ABSTRACT
        Using meta-analysis techniques, the study sought to
identify, integrate, and synthesize the literature from 61 articles
which review the efficacy of various treatments for hyperactive
children. The major objectives were to determine if drugs can be used
effectively with hyperactive children, what child and intervention
characteristics covary with and/or influence intervention
effectiveness, and what are the priorities and focus of future
research efforts or those questions which need further investigation.
The meta-analysis approach used for the project required locating all
studies or a representative sample of all studies on a given topic,
converting the results of each study to a common metric, coding the
various characteristics of studies that might have affected the
results, and using relational and descriptive statistical techniques
to summarize study outcomes and examine the covariation of study
characteristics with outcomes. Overall results of the study suggested
that drugs are a moderately effective treatment for hyperactivity;
however, a significant number of variables were identified in the
analysis which require further research, including age of children
and IQ. Among appendixes are references to and analysis of review
articles, a coding sheet for efficacy of drug treatments for
hyperactivity, coding conventions for efficacy of drug treatments for
hyperactivity, and procedures for contacting authors for additional
information. (Author/SW)

A Meta-Analysis of Previous Research on the Treatment of Hyperactivity

FINAL REPORT FOR
GRANT # HEW/OE/NIE-G-80-0008

By:

Karl R. White, Beverly Myette,
Richard Baer, and Cie Taylor

A Meta-Analysis of Previous Research on

the Treatment of Hyperactivity

FINAL REPORT FOR

GRANT # HEW/OE/NIE-G-80-0008

by

Karl R. White, Beverly Myette,

Richard Baer, and Cie Taylor

TABLE OF CONTENTS

ED 224 218   EC 150 601

AUTHOR   White, Karl R.; And Others
TITLE   A Meta-Analysis of Previous Research on the Treatment
                of Hyperactivity. Final Report.
INSTITUTION   Utah State Univ., Logan. Exceptional Child Center.
SPONS AGENCY   National Inst. of Education (ED), Washington, DC.
PUB DATE   [82]
GRANT   NIE-G-80-0008
NOTE   120p.
PUB TYPE   Reports - Research/Technical (143) -- Information
                Analyses (070)

EDRS PRICE   MF01/PC05 Plus Postage.
DESCRIPTORS   *Drug Therapy; Exceptional Child Research;
                *Hyperactivity; *Intervention; Literature Reviews;
                *Research Methodology
IDENTIFIERS   *Meta Analysis

ABSTRACT
        Using meta-analysis techniques, the study sought to
identify, integrate, and synthesize the literature from 61 articles
which review the efficacy of various treatments for hyperactive
children. The major objectives were to determine if drugs can be used
effectively with hyperactive children, what child and intervention
characteristics covary with and/or influence intervention
effectiveness, and what are the priorities and focus of future
research efforts or those questions which need further investigation.
The meta-analysis approach used for the project required locating all
studies or a representative sample of all studies on a given topic,
converting the results of each study to a common metric, coding the
various characteristics of studies that might have affected the
results, and using relational and descriptive statistical techniques
to summarize study outcomes and examine the covariation of study
characteristics with outcomes. Overall results of the study suggested
that drugs are a moderately effective treatment for hyperactivity;
however, a significant number of variables were identified in the
analysis which require further research, including age of children
and IQ. Among appendixes are references to and analysis of review
articles, a coding sheet for efficacy of drug treatments for
hyperactivity, coding conventions for efficacy of drug treatments for
hyperactivity, and procedures for contacting authors for additional
information. (Author/SW)

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

> Science is built up with facts,
> as a house is with stones; but a
> collection of facts is no more a
> science than a heap of stones is
> a house.--Poincare in L'Hypothese

Most researchers in the social sciences would agree that their business includes the collection of facts. Furthermore, policy makers, administrators, and practitioners expect these facts to lead step by step to improve our understanding and contribute to improvements in practice. Unfortunately, examples of where facts from many individual research studies have been logically fitted together resulting in major changes and improvements in the social sciences are difficult to find.

In recent years, the integration of research has received substantial attention (Cooper, 1982; Feldman, 1971; Glass, 1976; Jackson, 1980). This attention has probably stemmed in part from the frustration of individual researchers and funding agencies about the lack of cumulative knowledge stemming from individual research studies. More and more people have devoted their attention to the importance of doing high quality, integrative reviews. In 1978, Greg Jackson pointed out that even though the reviewing and synthesizing of empirical research on a given topic is a fundamental activity in social science research, there is an absence of well-defined methods or procedures for conducting such reviews. Jackson concluded that these circumstances presented a major limitation to the accumulation of knowledge. Based on an analysis of a random sample of review articles and correspondence with journal editors and officials from government and private organizations responsible for reviewing and synthesizing research in the social sciences,

Jackson concluded that there were "important weaknesses in the currently prevailing methods of integrative reviews" (p. 37). The most frequently reported problems cited by Jackson included:

(1) the failure of most reviews to consider a complete or representative sample of the available evidence;

(2) the tendency for the conclusions of such reviews to be misleading;

(3) the failure of reviewers to systematically consider the possible relationships between characteristics of review studies and their findings; and

(4) the failure of reviewers to draw inferences for either theory, policy, or practice from the results of the studies.

Feldman (1971) observed that "the half-hearted commitment in reviewing and integrating completed research might account in part for the relatively unimpressive degree of cumulative knowledge in many fields of the behavioral sciences" (p. 86).

Emphasizing the importance of integrative research, Glass (1976) pointed out:

> A good review is the intellectual equivalent of original
> research . . . . we need more scholarly effort concen-
> trated on the problem of finding the knowledge that lies
> untapped in completed research studies . . . . The best
> minds are needed to integrate this staggering number of
> individual studies. This endeavor deserves higher
> priority now than adding a new experiment or survey to
> the pile. (p. 4)

An important area of research in which it has been particularly difficult to integrate and draw conclusions from the findings of previous research concerns the use of drugs to ameliorate the symptoms of hyperactive children. According to Trites (1979), hyperactivity is the most frequent reason children are referred to clinics and special school services. During

the last decade hundreds of research studies have investigated various treatments for ameliorating the symptoms of hyperactivity. Also referred to as "hyperkinesis," "minimal brain dysfunction," and "attention deficit disorder," hyperactivity has been called "one of the major childhood disorders of our time" (Ross & Ross, 1976).

In general, a child is considered to be hyperactive if he or she consistently exhibits an excessively high level of activity in situations where it is clearly inappropriate and is unable to inhibit his or her activity on command. Hyperactivity is often characterized by other psychological, learning, and behavioral problems, such as impulsivity, low self-esteem, poor academic performance, aggression, and distractibility. Although there are no firm statistical data, Grinspoon and Singer (1973) estimate 4 to 10% of the U.S. elementary school children are hyperactive and point out that educators claim the incidence to be as high as 15 to 20% (see also Bussey, 1967; Miller, Palkes, & Steward, 1973; Sprague, 1979; Stewart, 1975). A fairly conservative estimate of 5% would suggest that 1.5 to 2 million elementary-aged children in the U.S. are hyperactive. Clearly, hyperactivity is a problem of significant proportions.

Drugs are the most frequently used treatment for hyperactivity. In spite of hundreds of completed research studies using drugs, there is little agreement about their effectiveness, whether different types of children respond differently, or how much the symptoms of hyperactivity can be reduced. Discrepant results in the research literature regarding the treatment of hyperactivity point up the need for a methodologically sound review aimed at identifying the reasons for these discrepancies and providing more definitive information about the effectiveness of using drugs for the treatment of hyperactivity.

Objectives

The major aims of the project described in this report were to:

1. Determine if drugs can be used effectively with hyperactive children to decrease activity level, aggressiveness, and impulsivity; and improve cognitive performance, attention, academic achievement, and behavior.

2. Determine what child and intervention characteristics (e.g., age of child, nature of intervention, involvement of family) covary with and/or influence intervention effectiveness.

3. Prioritize and focus future research efforts by identifying those research questions which need further investigation and replication as opposed to those questions which have been sufficiently investigated, documented, and replicated.

Significance

The problems associated with hyperactivity are pervasive. As Barkley (1979) noted:

> Hyperactive children are often described as inattentive, overactive, and impulsive (Safer & Allen, 1976) . . . . Many demonstrate problems in noncompliance to adult commands (Barkley & Cunningham, 1979a; Campbell, 1973; Campbell, 1975) as well as aggressiveness towards others. Academic underachievement (Cantwell & Satterfield, 1978) and problems in classroom conduct are also evidenced . . . . As they develop into later childhood and adolescence, school failure, poor peer relationships, trouble with the law, and secondary reactive emotional problems are likely to occur (Ross & Ross, 1976). Although their social conduct problems may lessen as they enter young adulthood (Weiss, Hechtman, Perlman, Hopkins, & Wener, 1979), many are still more restless and inattentive than their normal peers and may develop problems with alcohol abuse (Blouin, Bornstein, & Trites, 1978). Hence, the disorder of hyperactivity is a lifelong difficulty associated with chronic academic problems and poor social relations. (p. 412)

Hundreds of research studies have investigated the effectiveness of treating hyperactivity with drugs, but the results are disturbingly discrepant and reviewers cannot agree. For example, in commenting on the effectiveness of drug treatments, Wender (1971) stated "minimal brain dysfunction is probably the single most common disorder seen by child psychiatrists . . . the correct treatment is often dramatically effective and is always cheap and readily accessible" (p. 1). In contrast, Adelman (1977) concluded that "the widespread use of such drugs for treatment of [hyperactivity] is premature. . . and perhaps quite dangerous" (p. 401), and Eisenberg and Conners (1971) concluded that "the continuing use of drugs despite the frequently negative outcomes of controlled studies indicates that behavioral disorders in children are placebo responsive" (p. 397).

Additional examples of the contradictions in the research literature concerning the effectiveness of using drugs for the treatment of hyperactivity could easily be cited, but would serve little purpose. The current state of confusion was summarized well by Freeman (1976) who stated:

> There is only one phrase for the state-of-the-art and practice in the field of minimal brain dysfunction, hyperactivity, and learning disability in children: a mess. There is no more polite term which would be realistic. The area is characterized by rarely challenged myths, ill-defined boundaries, and a strangely seductive attractiveness. These categories and their management, because of massive support from frustrated parents, professionals, government, and the drug and remedial education industries, constitute an epidemic of alarming proportions.

Kinsbourne and Swanson (1979) noted that "so much is known about hyperactivity that the information has become confusing. Before more work is done, some simplifying generalizations are needed" (p. 1). One explanation for the contradictory conclusions regarding the efficacy of using drugs for the treatment of hyperactivity is that previous reviews have failed to conduct the type

of integrative review which could advance knowledge in the field. Given the
large numbers of children affected by hyperactivity, the millions of dollars
spent yearly on the treatment of hyperactivity, and the contradictions in
previous research studies and reviews, a high quality review and summarization
of existing research is urgently needed. As will be documented in the
following section, existing reviews suffer from major methodological weaknesses
that may account in large part for their contradictory findings. The methods
used in the project described herein avoid most of these previous problems and,
in so doing, yield information which is more credible and comprehensive.

2.0 REVIEW OF PREVIOUS EFFORTS TO INTEGRATE THE RESEARCH

Jackson (1980) suggested that the quality of a review, and hence the confidence one should place in the conclusions of the review, can be judged by examining how well the review meets criteria in six areas.

1. / Selecting A Topic--Was the topic appropriately 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. Selecting 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 that topic reviewed, so that results of the review are generalizable to the "population" of research studies?

4. Data Collection--Were data collected for each study (so far as possible) on common dependent variables (study outcomes) and independent variables (study or subject characteristics such as age of students, type of intervention, methodological quality)? Were data collection procedures specifically described and defended on rational and empirical grounds?

5. Data Analysis--Was the relationship between dependent and independent variables examined in both univariate and multivariate dimensions? Were appropriate analysis techniques utilized?

6. Interpretation and Reporting--Were results reported in such a way that the reader can tell exactly what procedures and operational definitions were used? Are conclusions sufficiently supported by the data? Could the investigation be replicated based on the information reported?

These criteria can be used as a yardstick in judging the quality of previous reviews on the effectiveness of various treatments for hyperactivity.

