyle	1982	126	1	10	92	1	11	0	11	0	3	1	24 0	3
		126	2	10	92	1	11	0	11	0	3	1	24 0	3
McHale & Simeonsson	1980	189	1	16	28	2	2	0	0	0	17	2	2 5	3
		189	2	16	28	2	2	0	0	0	17	2	2 5	3
Appolloni et al.	1982	119	1	2		15	6	0	0	0	6	2	72 0	3
		119	3	7		15	6	0	0	0	6	2	72 0	3
		119	2	10		15	6	0	0	0	6	2	72.0	3
Wallston et al.	1972	150	6	6	16	10	7	0	0	0	9	2		3
		150	5	11	16	10	7	0	0	0	9	2		3
		150	2	12	16	10	7	0	0	0	9	2		3
		150	3	19	16	10	7	0	0	0	9	2		2
		150	4	20	16	10	7	0	0	0	9	2		2
LeUnes et al.	1975	262	1	3	179	4	8	1	0	0	9	2		3
		262	2	3	178	4	8	1	0	0	9	2		3
Chinsky & Rappaport	1970	86	1	6	90	4	8	9	0	0	15	2	30 0	3
		86	2	6	90	4	8	9	0	0	15	2	30 0	3
		86	3	6	90	4	8	9	0	0	9	2	30 0	1
		86	4	6	90	4	8	9	0	0	9	2	30 0	3
Shaver & Scheibe	1967	228	1	6	16	9	8	1	0	0	12	2	84 0	1
Spiegel et al.	1973	241	1	6	40	6	8	1	0	2	3	2	22 0	3
		241	2	6	40	6	8	1	0	2	3	2	22 0	3
Cole	1970	198	1	8	45	1	8	1	0	1	3	2	5 8	1
		198	2	8	45	1	8	1	0	5	3	2	5 8	1
Clore & Jeffery	1972	113	1	1	44	13	8	1	0	0	3	3	1.0	3
		113	2	1	44	13	8	1	0	0	3	3	1.0	3
		113	3	1	44	13	8	1	0	0	10	3	1.0	3
		113	4	1	44	13	8	1	0	0	10	3	1.0	3
		113	5	1	44	13	8	1	0	0	10	3	1 0	2
		113	6	1	44	13	8	1	0	0	10	3	1 0	2
Wilson	1971	258	4	9		1	8	4	0	0	9	3	2 5	3
		258	3	9		14	8	4	0	0	9	3	2 5	1
		258	2	9		1	8	4	0	1	3	3	2 5	1
		258	1	9		14	8	4	0	1	3	3	2 5	1

AUTHOR	YEAR	ID#	ESID#	ATTOWARD	TOTN	SETTINGE	EDLEVEXP	UNIVMAJE	OCCUPEXP	COMINSTR	ASSESTYP	TRTECHE	TOTHRSE	ESAVAIL
Sadlick & Penta	1975	234	1	14	80	1	8	6	0	0	9	3	3	3
		234	2	14	80	·1	8	6	0	0	9	3	8	3
		234	3	14	80	1	6	6	0	0	9	3	8	3
		234	4	14	80	1	8	6	0	0	9	3	8	3
Daniels	1976	147	1	2		0	0	0	0	1	3	4		1
DuHoux	1979	197	1	1	150	1	2	0	0	0	3	7	7	3
		197	3	1	50	1	2	0	0	0	3	7	8	3
		197	4	1	53	1	2	0	0	0	3	7	7	3
Sasso et al.	1985	249	1	16	37	1	2	0	0	0	3	7	11 5	3
Dickson	1974	244	3	1	80	15	3	0	0	0	3	7		1
Simpson et al.	1976	2	4	1	28	1	3	0	0	1	3	7	8	1
		2	5	1	23	1	3	0	0	1	3	7	8	1
Dickson	1974	244	4	1	80	15	3	0	0	1	3	7		1
Anastasiow et al.	1978	82	1	1	88	15	4	0	0	0	3	7		1
		82	2	1	88	15	4	0	0	0	9	7		1
		82	3	1	28	15	4	0	0	0	3	7		1
		82	4	1	28	15	4	0	0	0	9	7		1
		82	5	1	20	15	4	0	0	0	3	7		1
		82	6	1	20	15	4	0	0	0	9	7		3
		82	7	1	84	15	4	0	0	0	3	7		3
		82	8	1	84	15	4	0	0	0	9	7		1
		82	9	1		15	4	0	0	0	3	7		3
		82	10	1		15	4	0	0	0	9	7		3
		82	11	1	101	15	4	0	0	0	3	7		3
		82	12	1	101	15	4	0	0	0	9	7		3
Apolloni et al.	1982	119	4	2		15	6	0	0	0	6	7		3
		119	6	7		15	6	0	0	0	6	7		2
		119	5	10		15	6	0	0	0	6	7		2
Nemon	1980	267	1	9	55	6	7	0	15	0	3	7	6 0	2
		267	2	9	55	6	7	0	15	0	3	7	6 0	2
		267	3	9	55	6	7	0	15	0	3	7	6 0	3
		267	4	9	55	6	7	0	15	0	3	7	6 0	2
		267	5	9	38	6	7	0	15	0	3	7	6 0	2
		267	6	9	38	6	7	0	15	0	3	7	6 0	1
		267	7	9	38	6	7	0	15	0	3	7	6.0	3
		267	8	9	38	6	7	0	15	0	3	7	6 0	2
Austin	1980	68	1	1		1	8	16	0	4	3	7	2	3
Johannsen et al.	1964	110	1	6	72	4	8	6	0	2	3	7		3
		110	2	6	72	4	8	6	0	2	3	7		3
		110	3	6	62	4	8	6	0	2	3	7		3
		110	4	6	62	4	8	6	0	2	3	7		2
		110	5	6	59	4	8	6	0	2	3	7		3
		110	6	6	59	4	8	6	0	2	3	7		2
		110	7	6	59	4	8	6	0	2	3	7		3
		110	8	6	59	4	8	6	0	2	3	7		2
Lewis & Cleveland	1966	125	1	6	134	4	8	6	0	2	3	7		3
		125	2	6	134	4	8	6	0	2	3	7		1
Naor & Milgram	1980	177	5	20	53	13	8	2	0	0	17	7	22 5	3
Richardson & Guralnick	1978	164	1	1	18	0	10	0	13	0	17	7	12 0	3
		164	2	1	18	0	10	0	13	0	17	7	12 0	3
Brooks & Bransford	1971	111	1	1	30	15	10	0	15	0	9	7		1