Besides the hundreds and hundreds of primary research studies on hyperactivity, dozens of reviews have also been completed. A computer-assisted literature search of Psychological Abstracts, Dissertation Abstracts International, CEC Abstracts, Social Science Search, Index Medicus, and Education Resources Information Center, followed by a hand search and

information derived from bibliographies of already obtained articles,
identified 61 articles which have reviewed the efficacy of various treatments
for hyperactivity. To qualify as a "review," the article had to meet <u>at least</u>
<u>one</u> of the following criteria:

1. "Review" was used in the title; or

2. At least 35 primary research studies were considered to examine the
   effectiveness of a particular treatment for hyperactivity; or

3. At least 10 primary research articles were considered to examine the
   effectiveness of a particular treatment for hyperactivity, and the
   main purpose of the article was not to report on primary research
   conducted by the authors.

The review articles identified and some additional descriptive information on
each review are listed in Appendix 1.

Procedures for Examining the Quality of Previous Reviews

Each of the 61 articles was coded as to how well it met criteria in each
of the six areas previously described. Questions included on this coding sheet
are listed below with additional explanation as necessary.

1. <u>Did the review article explicitly and specifically state and delimit</u>
   <u>the topics to be included?</u> Any statement of the reviewer which
   delimited or defined the types of articles to be included in the
   review was counted as meeting this criterion. For example, this item
   would have been coded "yes" if the author said "tnis review will
   consider all articles which have examined the effectiveness of
   behavioral interventions on hyperactivity which meet minimum standards
   of methodological quality."

2. <u>Did the reviewer cite articles that are described as previous reviews</u>
   <u>on the same topic(s) or on similar topics?</u> This question was coded

"yes" only if the reviewer described such articles in the text and referred to them as previous reviews. It was not coded "yes" if reviews were listed in the references but were not referred to in the text as reviews.

3. Did reviewer critique previous reviews?

4. Did reviewer state how the present review would differ from or extend previous reviews?

5. Did reviewer state how studies to be included in the current review were located? To be answered "yes," the reviewer needed to explain the procedures in enough detail so that someone else could replicate the review using the same or nearly the same studies. For example, it was not sufficient to say that an ERIC search was done. The item would be coded "yes" if the authors said an ERIC search was done and stated the descriptors used in conducting the ERIC search, the years which were covered, and whether additional techniques were used to identify articles.

6. What is the actual number of experimental studies from which results were used to address the questions posed? To be counted, an article had to be cited in the text as supporting or refuting a particular point of view about the effectiveness of some intervention for hyperactivity. Articles which referred only to methodological issues or cited different instrumentation which previous research has used, were not counted in this total unless they were also cited pertaining to the effectiveness of a particular treatment.

7. What is the total number of references cited in the bibliography?

8. Did the reviewer describe with at least 250 words the major methodological difficulties or shortcomings of the primary or integrative research on the given topic(s)?

9. Did the reviewer suggest desirable foci or methods for future primary or integrative research on the topics?

10. What methods were used to consider findings of individual studies? This question asked about how outcomes from individual studies were used to draw conclusions about the effectiveness of a particular treatment for hyperactivity.

Each study considered was categorized in one of six areas: (a) Effect Size--some type of standardized metric that was comparable across all studies, e.g., $(\overline{X}_E - \overline{X}_C) \div SD_C$, (b) statistical significance-- favoring, nonstatistical significance, or against, (c) single subject designs which were visually analyzed, (d) differences-- each individual study was reported as having found differences or no differences, but no reference was made to whether it was a statistic- ally significant difference or not, (e) differences (groups)--a group of articles were cited as having found differences but did not make reference to whether these differences were statistically significant and did not consider the studies individually, and (f) percent improved--percent of subjects showing improvement following the treatment. If all studies in a particular review used the percent improved method, it would have been counted as Effect Size since it was a standard metric for all studies. However, this was never done.

In other words, a reviewer might cite 20 studies. For 10 of these studies, the percent of subjects improved in each study might have been reported; 5 of the studies might have been reported as having found statistically significant differences; 3 of the studies were reported to have found differences; and 2 of the studies were reported to have found no differences.

11. <u>How were findings or results of the reviewed studies summarized to</u>
<u>draw general conclusions?</u> This question addressed how, after
collecting data about the effectiveness for a particular treatment
from the individual studies, the reviewer summarized the results to
draw conclusions. Each study was coded in one of three categories:
(a) general direction of findings--no explicit summarization technique
was used, but the author did draw conclusions about what the studies
seemed to be showing, (b) percentage of studies finding "X"--in this
case, the author considered all studies in the review and said that
such and such a percentage favored this treatment and such and such a
percentage favored that treatment, (c) Effect Size for each study--
after quantitatively summarizing the results of each study on a common
metric, the author used this common metric to summarize what could be
concluded from the research.

12. <u>How was covariance of outcomes with subject or study characteristics</u>
<u>analyzed?</u> Each article was coded in one of four ways: (a) data
based multivariate--the reviewer empirically considered the covariance
of subject characteristics with study outcomes for most of the
studies, and simultaneous covariance with more than one subject or
study characteristic was considered, (b) data based univariate--same
as (a) except that the author only considered one study or subject
characteristic at a time, (c) logically considered for major subset--
coded if the author logically presented information for a substantial
subset of the data but did not do it in a systematic, data based
fashion--for example, if out of 30 articles considered in the review,
the author pointed out that in 8 of the articles, younger children
appeared to do better on treatment X than older children but made no
effort to report the age of children in the other 22 articles, it

would have been coded "c", (d) not considered for major subset--coded if none of the other three were applicable--for example, an author may have pointed out that one research article had found that drug therapy worked better for hyperactive children who had organic brain damage. Unless some effort was made to either logically or empirically consider the influence of organic brain damage on outcomes of treatment for other studies, it would have been coded in this category.

13. <u>What best describes the sample of articles considered in the review to draw conclusions about the topic?</u> This question was coded in one of four categories: (a) reasonable approximation of all research--by considering the year in which the review was completed, an estimation could be made of whether the sample considered in the review was a reasonable approximation of all research on the topic. If the research considered in the review was not a comprehensive sample, it would have been considered in category (b) (representative) if the author gave explicit criteria or procedures that were followed to assure some degree of representativeness. (c) convenience sample-- unless the author provided explanation that would have placed the article in one of the other three categories or considered a large enough sample of articles that convinced us that it was a reasonable approximation of all research, it was coded as a "convenience" sample. (d) purposive/exemplary research--if the author stated that they intentionally limited their study to only those articles that met predetermined standards for methodological quality, it was coded in this category.

14. <u>Did the reviewer draw conclusions based on the results of the reviewed studies about theory, policy, or practice?</u>

Quality of Previous Reviews

The results of the coding for 61 review articles is shown in Table 1 for each of the questions coded. As can be seen from the data presented in Table 1, the 61 review articles do not meet most of the criteria suggested for high quality integrative research. More than half of the review articles delimit or specifically state the topics to be included. However, even though almost half cite previous reviews, critiques of previous reviews and explanations of how the present review will differ from or extend previous reviews are seldom done. No one explained how studies to be included in the review were located. In spite of the fact that literally hundreds and hundreds of studies have experimentally examined the effectiveness of various intervention techniques for hyperactivity, the median number of studies included in reviews was only 12.5. Extensive discussion of topics which should be central to any high quality review (e.g., major methodological difficulties or shortcomings of previous research; or suggested procedures for future primary or integrative research on the topics being reviewed) were done infrequently. The procedures used for analyzing the results of individual studies or summarizing results across studies and considering how outcomes covaried with subjects or study characteristics was seldom done appropriately. Almost all of the reviews considered a convenience sample of articles. In spite of these serious shortcomings, reviewers did not hesitate to draw conclusions about theory, policy, and practice based on their review. Unfortunately, the basis for such conclusions and the procedures used to reach those conclusions would make it impossible for the reader to have confidence in the credibility of the conclusions.

Kerlinger (1977) has stated "the basic purpose of scientific research. . . is to understand and explain theory. Science then really has no other purpose

1 J

Table 1

Results of How Well Reviews of Hyperactivity Meet Criteria for High Quality Research

		YES	YES Briefly	NO	OTHER		
1.	Did review article explicitly and specifically state and delimit the topic(s) to be included?	52%	-	48%			
2.	Did reviewer cite article(s) that are described as previous reviews on the same topic(s) or on similar topics?	45%	-	55%	Mean = 2.75 Median = .5 Range = 0-33		
3.	Did reviewer critique previous reviews?	0%	3.3%	96.7%			
4.	Did reviewer state how this review would differ from or extend previous reviews?	6.7%	-	93.3%			
5.	Did reviewer state how studies to be included were located?	0%	-	100%			
6.	What is the actual number of experimental studies cited in the review as support for a particular contention?	Mean = 24.5; SD = 37.0; Range = 0 to 184; Median = 12.5					
7.	What is the total number of references cited in the bibliography?	Mean = 70.3; SD = 109.9; Range = 1 to 803; Median = 45.5					
8.	Considering the experimental studies cited in reviews, what methods were used to present results or findings of individual studies?	Effect Size	Statistical Significance	Single Subject	Differences (each study)	Differences (groups)	Percent Improved
		1%	11.3%	4.8%	42.6%	40.0%	1.3%
9.	How were findings or results of the reviewed studies summarized to draw general conclusions?	General Direction of Findings		Percentage of Studies Finding "X"		Effect Size for Each Study	
		91%		7%.		2%	
10.	How was covariance of outcomes with subject or study characteristics analyzed?	Data Based Multivariate		Data Based Univariate	Logically Considered for Major Subset	Not Considered for Major Subset	
		0%		0%	5.0%	95.0%	
11.	What best describes the sample of articles considered in the review to draw conclusions about the topic?	Reasonable Approx. of All Research		Representative	Convenience	Purposive/ Exemplary	
		5.0%		0%	93.3%	1.7%	
12	Did reviewer draw conclusions, based on the results of the reviewed studies, about theory, policy, or practice?	YES	YES Briefly	NO			
		30%	70%				

Note: All percentages indicate the percentage of the 61 reviews except where noted.

than . . . understanding and explanation" (p. 5). Kerlinger's statement, similar to common and acceptable definitions of science, represents the basic motivation and rationale for conducting educational research. Apparently, those who support and conduct educational research believe that the results of such research will improve our understanding and explanation of the educational process and thereby lead to the development, implementation, dissemination, and adoption of practices which will improve the quality of our educational system.

During the last 20 years, we have witnessed an explosion in the amount of educational research being conducted. Hundreds of thousands of research studies are completed every year. Journals are flooded with articles reporting research. Archive systems such as ERIC and <u>Dissertation Abstracts International</u> have made the results of hundreds of thousands of unpublished and fugitive manuscripts more readily available, and computer-assisted bibliographic searches have made comprehensive and complicated searches for existing literature much more feasible. Still, an all too familiar criticism of educational research is that, in most cases, it has failed to have significant impact on improving the quality of educational practices (Clifford, 1973; Kerlinger, 1977; Shaver, 1979; Strike, 1979).

The analysis of reviews on hyperactivity suggest that one major explanation for this lack of cumulative knowledge is the methods that have typically been used in attempting to integrate completed research. Although reviewers frequently criticize the methodological quality of primary research studies and call for more studies with improved methodology to be conducted, this analysis suggests that the same criticism can be made of reviewers. If reviews would attempt to meet criteria in the same basic areas which they recommend for primary research, then more progress would be made in addressing the problems of educational research and drawing conclusions which can be defended and can lead to improvements in practice and policy.

3. PROCEDURES

As demonstrated in the previous section, many of the reviews which have
examined the effectiveness of various treatments for hyperactivity suffer from
serious deficiencies. Although a number of approaches have been suggested for
reviewing literature, most of them also fail to meet the criteria for high
quality research. A brief summary of some of the more frequently used
techniques for reviewing research is presented below. Based on this summary,
it is suggested that the meta-analysis approach recommended by Glass and his
colleagues (Glass, 1978; Glass, 1980; Glass, McGaw, & Smith, 1981; Glass &
Smith, 1978, 1979; Smith & Glass, 1980) provides the best approach for
reviewing and drawing conclusions about previously completed research in areas
such as the treatment of hyperactivity.

Alternative Approaches for Integrating Completed Research

The most commonly used technique for reviewing research is a narrative
approach based on a group of easily accessible articles from fairly prominent
journals or other publications. Using 20 to 40 research articles, the reviewer
offers a verbal synopsis of each article, sometimes critiquing the methodology
and credibility of the conclusions, and often concluding that the existing
research is inconclusive--sometimes researchers reach one conclusion, sometimes
another. A call is then made for additional research using better techniques
and more precise methodology so that the truth of the matter can be
discovered.

In a slight variation of the narrative review approach, the reviewer
begins with a similar group of articles but eliminates all but a small number
because of supposed design or analysis flaws. The findings of the remaining
"acceptable" studies are presented as the truth of the matter. Unfortunately,
a judgment as to what constitutes a good article frequently differs from

reviewer to reviewer, and criteria for selecting "methodologically superior" articles are often overly restrictive and result in very small, and frequently nonrepresentative, samples of articles being considered. Moreover, as Smith and Glass (1980) have pointed out, "methodologically good" studies often report contradictory findings which can create considerable difficulty in evaluating what conclusions should be reached.

A more systematic approach to integrating the outcomes of primary research is what Light and Smith (1971) refer to as "the voting method". In the voting method, a relationship between a dependent and an independent variable is tallied as positively statistically significant, negatively statistically significant, or non-statistically significant. Studies are not usually weighted according to the size of the sample utilized in conducting the research. Since larger sample sizes lead to a greater probability of concluding that results are statistically significant, the voting method systematically discriminates against studies with small samples. Consequently, the true relationship may never be detected, and/or misleading conclusions may be drawn. Additionally, the voting method incorrectly implies that inferential statistics reveal the degree or importance of relationship, and that artifacts of measurement, bias, and the issues of experimental validity are controlled for adequately in all studies. As Glass (1977) pointed out, nine small sample studies may yield not-quite-significant results in one direction while a tenth large sample study yields statistically significant results in the opposite direction. The vote in this case is one for and nine against--a conclusion quite at odds with one's best instincts.

In an effort 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, original data from studies is frequently extremely difficult to obtain, and this procedure must disregard any researcher's data which is not obtainable. Glass (1977) reported that Wolins (1962) wrote to 37 authors asking for their data from studies 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.

The approach used in this project is referred to as meta-analysis and was first proposed by Glass (1976). Properly implemented, the "meta-analysis" approach meets all of the criteria for high quality integrative reviews proposed by Jackson (1978). Conducting a meta-analysis requires the location of either all studies or a representative sample of all studies on a given topic, converting the results of each study to a common metric, coding the various characteristics of studies that might have affected the results, and using relational and descriptive statistical techniques to summarize study outcomes and examine the covariation of study characteristics with outcomes. In his critique of previous efforts to integrate the findings of social science research, Jackson (1978) concluded that the "meta-analysis approach is a very important contribution to the social science methodology . . . . it will often prove to be quite valuable when applied and interpreted with care" (p. 47).

Since its introduction, the meta-analysis approach has been used to integrate research findings on a wide variety of topics including the relation of class size to achievement (Glass & Smith, 1979; Smith & Glass, 1980); the relation of socioeconomic status and academic achievement (White, 1982); the effectiveness of training and reinforcement on standardized test.results (Taylor & White, 1982); and neuropsychological assessment for brain damaged children (Davidson, 1978). More than 100 completed meta-analysis studies suggest that meta-analysis techniques are accepted as a useful methodology by substantial numbers of professionals.

It should be noted that some educational researchers have raised questions about the usefulness of meta-analysis (Educational Research Service, 1980; Eysenck, 1978; Gallo, 1978, Mansfield & Bussey, 1977; Shaver, 1979; Simpson, 1980). Some of these have questioned the results of a specific meta-analysis, while others have raised cautions or concerns about the meta-analysis approach per se. Most of these criticisms and cautions have been responded to in the literature (Glass, 1978, 1980; Glass et al., 1981; Glass & Smith, 1978). The most important point the concerns and questions have demonstrated is that meta-analysis, like all other research procedures, is not a fail-safe approach. If applied carelessly, many problems will occur. However, the meta-analysis approach, if properly implemented, has excellent potential as a tool for integrating research about the effectiveness of various treatments for hyperactivity.

Procedures for the Hyperactivity Meta-Analysis

The specific activities and procedures used in conducting the meta-analysis of the research on the treatment of hyperactivity are described below for each of the six areas suggested by Jackson (1980) for determining whether an integrative review is of high quality. Examples from previous meta-analyses are used to provide supporting evidence for the advantages of the meta-analysis approach and additional detail on the procedures to be used.

1. _Selecting and delimiting the topic._ The way in which the investigation of any research topic is defined determines in a large part the questions which will be answered. A topic which is too narrowly defined may only be able to answer trivial questions or may overlook important conclusions revealed by previous research. A topic which is too broadly defined may lead to the consideration of studies which are so divergent as to be uninteresting. Included in this integrative review were all those studies that have

empirically investigated the efficacy of drug treatments for hyperactivity.

Key terms in the preceding statement are defined below:

> Drug Treatment - Any treatment which attempts to ameliorate the symptoms of hyperactivity by administering a drug or chemical substance to the subject.

> Hyperactivity - Any pattern of behavior or activity level demonstrated or considered to be excessive. This definition is necessarily broad. Researchers employ a wide variety of criteria for defining hyperactivity. These range from accepting the opinion of a parent, teacher, or physician that a subject is hyperactive, to making systematic observations of subjects or using electro-mechanical devices to measure motor activity. A system for coding these various methods of defining hyperactivity was used as a part of the coding system described below.

Although some have argued that integrative reviews should only consider methodologically superior studies, our experience has been that this frequently fails to consider studies which can provide important information. The relation between study outcomes and methodological adequacy can be empirically assessed as a part of the meta-analysis. Then, if it is determined that the methodological adequacy of studies is confounding the results, appropriate adjustments can be made.

It should be noted that decisions concerning what to do about methodological inadequacies are different for a person conducting a primary research study than for a person integrating the results of previous studies. As Glass (1977) has noted, a researcher does not set out to perform a study deficient in some aspect of measurement or analysis, but it hardly follows that after a less than perfect study has been done, its findings should not be considered.

> 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 also possible that methodologically weak studies will

yield biased or misleading results. For example, as explained in the results

section, from the hyperactivity data considered in this project, drugs appeared

to be substantially more effective in reducing the symptoms of hyperactivity

when all studies were considered than when the analysis was limited to those

studies which used control groups, met minimum standards of internal validity,

and used objective measures to select hyperactive children for the study and to

measure outcomes.

As these results demonstrated, the best approach for determining whether

"weak" studies yield biased results is empirical. Each of the studies included

in this meta-analysis was classified according to well defined criteria which

are thought to impact on methodological quality (e.g., type of control group,

reliability or fakability of outcome measures, "blinding" of judges, duration

of intervention). Because "weaker" studies yielded different outcomes than

"stronger" studies, more credence was placed in the results of the "stronger"

studies. However, if the results had been similar, the inclusion of additional

studies would have allowed other important questions (e.g., the influence of

age of child or duration of treatment) to be examined more completely.

In summary, any study which investigated the efficacy of drug treatments for hyperactivity was considered in the meta-analysis. By considering all of these studies, questions of whether methodological adequacy covaries with results were examined empirically while at the same time substantially expanding the data base so that questions of how other study characteristics covary with study outcomes could be considered more adequately. This approach does not in any way condone future experiments that have weaknesses of design, analysis, or measurement.

2. Reviewing previous work on the same topic. As has already been noted, one part of this project was to examine previous reviews which have attempted to integrate the research literature on hyperactivity. In addition to demonstrating the need for a project such as this, the analysis of previous reviews often provides important information which can be the key to making sense out of the research literature. For example, in his meta-analysis of the research literature which investigated the relationship between socioeconomic status and academic achievement, White (1982) found that the unit of analysis used in computing the correlation between SES and achievement accounted for almost 40% of the variance in previously obtained correlation coefficients. As shown in Figure 1, those studies which had used individual students as a unit of analysis had a median correlation coefficient of .22, while those studies which had used group means in computing the correlation coefficients had a median correlation of .73. This one factor alone did much to clear up the confusion about how strongly SES is related to academic achievement. However, the unit of analysis used in computing the correlation coefficient was not an obvious factor to consider in conducting integrative review of the SES-Achievement correlation. Indeed, the "unit of analysis" variable was included, based on the suggestion of another reviewer even though the previous

reviewer had not presented enough evidence to substantiate the importance of the variable. If "unit of analysis" had not been considered, important questions regarding the relation between SES and achievement would not have been resolved.
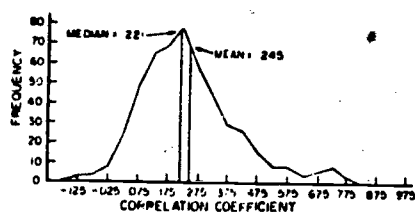
a) Using students as the unit of analysis
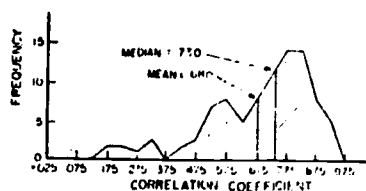
b) Using groups as the unit of analysis

Figure 1. Distribution of obtained correlation coefficients of the relationship between socio-economic status and academic achievement from 100 students.

The same principle applied to integrating the literature on the effectiveness of various treatments for hyperactivity. For example, suppose 100 research studies are considered, 50 of which implemented an intervention for hyperactivity and measured the outcome in a structured setting and 50 of which implemented the intervention and measured the outcome in an unstructured setting. Further suppose that those interventions in structured settings were very successful and all of the intervention programs in unstructured settings were completely unsuccessful. Finally, suppose that degree of structure in the setting where the intervention was implemented was not systematically considered in trying to organize and interpret previous research results. In this admittedly oversimplified and exaggerated example, the reviewer would probably conclude that the research concerning intervention

for hyperactivity is inconclusive--sometimes the intervention is effective, sometimes it is not.

Such an obviously wrong conclusion would occur because the correct concomitant variable was not considered. In spite of how obvious such an oversight appears when presented in this manner, this is exactly the type of mistake that almost all other reviewers of the hyperactivity literature have made. The best way to be sure that critical factors are included for consideration in the meta-analysis is to conduct a thorough review of what other people have suggested as potentially important factors and then to consider each of these factors to the degree possible in all of the primary research studies. Those factors which the analysis of previous reviews suggested are important for the hyperactivity research literature are included on the coding sheet used for the project which is included in Appendix 2.

Another reason for doing such an extensive analysis of previous reviews as the first step in conducting the meta-analysis is that it provided historical information (with specific references) about the most important issues that should be resolved by the meta-analysis. Conclusions of the meta-analysis regarding such issues can be referenced back to these contentions to either confirm or reject existing notions or hypotheses.

3. <u>Selecting studies for inclusion in the review</u>. The studies considered in the meta-analysis were identified by doing a computer search of the following indexes--<u>Psychological Abstracts</u>, <u>Dissertation Abstracts International</u>, <u>CEC Abstracts</u>, <u>Social Science Search</u>, <u>Science Search</u>, <u>Index Medicus</u>, and <u>Education Resources Information Center</u> (ERIC). Approximately 300 articles were identified as relevant for the meta-analysis. As each article identified through the computer search was read and coded for the meta-analysis, the bibliography of that article was examined to see if additional articles were

referenced which would be appropriate for the meta-analysis. These articles were then obtained and included in the meta-analysis. Not all of these articles could be included in the meta-analysis because of insufficient information being reported.