AUTHOR	YEAR	ID#	ESID#	ATTOWARD	TOTN	SETTINGE	EDLEVEXP	UNIVMAJE	OCCUPEXP	COMINSTR	ASSESTYP	TRTECHE	TOTHRSE	ESAVAIL
Distefano & Pryer	1975	95	1	6	15	6	10	0	13	2	3	7	36 0	2
		95	2	6	15	6	10	0	13	2	3	7	36 0	2
Distefano & Pryer	1970	87	1	6	71	4	11	0	4	2	3	7	200 0	1
		87	2	6	71	4	11	0	4	2	3	7	200 0	2
		87	3	6	48	4	11	0	4	2	3	7	200 0	1
		87	4	6	22	4	11	0	4	2	3	7	200 0	1
Distefano & Pryer	1975	95	3	6	28	6	11	0	15	2	3	7	24 0	3
		95	4	6	28	6	11	0	15	2	3	7	24 0	2
Terrell	1981	254	1	8	27	1	3	0	0	0	3	8	4 5	3
Wyri[illegible]	1967	160	1	12	19	1	8	0	0	1	3	8	5	2
		160	2	12	19	1	8	0	0	1	3	8	5	2
[illegible]iels	1976	147	2	2		0	0	0	0	1	3	9		1
Handlers & Austin	1980	102	1	1	18	1	6	0	0	0	3	9	6 0	3
		102	2	1	18	1	6	0	0	1	3	9	6 0	1

NUMBER OF CASES READ = 167 NUMBER OF CASES LISTED = 167

APPENDIX G

CONVERTING D TO r_{pb}

Converting $\underline{D}$ to r_{pb}

Formula: $$r_{pb} = \frac{\underline{D}}{\sqrt{\underline{D}^2 + 4}}$$ (Rosenthal, 1984, p. 25)

Sample Values*

D	r_{pb}	r^2_{pb}
.00	.00	.00
.1	.05	.00
.2	.10	.01
.3	.15	.02
.4	.20	.04
.5	.24	.06
.6	.29	.08
.7	.33	.11
.8	.37	.14
.9	.41	.17
1.0	.45	.20
1.1	.48	.23
1.2	.51	.26
1.3	.54	.30
1.4	.57	.33
1.5	.60	.36
1.6	.62	.39
1.7	.65	.42
1.8	.67	.45
1.9	.69	.47
2.0	.71	.50
2.2	.74	.55
2.4	.77	.59
2.6	.79	.63
2.8	.81	.66
3.0	.83	.69
3.2	.85	.72
3.4	.86	.74
3.6	.87	.76
3.8	.88	.78
4.0	.89	.80

*Cohen (1977, p. 22) has provided a similar, but more detailed, table.