Included in the articles to be coded for the proposed study were both published and unpublished research reports. The importance of considering research from a variety of sources is clearly demonstrated by the information in Table 2 which was taken from Glass, Smith and Barton (1979). As can be seen, in nine different meta-analyses, the results often varied substantially

Table 2

Average Effect Sizes in Results of Studies
from Different Sources

Investigator	Topic	Source of Publication			
		Journal	Book	Thesis	Unpublished
Kavale (79)	Psycholinguistic training	.50		.30	.37
Hartley (77)	Computer-based instruction	.36		.28	.54
	Tutoring	.77		.40	1.05
Rosenthal (76)	Experimenter bias	1.02		.74	
Smith (78)	Sex bias in psychotherapy	.22		-.24	
Smith (80)	Effects of aesthetic education on basic skills	1.08		.48	.50
Carlbert (79)	Special class vs. regular class	-.09	-.01	-.16	-.14
	Resource room vs. regular class	.32	-.09		
Miller (79)	Drug therapy of psych disorders	.49	.56		
Hearold (79)	TV and anti-social behavior	.40	.14	.18	.23
Smith, Glass, Miller (80)	Psychotherapy	.87	.80	.66	1.96

Note: An Effect Size (ES) is defined as the standardized mean difference between two groups. Mathematically, ES $= (\overline{X}_E - \overline{X}_C) \div SD_C$.

from source to source. Note, for example, the results of Smith (1978) who considered the presence of sex bias in psychotherapy. Studies which had been published in journals showea that women were systematically discriminated against during psychotherapy, whereas studies reported in theses showed a bias in favor of women durirg psychotherapy.

Questions of Type I errors, bias, and quality of research reported in different sources are too important to be ignored when considering the questions of what previous research has really concluded about a specific topic. Research that is unpublished or reported in government project reports is usually reported regardless of the results, whereas some have suggested that research has a better chance of being published in journals or books if the results show statistically significant differences or agree with contemporary points of view. The primary objective of the meta-analysis of the hyperactivity literature was not to resolve questions about publication bias. However, the necessity of considering previously completed research from all sources is clearly evident if one is to draw valid conclusions about what can be concluded concerning the effectiveness of treatments for hyperactivity. For example, as reported in the results section, we found that the average effect size for studies supported in whole or part by commercial drug companies was .55 (n = 468), while the average effect size for those studies not supported by commercial drug companies was only .18 (n = 118). Clearly, any interpretation of the hyperactivity literature must at least consider who sponsored the research.

To assure that as many studies as possible were included in the meta-analysis, some articles were obtained from sources other than USU. The library system at USU was sufficient for obtaining the majority of the articles identified. However, additional efforts were sometimes necessary, including

utilization of the Interlibrary Loan System; requests to Dissertation Abstracts International; letters to authors requesting copies of unpublished reports; and requests to government archive systems. Although this was a time consuming and consequently an expensive undertaking, it was an important one if meaningful conclusions were to be made about what could be concluded from previous research.

4. Data collection. The key to a successful meta-analysis is the appropriate development of the coding/classification system. The basic concept behind the meta-analysis approach is to quantify the outcome of all research studies on a metric that can be used for all studies in the sample, and then to code the various study characteristics which may covary with outcomes in order to determine whether or not studies with certain characteristics consistently result in one outcome while another type of study produces another outcome. The classification system used to code the studies is the basic data collection instrument. This classification system must be comprehensive enough to include those factors which are contributing to the variance among different studies, but cannot be so complex that coding studies becomes an overly burdensome task.

The development of the coding/classification system is an extremely important task in conducting a meta-analysis. Appendix 2 shows an example of the coding/ classification system used in coding studies on drug treatment of hyperactivity. As can be seen, this classification system includes information about dozens of factors that other reviewers, researchers, and the project team thought might be important in explaining the results of research investigating the efficacy of drugs for the treatment of hyperactivity. As can be seen in Appendix 2, the coding sheet was divided into eight sections:

1. Identifying information on the article being coded,

2. Description of the research sample,

3. Description of how subjects were classified as hyperactive,

4. Description of the type of treatment given to subjects,

5. Description of the research design,

6. Description of potential threats to study validity,

7. Description of the research outcome and conclusions, and

8. Description of the specific drug treatment employed.

Another important step in the meta-analysis is the development of the conventions by which decisions are to be made in classifying each of the variables in the coding/classification system. In other words, it is not enough to say that factors related to the internal validity of a study will be coded. One must also specify the basic decision rules which will be used in determining, for example, whether selection bias is a threat to the internal validity of a study. For some factors, these decision rules are obvious and need not be specified in great detail. For others, it is critical that the decision rules be explicitly specified so that replication could occur. The basic conventions used for this project are included in Appendix 3.

Another important part of the procedures for coding individual studies was the development of examples which clarified the basic conventions. After coding was initiated, many situations were encountered which were not covered by the basic conventions. As coding proceeded, examples of how specific conventions were interpreted were noted in an example notebook. In this way, the rationale used for past decisions was documented and served to keep future decisions consistent. The following examples of how "instrumentation" threats to internal validity were coded will clarify this procedure.

Code #1 (minor threat): Dependent variable for project consisted of continuous 15-minute segment of observation data for each child gathered only once during 6 weeks of intervention. Threat of sampling error.

Code #1 (minor threat): Dependent variable was observation data collected at different times by different judges for experimental and control groups. Ratings required moderately high inference judgments with vaguely stated criteria. Raters were "blind".

Code #2 (moderate threat): Dependent variable consisted of pre/post opinion ratings as to degree of improvement. Staff members knew some children were receiving treatment but did not know which ones.

Code #2 (moderate threat): Dependent variable was observation data with fairly well specified low inference rating system. However, raters were not blind as to who was receiving treatment and had some cause to be biased.

Code #3 (major threat): Outcome measured on pre/post design with dependent variable being opinion of staff as to degree of improvement with no criteria. Staff knew subjects were being treated for hyperactivity.

The most important piece of information to be coded for each article was the outcome of the research. The basic outcome measure for each study examined in the meta-analysis was an effect size (ES) defined as $\overline{X}_E - \overline{X}_C \div SD_C$. In other words, the ES or outcome for each study was defined as the difference between the means of the treatment (i.e., "experimental") and the control subjects on a given dependent variable divided by the standard deviation of the control group on that variable. Thus, an ES of +1.0 as indicated in Figure 2 would indicate that the average person in the treated group is one standard deviation above the mean of the control group on that particular measure. This measure of ES avoids many of the problems encountered in using statistical significance as a measure of the outcome, since it is independent of the size of sample and has similar meaning across all studies and dependent variables. Quantifying the results of each individual study into an ES which has similar meaning across all studies allowed comparisons and cross tabulations with other study characteristics. Thus, questions about whether certain types of studies are "more effective" could be answered.
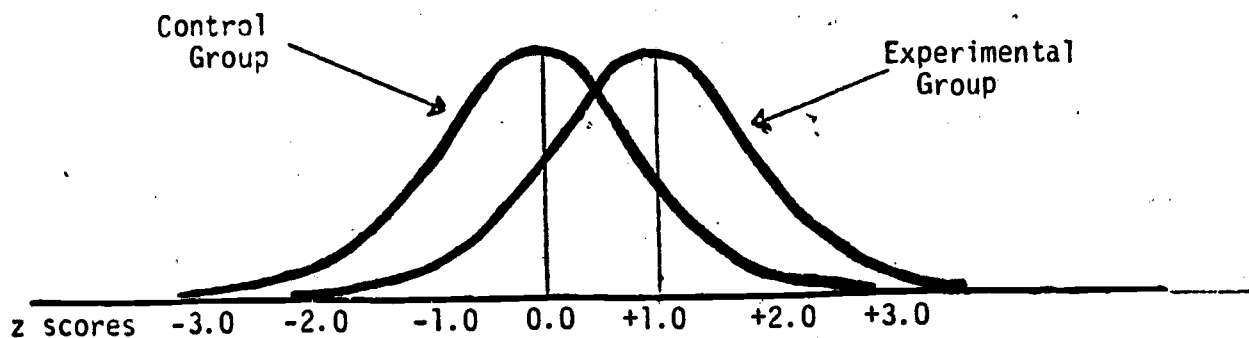
Figure 2. Graphic representation of distribution for experimental and control groups with an Effect Size (ES) of +1.0.

An obvious problem in this approach is that many reports of research do not provide sufficient information to calculate in ES using the $\overline{X}_E - \overline{X}_C \div SD_C$ definition given above. Where means and standard deviations were not reported, it was frequently possible to obtain estimates of the ES using information from reported statistics (e.g., $\underline{F}$ ratios, t values, r_{xy}, etc.) For example, if a study failed to report the standard deviation of the control group for a particular dependent variable, the square root of the within cell mean square (MS) from a one-way analysis of variance could be used as an estimate of the standard deviation of the control group for that dependent variable (Glass et al., 1981). Another example--suppose a study reported the obtained $\underline{t}$ value for a particular comparison between two groups but did not report means and standard deviations for the groups. Assuming that the variances between the two groups are equal (a standard assumption of the $\underline{t}$ test), the equation for the obtained $\underline{t}$ ratio was solved to yield an estimated effect size as follows:

$$t = \frac{\bar{x}_e - \bar{x}_c}{\sqrt{\frac{s^2(n_1-1) + s_2^2(n_2-1)}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

if $SD_1^2 = SD_2^2$

$$t = \frac{\bar{x}_e - \bar{x}_c}{\sqrt{\frac{SD^2[(n_1-1) + (n_2-1)]}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$t = \frac{\bar{x}_e - \bar{x}_c}{\left(\sqrt{SD^2}\right)\left(\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)}$$

$$t\left(\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) = \frac{\bar{x}_e - \bar{x}_c}{SD} = ES$$

Other equations for estimating effect sizes (ES) from analysis of variance summary tables (either one way or factorial designs), reported $\underline{F}$ ratios, probability levels, analysis of covariance results, matched pairs $\underline{t}$ tests, and other summary statistics were used as outlined by McGaw and White (1981).

Many times, enough information was not reported or the information which was reported was reported in such a way that it was impossible to estimate an effect size (e.g., a probability level from a chi-square test, or an $\underline{F}$ ratio with no supporting ANOVA summary table from a repeated measures ANOVA design). In these cases, authors were contacted to obtain information about means and standard deviations to calculate the effect sizes. The procedures used for writing to authors for additional information are shown in Appendix 4.

5. Data analysis. There is very little which is complex or statistically unique about the data analysis of the information produced from the coding of studies in a meta-analysis. As Glass et al. (1979) have noted:

> The approach to research integration referred to as "meta-analysis" is nothing more than an attitude of data analysis applied to quantitative summaries of individual experiments. By recording the properties of studies and their findings in quantitative terms, the meta-analysis of research invites one who would integrate numerous and diverse findings to apply the full power of statistical methods to the task. Thus, it is not a technique. Rather it is a perspective that uses many techniques of measurement and statistical analysis.

The most useful data-analysis techniques in meta-analysis studies are frequently the most simple. After coding all of the study characteristics and outcomes of the studies, frequencies and mean effect sizes were computed for each variable. Next, cross-tabulations with the effect size were computed for each of the relevant study characteristics which have been coded. For example, an average effect of .85 for 100 effect sizes of methylphenidate and an average effect size of .25 for 150 effect sizes of dextroamphetamine would indicate that using methylphenidate results in approximately six tenths of a standard deviation better gain across all dependent variables than dextroamphetamine. This finding could be broken down further to see if the advantage of methylphenidate is constant for subjects at all age levels, e.g., 4 to 6 years, 7 to 9 years, and 10 to 12 years. The results could be broken down still further in a three-way tabulation to look at the general methodological quality of the study, as it interacts with these other two variables. In this manner, various combinations of the study characteristics were examined to determine how outcomes covary with the characteristics.

     6. <u>Interpreting and reporting the results</u>. Scientists generally have given much import to the interpretation and reporting of their research. Reports of research are supposed to be thorough enough to allow other people to judge the validity of the findings and interpretations, and to replicate the research should they so desire. It is generally believed that reports of primary research ought to indicate at least the sampling procedures, essential design characteristics, the data collection techniques, the methods of analysis, and the findings. These same standards ought to be applied to the reporting of integrative reviews, but frequently are not.

     The systematic procedures for collecting and analyzing data in the meta-analysis allows the results to be reported in enough detail so that others

can judge the plausibility and validity of the findings. The explicit and systematic manner in which the meta-analysis was conducted also helps to ensure that interpretations do not overstep the quality of the data which have been collected. In many reviews, the degree to which the conclusions are supported by the data is difficult to determine since the procedures and techniques used to collect, analyze, and interpret the data exist mostly in the mind of the reviewer rather than being explicitly stated as procedures.

4. RESULTS AND DISCUSSION

Characteristics of Data Set

Seven hundred and fifteen effect sizes (ES) were obtained from the meta-analysis coding. As explained previously, an effect size was defined as the mean of the experimental group minus the mean of the control group divided by the standard deviation of the control group as shown in Formula 1:

$$ES = \overline{X}_E - \overline{X}_C \div SD_C \qquad (1)$$

This definition of effect size allowed results from one study to be compared with results of another study, or results from one outcome measure to be compared with results of other outcome measures, without being confused by artifacts of statistical significance or scaling.

The distribution of the magnitude of these effect sizes is shown in Figure 3. As can be seen, when all 715 effect sizes are considered without regard to other subject or study characteristics, the mean effect size resulting from the treatment of hyperactivity with drugs is .44 with a median of .40. In other words, children who received drugs for the treatment of hyperactivity are, on the average, .4 of a standard deviation better off than children who are not treated with drugs. This effect size of .40 indicates that a child who has received drugs would score at the 66th percentile of a group of children who did not receive drugs.

Before examining the interactions of various other subject characteristics with outcome, some of the characteristics of the data set from which these effect sizes were obtained will be described. Overall, the quality of research which has examined the effectiveness of treating hyperactivity with drugs is better than many have supposed. Of the 715 effect sizes, 567 or 73% came from
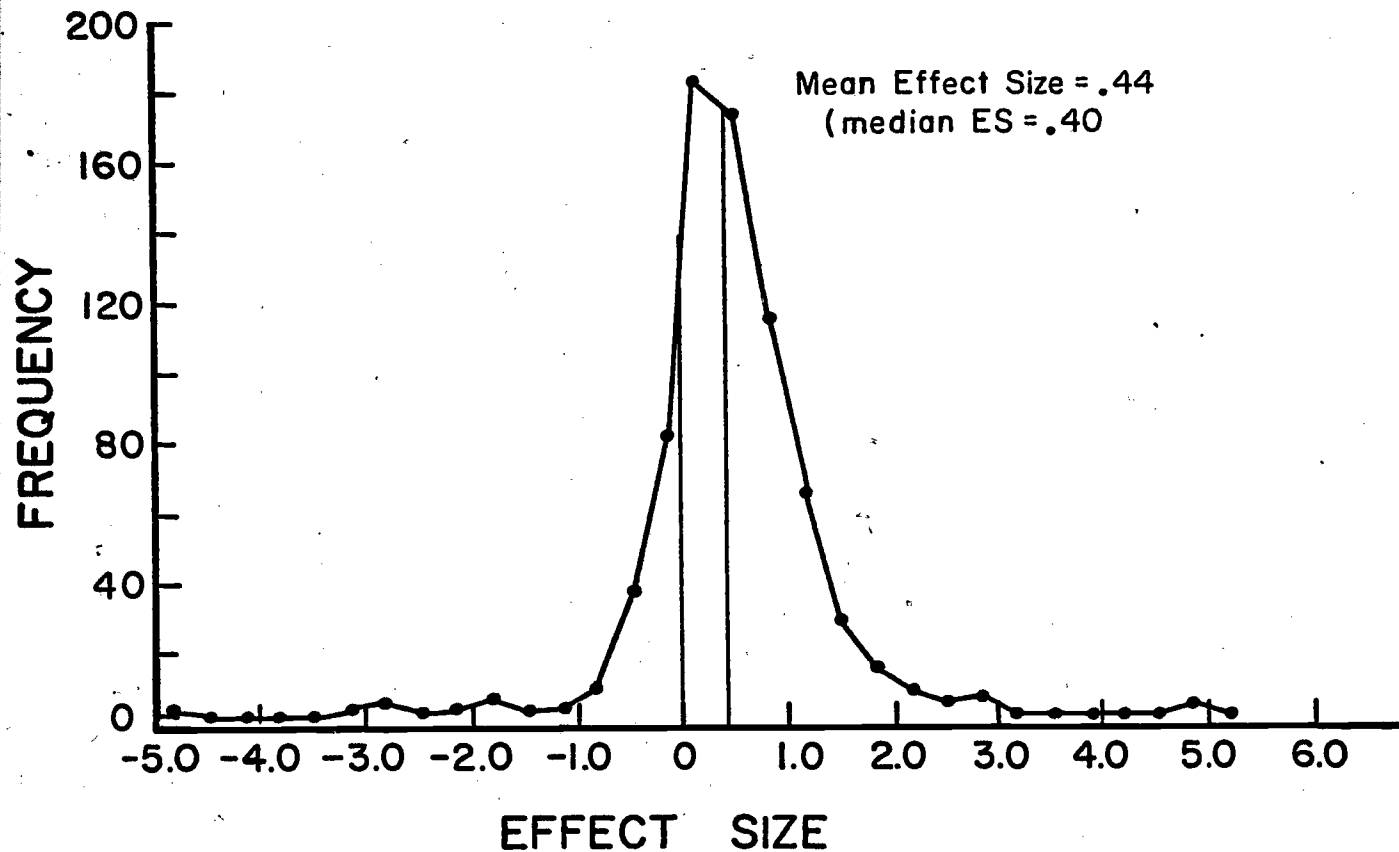
Figure 3. Distribution of Effect Sizes ($\overline{X}_E - \overline{X}_C \div SD_C$) for studies which have examined the effectiveness of treating hyperactivity with drugs (n = 715).

42

studies with cross-over designs or where subjects were randomly assigned to experimental and control groups. Fifty-one percent of the effect sizes were obtained from studies which employed a placebo in the control group, and 78% of those placebos were judged to be high quality placebos. Most effect sizes (approximately 80%) were obtained from studies which took some measures to assure that subjects, treatment implementors, and data collectors were blind as to which group of children was receiving the treatment. The quality of these blinding procedures was often quite good, although improvements would have been desirable (e.g., 46% of the effect sizes had good blinding for subjects, 31% had good blinding for the treatment implementor, and 31% had good blinding for the data gatherer). Forty-two percent of the effect sizes were obtained from studies which had excellent or good ratings of methodological quality. Thirty-three percent of the effect sizes came from studies with fair ratings of methodological quality, and 25% came from studies with either poor or very poor ratings of methodological quality. Although these ratings do indicate that there is need for further improvement and rigor in the research which examines the treatment of hyperactivity, the ratings also indicate that a good many high quality studies are available upon which to base conclusions.

Most of the studies considered in this meta-analysis were conducted in the 1970s. A few studies occurred as early as 1945 with substantial increase in research activity occurring in the 1970s. The median year in which effect sizes from this meta-analysis came was 1974. As shown in Table 3, most studies came from "medical" instead of "educational" journals. However, differences in average effect sizes between these two categories were trivial. The number of subjects included in experimental groups ranged from 2 to 217 with a median sample size for experimental groups being 29. In summary, the conclusions which follow regarding the treatment of hyperactivity with drugs are based on a

Table 3

Frequency and Average Magnitude of Effect Sizes
in Different Types of Journals

	ES	n	SEM
Educational journals	.43	(207)	.05
Medical journals	.45	(508)	.02

large number of studies in which literally thousands of children were examined in experimental treatments. The studies cut across a broad range of years and considered many different outcomes which might be affected by treating hyperactivity with drugs.

Potentially Confounding Variables

The overall statement of the effectiveness of using drugs for the treatment of hyperactivity (i.e., a median effect size of .40) indicates that drugs do have a moderate but positive effect on ameliorating the symptoms of hyperactivity. However, the real power in the meta-analysis approach is that it allows the examination of various factors which may interact with this general statement of effectiveness. Most interesting are those study and subject characteristics which covary with effect size. For example, are certains of drugs more effective? Or, do drugs work better with younger rather than older children? Or, do drugs have greater impact on certain types of dependent measures? However, before examining these questions, it is important to consider whether there are variables which may be confounding the relationship between study characteristics and effect size and thus, mislead researchers.

<u>Support from commercial companies</u>. For approximately 65% of the effect

sizes obtained, some type of commercial support from drug companies was

provided to the study. Table 4 shows the average effect size for studies which

received various levels of financial support from commercial companies. Those

studies which received complete financial support yielded dramatically higher

average effect sizes (.79) than those studies receiving no support (.18).

Table 4 breaks these results down further by quality of research design. As

can be seen in the panel to the right of that table, the trend for high effect

sizes being associated with support from commercial drug companies holds true

for both high quality and low quality research, although the differences are

more dramatic for research which was of high quality. Data in Table 4 suggest

that one must be cautious in interpreting the results of research which is

supported by commercial drug companies. Although certainly not definitive

evidence, these data do suggest that support for research from commercial

companies may bias the results of the research. Notice particularly the fact

that studies which received no commercial support and where the research was of

high quality had an average effect size very close to zero.

Some explanation in interpreting the data displayed in Table 4 will be

helpful in interpreting the remaining data displays because all tables have

been constructed using a similar format. $\overline{ES}$ indicates the mean effect size

for a particular cell. N indicates the number of effect sizes on which that

mean is based. SEM is the standard error of the mean for $\overline{ES}$. This was

obtained by dividing the standard deviation for the distribution of effect

sizes in that particular cell by the square root of n.

The standard error of the mean helps one determine if apparent differences

are real or only the result of sampling fluctuation. For example, in Table 4,

the differences in average effect size for those studies receiving complete

Table 4

Average Effect Size for Studies Broken Down by
Amount of Financial Support and Quality
of Research Design

Financial Support from Drug Company	Overall ES			Quality of Research Design					High Quality (1,2)			Low Quality (3,4,5)		
	$\overline{ES}$	n	SEM	1	2	3	4	5	$\overline{ES}$	n	SEM	$\overline{ES}$	n	SEM
Complete	.79	(17)	.16			.51 (12)		1.47 (5)	-	-	-	.79	17	.16
Partial	.54	(451)	.04	.43 (8)	.51 (177)	.58 (152)	.58 (40)	.55 (74)	.51	185	.05	.57	266	.04
None	.18	(118)	.10	.56 (20)	-.23 (52)	.33 (34)	1.11 (6)	.64 (6)	-.01	72	.08	.47	46	.10

39

40

4

support to those studies receiving no support from commercial companies runs from .79 to .54 to .18 when all effect sizes are considered. The larger the number of effect sizes used in calculating an average effect size, the smaller the standard error of the mean will be, all other things being equal. A good rule of thumb is that if a "confidence interval" of 2 standard errors of the mean around each $\overline{ES}$ do not overlap each other, then there is a good chance that the differences are due to sampling fluctuation. For example, in Table 8, the average effect size for studies receiving partial support from commercial companies was .54 with a standard error of the mean equaling .04. This indicates that the best estimate of the true mean for those studies receiving partial support from drug companies is somewhere between .46 and .62 (.54 ± .08). The best estimate for the average effect size for those studies receiving no commercial support is between -.02 and .38 (.18 ± .20). Since these "confidence intervals" do not overlap, one can be reasonably confident that the differences in average effect size between those companies receiving partial support and no support are not due to sampling fluctuation. .

Many of the tables reported in the remainder of this section break down overall effect sizes by quality of research design. This has been done because quality of research for many variables was found to confound the interpretation of overall effect sizes. For each study considered, quality of research design was coded from 1 (high quality research) to 5 (low quality research). In addition to indicating the average effect size for each rating of research designs, the right-hand panel in Table 4 categorizes the studies into those studies that received either excellent or good ratings (1 or 2) as opposed to those studies which received moderate, poor, or very poor ratings (3, 4, or 5). When no other indication is given, numbers in parentheses indicate the number

of effect sizes upon which an estimate was based. The numbers in bold-faced type indicate the average effect size as is done in the middle panel in Table 4. Finally, any estimates of average effect size which were not based on more than five effect sizes have generally been eliminated from these tables. A few exceptions have been made where not including an estimate based on a low number of effect size would have been misleading.

Procedures for classifying children. As shown in Table 5, a significant problem affecting many of the studies was the frequently inadequate procedures used in assuring that children selected for hyperactivity research were truly hyperactive. Overall, approximately half of the effect sizes were obtained from studies where classification procedures were considered poor (e.g., no objective measures such as systematic observations or well-defined ratings were used to classify the children as hyperactive). The problem with not using better procedures to assure that children selected for such research are truly hyperactive is emphasized by the fact that the average effect size for those children where the procedures for classification were fair to good was only .34 (n = 321), and the average effect size for those studies where the classification procedures were poor was .56 (n = 278).

Table 5

Average Effect Size for Classifying Children as
Hyperactive by the Quality of Classification

	Quality of Classification					
	Good or Fair			Poor		
	ES	n	SEM	ES	n	SEM
General Hyperactivity	.34	(321)	.04	.60	(259)	.04
Activity Level, Attention, or Aggression				.09	(19)	.16

As can be seen in Table 5, the most frequently used basis for classifying the children as hyperactive for inclusion in the study was a general measure of hyperactivity which was used in 81% of all effect sizes obtained. More specific measures such as activity level or attention which are supposedly some of the defining characteristics of hyperactivity were used very infrequently. The data in Table 5 indicate that some of the apparent improvement in children treated with drugs may be due to the fact that many of the children included in such research are not really hyperactive in the first place. General hyperactivity level was the most frequently used basis for classifying children to be included in the study, and a large portion of those effect sizes came from classification procedures that were quite poor. More specific detail on how these ratings were made is shown on the coding sheet and the conventions in Appendices 2 and 3.

The possibility that many subjects included in the research may not have been truly hyperactive is underscored by information obtained from the ratings of the severity of hyperactivity. Those subjects who exhibited mild symptoms of hyperactivity had an average effect size of .57 (n = 356 effect sizes) whereas those subjects who exhibited extreme cases of hyperactivity had an average effect size of .35 (n = 59 effect sizes). Again, the much higher average effect size for milder cases suggests that even though drugs do have a positive impact, the true magnitude of the impact may be overestimated because some of the children included in such treatments may not be truly hyperactive.

Quality of research. Another major area of concern is the quality of the research upon which effect sizes are based. Each study considered in the meta-analysis was rated on various factors which might have threatened the internal validity of the study. These factors generally followed the Campbell-Stanley (1966) paradigm for internal and, to some degree, external

validity of research. As can be seen in Table 6, studies with more serious threats generally resulted in higher average effect sizes. In addition, Table 6 shows that instrumentation and mortality were the most frequently occurring threats to the validity of the study.

Table 6

Average Effect Size for Studies with Various Threats
to Internal Validity

	No Threat	Minor Problems	Moderate Problems	Serious Problems
Maturation	.36 (511)	.62 (136)	.66 (66)	
History	.40 (565)	.53 (76)	.58 (72)	
Testing	.52 (517)	.25 (139)	-.04 (44)	.60 (13)
Instrumentation	.37 (338)	.60 (206)	.39 (170)	
Regression	.41 (485)	.49 (196)	.58 (34)	
Selection	.44 (546)	.46 (112)	.40 (56)	
Mortality	.38 (443)	.58 (230)	.28 (42)	
Novelty	.44 (597)	.43 (116)	.07 (2)	
Experimenter Effect	.47 (501)	.33 (174)	.51 (39)	

Note: See coding sheet and conventions for more complete explanation of how each threat to internal validity was rated.

The ratings of individual threats to the validity of a study were used in determining a general index of validity for the study (procedures for doing this are contained in Appendix 3). Table 7 shows the average effect size for those studies which had good ratings (1 or 2 on a 5-point scale with 1 being high and 5 being low) as opposed to moderate or poor ratings. As can be seen, the average effect size for studies with high quality research designs is somewhat lower than those studies with moderate or poor research designs. However, as indicated in the schematic at the bottom of that table, a 95% confidence interval (2 standard errors of the mean) for each of these estimates is slightly overlapped. Although there is a trend for better research to show lower results, one must be cautious in over-interpreting these results.

Table 7

Average Effect Size for Studies with Good Research
Designs Versus Those with Moderate or Poor
Research Designs

Quality of Design	ES	n	SEM
Good (1, 2)	.36	(298)	.04
Moderate or Poor (3, 4, 5)	.50	(417)	.04

```
 +   +   +   +   +   +   +   +   +   +   +   +   +
.20 .24 .28 .32 .36 .40 .44 .48 .52 .56 .60 .64 .68
```
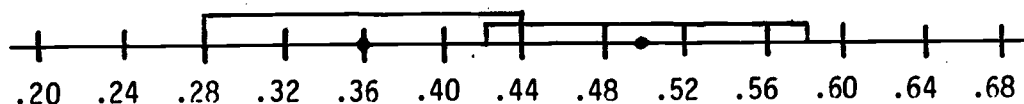
Instruments used in collecting outcome data. Also related to the quality of the research was the type of instrument used to collect data on the outcome measure. Shown in Table 8 are the average effect sizes for those studies in which the outcome data were based on someone's opinion as opposed to some sort of systematic rating procedure. Studies which obtained data using an

Table 8

Average Effect Size for Studies Which Used Different
Instruments to Select Children for Study

Type of Instrument	ES	n ~	SEM
Opinion	.51	(536)	.03
Rating	.18	(101)	.06

opinion have substantially higher average effect sizes than studies which
obtained outcome data via a rating. These data suggest strongly that the
apparent effectiveness of using drugs for the treatment of hyperactivity may be
confounded by the rigor with which data is collected regarding the outcome.

Placebo effect. A related but different concern is shown in Table 9.
These data show the average effect size for those studies which used a "no
treatment" control group as opposed to those studies which used a "placebo"
control group. The differences in average effect sizes between these two
groups indicates that substantially lower effect sizes are obtained when a
placebo was used as opposed to when a no-treatment control group was used. The
lower effect sizes obtained with placebos indicate that drug treatment of
hyperactivity is to some degree placebo-responsive, as has been suggested by
some previous reviewers. The best estimate of the magnitude of the placebo
effect is approximately 1/4 standard deviation (the difference between the
average effect size for no-treatment groups and placebo groups). Although the
magnitude of this placebo effect is substantial, it is not enough to account
for the apparent effectiveness of utilizing drugs for the treatment of
hyperactivity.

Table 9

Placebo Effect of Using Drugs for Treatment
of Hyperactivity

"Treatment" Used for Control Group	ES	n	SEM
No Treatment	.71	(131)	.06
Placebo	.48	(356)	.03

Degree of treatment implementation. Another methodological

consideration in interpreting the results from drug research is the degree to

which the treatment was actually implemented. In some research studies, extra

precautions are taken to make sure that the subject actually receives the drug

in the appropriate dosage and at appropriate times. In other studies, no such

precautions are taken. In those studies considered in the meta-analysis,

ratings were made of the degree to which one could be confident that treatment

implementation actually occurred. Table 10 shows the average effect size for

those studies where there was complete implementation or only minor problems

with implementation as opposed to those studies where there were major problems

with implementation. As can be seen, there is a substantial difference in

average effect size obtained. In what may seem to be counter-intuitive, given

the results reported above, the average effect size where there were major

problems with implementation was substantially higher than where there were

very few problems with treatment implementation. This may have occurred

because those studies which had major problems with treatment implementation

also had many other problems in terms of research quality and outcome measures.

Table 10

Average Effect Size for Studies with
Different Degrees of Treatment Implementation

Degree of Treatment Implementation	ES	n	SEM
Complete Implementation or Only Minor Problems	.12	(212)	.05
Major Problems with Treatment Implementation	.55	(337)	.03

In other words, whether or not checks were made on treatment implementation may be indicative of the general quality of the research being conducted and lower quality research generally found higher effect sizes.

Reliability of outcome measures. A final note on methodological quality and the ways in which it may interact with estimates of treatment impact is shown in Table 11. Here, the average effect size is shown for studies which used highly reliable outcome measures as opposed to those which used less reliable outcome measures. As can be seen from the number of effect sizes considered in each case, there were many studies for which no estimate about the reliability of the instrument could be made. However, in those cases where estimates could be made, studies which had highly reliable instruments tended · to show lower effect sizes than those studies which had unreliable instruments.

Summary about potentially confounding variables. The data presented in Tables 4 through 11 are important because they must be used in interpreting the results of the following section which considers the effectiveness of using drugs for the treatment of hyperactivity. These data indicate that there are a number of factors that may confound the results reported below. In doing the

Table 11

Average Effect Size for Instruments with
Different Levels of Reliability

Reliability of Outcome Instrument	ES	n	SEM
1.0 - 8.0	.24	(41)	.07
.60 - .79	.48	(103)	.05
0 - .59	.83	(10)	.17

analyses reported in the next section, these potentially confounding factors
have been accounted for wherever possible. However, the fact that such
confounds are present in the data and that substantially different magnitudes
of effect sizes are associated with different levels of these potentially
confounding variables makes one more cautious about the results reported
below.

Effectiveness of Drug Treatment for Hyperactivity

The most important questions concerning the treatment of hyperactivity
with drugs are questions such as which drugs are most effective for the
treatment of hyperactivity, do drugs have differential impact on different
types of outcomes, and are drugs more effective with certain types of children.
Data presented in the preceding section are useful in helping to interpret
answers to questions such as these.

Relative effectiveness of different drugs. The data presented in Figure 3
suggest that in general, drugs do have a positive effect on the symptoms of
hyperactivity, but which drugs are most effective? Table 12 presents data

which indicate the relative effectiveness for the most frequently used drugs. Only six different drugs appear in this table even though more than 50 drugs were identified and coded in the meta-analysis. These six drugs represent those drugs which were most frequently used. As can be seen, dextroamphetamine and methylphenidate are the most frequently used drugs; and, according to these data, methylphenidate is the most effective. However, differences between the various drugs are not great. Given the relatively small number of effect sizes upon which these estimates are based, one can be completely confident about this conclusion.

Table 12

Average Effect Size for Different Drugs (Versus Control or Non-Treatment Group) Broken Down by Quality of Research Design

Type of Drug	Quality of Research High (1,2)			Low (3,4,5)			All Studies		
	ES	n	SEM	ES	n	SEM	ES	n	SEM
Chlorpromazine	-.38	10	.22				-.38	10	.22
Thioridazine	.35	12	.20	.25	6	.29	.32	18	.16
Dextro-amphetamine	.21	55	.09	.85	44	.11	.49	99	.07
Methylphenidate	.44	84	.08	.42	157	.06	.43	241	.05
Imipramine				.46	9	.15	.46	9	.15
Magnesium Remoline				.50	30	.13	.50	30	.13

More information on the relative effectiveness of drugs is given in Table
13. These data are taken from those studies where one drug was compared to
another drug. In general, the data in Table 13 support the conclusion that
methylphenidate is the most effective drug in reducing the symptoms of
hyperactivity. For example, in those studies which compared methylphenidate to
dextroamphetamine, children receiving methylphenidate were .2 of a standard
deviation higher across all dependent variables considered than were children
receiving dextroamphetamine. These results also must be interpreted
cautiously, however, since only 18 effect sizes are included in this
calculation. However, given the research which has been conducted, the data in
Table 12 and 13 suggest that methylphenidate is the medication of choice for
treating hyperactive children in terms of improvement on dependent measures.
Given the slim margin of apparent benefit, however, other considerations such
as cost, side effects, and feasibility of administration should be considered
carefully in making choices.

Table 13

Average Effect Size for Comparisons
of One Drug with Another

"Experimental" Group	"Control" Group	$\overline{ES}$	n	SEM
Methylphenidate vs. Dextroamphetamine		.20	(18)	.16
Methylphenidate vs. Imipramine		.38	(11)	.21
Methylphenidate vs. Thioridazine		.69	(24)	.14
Dextroamphetamine vs. Magnesium Pemoline		.41	(30)	.13

Age of child. Age has frequently been suggested as an important variable in understanding hyperactivity because many authors have suggested that most children will "grow out of" being hyperactive by the time they enter adolescence. In the meta-analysis, children were categorized into four age groups as shown in Table 14. As can be seen, the average effect size for children under 9 years of age is approximately double that of children from 9 to 12 years of age. Although these data are cross-sectional rather than longitudinal in nature, they do suggest that drugs are more effective with younger children. This may be because in those children where hyperactivity persists to the later ages, the condition is more severe and thus, less responsive to treatment than hyperactivity in younger children.

Table 14

Average Effect Size for Different
Ages of Children

Age	ES	n	SEM
0 - 84 mos (0 - 8 yrs)	.40	(95)	.08
85 - 108 mos (8 - 9 yrs)	.50	(441)	.03
109 - 120 mos (9 - 10 yrs)	.28	(107)	.07
121 - 144 mos (10 - 12 yrs)	.26	(60)	.11

Gender. Another consistently reported finding in the literature is that more boys than girls exhibit hyperactivity. Table 15 shows the average effect size for studies which had differing percentages of boys included in their experimental samples. As can be seen, the studies which included almost all boys had an effect size roughly double that of studies which had only 50% boys. Furthermore, it can be seen that most effect sizes came from studies which were composed primarily of boys. These data could be interpreted in a number of ways. Perhaps, drugs are more effective with boys than with girls. Alternatively, data may suggest that since educators are convinced that more boys are hyperactive than girls, they are more likely to misidentify boys as being hyperactive than they are girls, and the larger effect size for boys has been inflated by spontaneous remission. In any case, the differences are substantial.

Table 15

Average Effect Size Broken Down by Percentage
of Male Subjects in Experimental Group

% Males in Experimental Group	ES	n	SEM
0 - 50%	.25	86	.08
51 - 85%	.33	162	.05
86 - 100%	.52	467	.001

Relationship of socioeconomic status. The meta-analysis also examined whether children in different socioeconomic groups responded differently to drug treatment. As can be seen in Table 16, small differences were identified between high and low socioeconomic groups, but these differences were not large enough to be attributed to anything more than sampling fluctuation.

Table 16

Average Effect Size for Different
Levels of SES

SES	ES	n	SEM
High	.67	(37)	.07
Medium	.60	(232)	.04
Low	.54	(53)	.06

Length of treatment. Length of treatment is another important consideration in administering drugs to children for hyperactivity. Is hyperactivity a condition that can be "cured" by the administration of drugs like pneumonia; or does administering drugs only suppress the symptoms but not ameliorate the condition? Some evidence on this important question is contained in the average effect size for children who received drugs for varying lengths of time. As can be seen in Table 17, average effect sizes tended to increase the longer the treatment was given up to 6 1/2 months. This trend was more pronounced when the data were limited to only high quality studies. The fact that the trend does not hold true after 6 1/2 months is probably attributable to the low number of effect sizes in those instances.

Table 17

Average Effect Size for Drug Treatment Based on Length of Treatment
and Broken Down by Quality of Research Design

Duration of Treatment	Overall ES			Quality of Research Design High ---- Low					High Quality (1,2)			Low Quality (3,4,5)		
	ES	n	SEM	1	2	3	4	5	ES	n	SEM	ES	n	SEM
1 month	.18	(148)	.06	.43 (8)	.02 (84)	.36 (49)	.47 (7)		.06	(92)	.07	.37	(56)	.12
1.5 months	.49	(189)	.05		.46 (104)	.53 (42)	.69 (25)	.34 (18)	.46	(104)	.07	.54	(85)	.08
3.3 months	.63	(188)	.05		.56 (41)	.62 (105)	.69 (12)	.71 (30)	.56	(41)	.11	.64	(147)	.06
6.6 months	.90	(36)	.12		.88 (18)			.99 (16)	.88	(18)	.16	.99	(16)	.18
12 months	.01	(9)	.23			.01 (9)						.01	(9)	.23
5 years	.51	(57)	.09	.29 (14)		.36 (7)	.24 (7)	.72 (29)	.29	(14)	.19	.58	(43)	.11

Additional information on this question is presented in Table 18 which examined length of treatment by type of drug used. Only dextroamphetamine and methylphenidate had sufficient number of effect sizes to be considered. As can be seen, however, the same trend appears to hold.

Table 18

Average Effect Sizes for Dextroamphetamine and Methylphenidate
According to Length of Treatment

Drug	Duration of Treatment					
	1 month	1.5 months	3.3 months	6.6 months	12 months	5 years
Dextroamphetamine	-.15 (36)	.25 (24)	.67 (71)	.99 (9)		
Methylphenidate	.22 (75)	.55 (89)	.70 (45)	.84 (8)		.49 (57)

Type of outcome. As noted in the beginning of this report, hyperactivity is characterized by a variety of problems. What is the effect of drugs on these various problems? This important question is answered to some degree by the data in Table 19 which examined the average effect size for different types of outcomes. The outcomes included in the table range from general hyperactivity and general behavior to indications of impulsivity, attention/vigilance, and IQ and academic achievement measures. As can be seen, the most substantial effects are found for those outcomes which are most subjectively measured and hence, most suspect to bias. However, when we limit these studies to those that were of high quality, substantial effects are still present for general hyperactivity, general behavior, activity level, and academic achievement and

attention/vigilance. Much lower effects are seen for aggression, impulsivity, IQ, and affective outcomes such as self-concept. Unfortunately, many of these estimates were based on rather small numbers of effect sizes. However, given the currently available data. Table 19 contains important information about the outcomes for which drugs are most effective.

Table 20 should be considered in conjunction with Table 19. As can be seen in Table 20, the average effect size for outcomes that are gathered via opinions is substantially higher than any of the other ways of collecting data. Those methods for gathering data which are least subject to bias such as systematic observation, actometers, and experimental tasks, show much lower effect sizes in general than measures which are more subject to bias such as opinions and ratings.

Degree of structure in setting where outcome measured. An important question has been raised in the literature about the degree to which the results of treating hyperactivity with drugs varies depending on the type of setting in which data are collected. Some researchers have suggested that drugs are primarily useful because they help the child to control their impulsivity and to remain on-task. These people have argued that in unstructured or free play settings, drugs may not have such a noticeable effect. If true, the lack of drug effect in unstructured setting would result because hyperactivity is more a problem of impulse control than of excessive activity; and unstructured settings do not require as much impulse control. Table 21 shows the results of studies in which children were observed in structured and unstructured settings. These data tend to support the hypothesis that hyperactivity is more a function of impulse control than of excessive activity. As noted, the average effect size obtained in structured settings is more than twice as high as the average effect size in unstructured settings.

Table 19

Average Effect Size for Different Outcomes Broken
Down by Methodological Quality of Study

Type of Outcome	Overall ES			Quality of Research Design High ← → Low					High Quality (1,2)			Low Quality (3,4,5)		
	ES	n	SEM	1	2	3	4	5	ES	n	SEM	ES	n	SEM
General Hyperactivity	.64	(175)	.05		.66 (48)	.60 (69)	.73 (30)	.61 (28)	.66	(48)	.10	.63	(127)	.06
General Behavior	.74	(105)	.05		.79 (40)	.63 (40)		.89 (23)	.79	(40)	.11	.72	(63)	.09
Activity Level	.64	(43)	.11	1.21 (6)	.26 (11)	.71 (20)			.59	(17)	.17	.66	(25)	.14
Achievement	.63	(27)	.13		.62 (20)				.62	(20)	.16			
Aggression	.47	(18)	.16		.20 (8)	.46 (8)			.20	(8)	.25	.46	(8)	.25
Impulsivity	.41	(32)	.12		.21 (11)	.41 (14)			.21	(11)	.21	.41	(14)	.19
Attention/ Vigilance	.23	(91)	.07		.45 (23)	.38 (32)	-.23 (29)		.45	(23)	.15	.16	(66)	.09
IQ	.20	(117)	.06	.27 (10)	-.10 (50)	.41 (22)	.05 (16)	.32 (7)	-.05	(60)	.09	.33	(35)	.12
Affective	.11	(33)	.12		.15 (22)			.17 (7)	.15	(22)	.15	.17	(7)	.26

60

67

Table 20

Average Effect Size for Different Types
of Instruments Used to Measure Outcomes

Type of Instrument	$\overline{ES}$	n	SEM
Opinion	1.15	(27)	.13
Rating	.58	(360)	.04
Systematic Observation	.22	(28)	.13
Actometer	.49	(34)	.12
Standardized Test	.37	(107)	.07
Experimental Task	-.08	(123)	.06

Table 21

Average Effect Size for Ratings Collected in
Structured and Unstructured Settings

Type of Setting	Quality of Research					
	High (1,2)			Low (3,4,5)		
	$\overline{ES}$	n	SEM	$\overline{ES}$	n	SEM
Structured	.79	(35)	.12	.77	(27)	.13
Unstructured	.33	(63)	.09	.55	(97)	.07

Methodological Notes

In conducting a meta-analysis of this scope, some interesting findings regarding the methodology of research integration are often identified. One important finding that emerged from this study concerns the data used in computing effect sizes. The notion of using a common metric for comparing outcomes across studies and types of dependent variables provides substantial flexibility and power in examining the results of previous research. Unfortunately, as pointed out in the Procedures section, many studies do not report the means and standard deviations necessary for computing an effect size. In those cases, alternative estimation methods have been proposed and are frequently used in meta-analysis studies. These estimation methods employ various assumptions which are infrequently, if ever, checked. In Table 22, the average effect size obtained for different ways of computing effect sizes is reported. The fact that the average effect size for studies which reported means and standard deviations is much lower than studies in which effect sizes had to be estimated using one of the other approaches is concerning. These data suggest that when effect sizes may be somewhat inflated when they are calculated from t or F ratios, t or F probabilities, or percentage of the sample exceeding a given criterion (e.g., percentage improved). This question deserves further investigation.

Table 22

Average Effect Size for Different Ways of
Constructing Effect Size Estimates

Information Used to Construct Effect Size	ES	n	SEM
$\bar{X}$s and SD (either control, pooled, published)	.29	(452)	.03
t or F ratio or probability	.70	(81)	.05
Percentage Improved ("Probit" Transformation)	.82	(120)	.06

Table 23 contains information regarding the relationship between authors' conclusions and computed effect size. Many critics of research literature have suggested that authors place entirely too much import on the statistical significance of findings and often conclude that a treatment has worth based on statistical significance when the educational importance of the finding is trivial. This argument is contradicted by the average effect size for those studies where authors concluded that the treatment worked, could not be determined, or did not work. Based on these averages, it appears that most researchers have a fairly good understanding of what constitutes educational or clinical significance.

Table 23

Average Effect Size Associated with
Different Conclusions by Author(s)

	ES	n	SEM
Treatment works	.86	(297)	.03
Cannot determine	.42	(36)	.07
Treatment does not work	.04	(247)	.04

Summary

The results of the meta-analysis suggest that drugs are a moderately effective treatment for hyperactivity in children. Of those drugs available, methylphenidate appears to be the most effective, but the margin of advantage is very slim. Furthermore, a number of drugs which appear promising have not

been investigated sufficiently to draw firm conclusions as to the relative advantages of methylphenidate.

The general conclusion that drugs are moderately effective in treating hyperactivity must be placed in the appropriate context. A fairly large number of variables were identified which indicate that the apparent benefits of drugs for treating hyperactivity are somewhat overestimated. For example, accounting for factors such as poor procedures in classifying children as hyperactive, low quality research designs, unreliability and bias in outcome measures, and suggested bias by those people supporting much of the research being conducted, all tend to reduce the obtained effect size of drug treatment.

The data also suggest that drugs are more effective with younger children than with older children, with boys than with girls, and when continued for longer periods. The greatest effect of drugs is for outcomes which are more generally defined. Those studies which have considered the effect of drugs on such variables as IQ, impulsivity, attention, have found much smaller effects. There is support in the literature for the suggestion that hyperactivity is more a problem of impulse control rather than excessive activity because drugs are substantially more effective in structured settings than in unstructured settings.

Although the meta-analysis has done much to clarify the results of previous research on whether drugs are an effective treatment for hyperactivity, many questions remain. The data are suggestive about the relative effectiveness of different drugs, but by no means definitive. Further research needs to be done comparing which drugs are most effective for which children. Also, the data regarding the degree of structure in which the outcome is measured raises important questions about the definition and ideology of hyperactivity. These questions need further investigation.

Much of the data from the meta-analysis also suggest guidelines for future research. Although a good deal of research has been done well, much of the research on which the meta-analysis is based suffered from one or more important problems. The results of the current meta-analysis suggest that future research should be more careful to control the following variables: bias in the support of the research, bias in the collection of data, rigorous classification of children as hyperactive, the use of placebos, and procedures for assuring that implementation has occurred. Finally, the meta-analysis data suggest that questions such as age, sex differences, and length of treatment ought to be investigated further. Of particular importance in such ongoing investigations would be longitudinal studies and follow-up studies of children who have previously received drug treatment for hyperactivity.

REFERENCES

Adelman, H. S., & Compas, B. E. Stimulant drugs and learning problems.
    Journal of Special Education, 1977, 11, 377-416.

Barkley, R. A. Using stimulant drugs in the classroom. School Psychologist,
    1979, 8, 412-425.

Barkley, R. A., & Cunningham, C. E. The parent-child interactions of
    hyperactive children and their modification by stimulant drugs. In R.
    Knight & D. Bakker (Eds.), Rehabilitation, treatment, and management of
    learning disabilities. Baltimore, Md.: University Park Press, 1979a.

Blouin, A. G., Bornstein, M. A., & Trites, R. L. Teenage alcohol abuse among
    hyperactive children: A five-year follow-up study. Journal of Pediatric
    Psychology, 1978, 3, 188-194.

Bussey, H. R. Study of the prevalence and therapy of the choreatiform syndrome
    and hyperkinesis in rural Vermont. Acta Paedopsychiatry, 1967, 34,
    130-135.

Campbell, S. Mother-child interaction in reflective, impulsive, and
    hyperactive children. Developmental Psychology, 1973, 8, 341-347.

Campbell, S. Mother-child interaction: A comparison of hyperactive, learning
    disabled, and normal boys. American Journal of Orthopsychiatry, 1975, 45,
    51-57.

Cantwell, D. P., & Satterfield, J. The prevalence of academic underachievement
    in hyperactive children. Journal of Pediatric Psychology, 1978, 3,
    168-171.

Clifford, G. J. A history of the impact of research on teaching. In R. M. W.
    Travers (Ed.), Second handbook of research on teaching. Chicago: Rand
    McNally, 1973.

Cooper, H. M. Scientific guidelines for conducting integrative research
    reviews. Review of Educational Research, 1982, 52(2), 291-302.

Davidson, T. B. Meta-analysis of the neuropsychological assessment of
    children. Doctoral dissertation, University of Denver, 1978.

Educational Research Services. Class-size research: A critique of recent
    meta-analyses. Kappan, 1980, 239-241.

Eisenberg, L., & Conners, C. K. Psycho-pharmacology in childhood. In N. B.
    Talbot, J. Nagan, & L. Eisenberg (Eds.), Behavioral science in pediatric
    medicine. Pennsylvania: Saunders, 1971.

Eysenck, H. J. An exercise in mega-silliness. American Psychologist, 1978,
    33, 517.

Feldman, K. A. Using the work of others: Some observations on reviewing,
    integrating, and consolidating findings. Sociology of Education, 1971,
    44, 86-102.

Freeman, R. D. Minimal brain dysfunction, hyperactivity, and learning disorders: Epidemic or episode. School Review, 1976, 85, 5-30.

Gallo, P. S. Meta-analysis--a mixed meta-phor. American Psychologist, 1978, 33, 515-517.

Glass, G. V. Integrating findings: The meta-analysis of research. Review of Research in Education, 1977, 5, 351-379.

Glass, G. V. Primary, secondary, and meta-analysis of research. Educational Researcher, 1976, 5, 3-8.

Glass, G. V. Reply to Mansfield and Bussey. Educational Researcher, 1978, 7, 3.

Glass, G. V, McGaw, B., & Smith, M. L. Integrating research studies: Meta-analysis of social research. Beverly Hills, Ca.: Sage Publications, 1981.

Glass, G. V, & Smith, M. L. Meta-analysis of research on the relationship of class size and achievement. Evaluation and Policy Analysis, 1979, 1, 2-16.

Glass, G. V, & Smith, M. L. Reply to Eysenck. American Psychologist, 1978, 33, 517-518.

Glass, G. V, Smith, M. L., & Barton, M. A. Methods of integrative analysis. (Annual report, NIE Grant No. G-78-0148). Boulder, Co.: Laboratory of Educational Research, University of Colorado, August 1979.

Grinspoon, L., & Singer, S. B. Amphetamines in the treatment of hyperkinetic children. Harvard Educational Review, 1973, 43, 515-555.

Jackson, G. B. Methods for reviewing and integrating research in the social sciences. Final report to the National Science Foundation for Grant No. DIS 76-20398. Washington, D.C.: Social Research Group, George Washington University, April 1978.

Jackson, G. B. Methods for integrative reviews. Review of Educational Research, 1980, 50(3), 438-460.

Kerlinger, F. N. The influence of research on education practice. Educational Researcher, 1977, 6, 5-12.

Kinsbourne, M., & Swanson, J. M. Models of hyperactivity implications for diagnosis and treatment. In R. L. Trites (Ed.), Hyperactivity in children. Baltimore: University Park Press, 1979.

Light, R. J., & Smith, P. V. Accumulating evidence: Procedures for resolving contradictions among different research studies. Harvard Educational Review, 1971, 41, 429-471.

Mansfield, R. S., & Bussey, T. V. Meta-analysis of research: A rejoinder to Glass. Educational Researcher, 1977, 6, 3.

Miller, Jr., R. G., Palkes, H. S., & Steward, M. A. Hyperactive children in suburban elementary schools. Child Psychiatry and Human Development, 1973, 4(2).

Ross, D. M., & Ross, S. A. Hyperactivity. New York: Wiley, 1976.

Safer, R., & Allen, D. Hyperactive children: Diagnosis and management. Baltimore: University Park Press, 1976.

Shaver, J. P. The usefulness of educational research in curricular/instructional decision-making in social studies. Theory and Research in Social Education, 1979, 7, 21-46.

Simpson, S. N. Comment on "meta-analysis of research on class-size and achievement". Educational Evaluation and Policy Analysis, 1980, 2, 81-83.

Smith, M. L. Sex bias in counseling and psychotherapy. Boulder, Co.: Laboratory of Educational Research, University of Colorado, October 1978.

Smith, M. L., & Glass, G. V. Meta-analysis of research on class size and its relationship to attitudes and instruction. American Educational Research Journal, 1980, 17(4), 419-434.

Sprague, R. L. Discussion of prevalence of hyperactivity in Ottawa. In R. L. Trites (Ed.), Hyperactivity in children. Baltimore: University Park Press, 1979.

Stewart, M. A. Is hyperactivity abnormal? and other unanswered question. School Review, 1975, 85, 31-42.

Strike, K. A. An epistemology of practical research. Educational Researcher, 1979, 8(1), 10-16.

Taylor, C., & White, K. R. The effect of reinforcement and training on group standardized test behavior. Journal of Educational Measurement, 1982, in press.

Trites, R. L. Prevalence of hyperactivity in Ottawa, Canada. In R. L. Trites (Ed.), Hyperactivity in children. Baltimore: University Park Press, 1979.

Weiss, G., Hechtman, L., Perlman, T., Hopkins, J., & Wener, A. Hyperactive children as young adults: A controlled prospective 10-year follow-up of the psychiatric status of 75 hyperactive children. Archives of General Psychiatry, 1979.

Wender, P. H. Minimal brain dysfunction in children. New York: Wiley - Interscience, 1971.

White, K. R. An integrative review of the relationship between socioeconomic status and academic achievement using meta-analysis techniques. Psychological Bulletin, 1982, in press.

Wolins, L. Responsibility for raw data. American Psychologist, 1962, 17, 657-658.

Appendix 1

References to and Analysis of

Review Articles

CHARACTERISTICS OF HYPERACTIVITY REVIEWS

Review References	Type of Sample [a]	Method of Selection Specified? [b]	Number of Primary Studies Cited [c]	Previous reviews cited, critiqued, and expanded? [d]	Outcomes of individual studies reported in terms of [e]	How were concomitant variables considered? [f]
Adelman, H. S., & Compas, B. E. Stimulant drugs and learning problems. Journal of Special Education, 1977, 11, 377-416.	Convenience	No	21	No	Statistical significance; differences study by study; brief differences	Not considered
Allen, R. P., Safer, D., & Covi, L. Effects of psychostimulants on aggression. The Journal of Nervous and Mental Disease, 1975, 160, 138-145.	Convenience	No	3	No	Differences study by study	Not considered
Bakwin, H. Benzedrine in behavior disorders of children. The Journal of Pediatrics, 1948, 32, 215-216.	Convenience	No	3	No	Differences study by study	Not considered
Barkley, R. A. A review of stimulant drug research with hyperactive children. Journal of Child Psychology and Psychiatry, 1977, 18, 137-165.	Convenience	No	37	No	Brief differences	Not considered
Barkley, R. A. Predicting the response of hyperkinetic children to stimulant drugs: A review. Journal of Abnormal Child Psychology, 1976, 4, 327-348.	Convenience	No	78	Yes	Brief differences	Not considered
Barkley, R. A. Recent developments in research on hyperactive children. Journal of Pediatric Psychology, 1978, 3, 158-163.	Convenience	No	0	No	Not reported	Not considered
Barkley, R. A., & Cunningham, C. E. Do stimulant drugs improve the academic performance of hyperactive children? Clinical Pediatrics, 1978, 17, 85-92.	Convenience	No	42	No	Statistical significance; differences study by study; brief differences	Logical for subset
Bradley, E. Academic, behavioral, and psychological responses of hyperactive children to stimulant medication. Unpublished master's thesis, Northeastern Illinois University, August 1975. (ED 116413)	Convenience	No	23	No	Statistical significance; differences study by study; brief differences	Not considered
Bower, K. B. Hyperactivity: Etiology and intervention techniques. The Journal of School Health, 1975, 45, 195-202.	Convenience	No	18	No	Statistical significance; single subject; differences study by study; brief differences	Not considered
Calhoun, G. Hyperactive emotionally disturbed and hyperkinetic learning disabilities: A challenge for the regular classroom. Adolescence, 1978, 13, 335-338.	Convenience	No	1	No	Differences study by study	Not considered

Key for interpreting each column appears at the end.

74

Review References	Type of Sample [a]	Method of Selection Specified? [b]	Number of Primary Studies Cited [c]	Previous reviews cited, critiqued, and expanded? [d]	Outcomes of individual studies reported in terms of [e]	How were concomitant variables considered? [f]
Cole, S. O. Hyperkinetic children: The use of stimulant drugs evaluated. American Journal of Orthopsychiatry, 1975, 45, 28-37.	Convenience	No	11	No	Single subject; differences study by study; brief differences	Not considered
Conners, C. K. Recent drug studies with hyperkinetic children. Journal of Learning Disabilities, 1971, 4, 476-483.	Convenience	No	14	Yes	Statistical significance; differences study by study; brief differences	Not considered
Cronin, J. P. The use of psychopharmaceutical stimulants for the control of childhood hyperkinesis. Minneapolis, Minnesota: University of Minnesota, 1975. (ERIC Document Reproduction Service No. ED126 654)	Convenience	No	4	No	Statistical significance; single subject; differences study by study; brief differences	Not considered
Cunningham, C. E., & Barkley, R. A. The role of academic failure in hyperactive behavior. Journal of Learning Disabilities, 1978, 11, 15-21.	Convenience	No	17	No	Single subject; differences study by study; brief differences	Not considered
Conners, C. K. Pharmacotherapy of psychopathology in children. In H. Quay & J. Werry (Eds.), Psychopathological disorders of children. New York: Wiley & Sons, 1972.	Convenience	No	84	No	Statistical significance; differences study by study; brief differences	Not considered
DeLong, A. R. What have we learned from psychoactive drug research on hyperactives? American Journal of the Disabled Child, 1972, 123, 177-180.	Convenience	No	0	No	Not considered	Not considered
Dyck, N. J. Educational management of hyperactive children. In M. J. Fine (Ed.), Principles and techniques of intervention with hyperactive children. Springfield, Illinois: Charles C. Thomas, 1977.	Convenience	No	12	No	Statistical significance; differences study by study; brief differences	Not considered
Eisenberg, L., & Conners, C. K. Psychopharmacology in childhood. In N. B. Talbot, J. Wagan, & L. Eisenberg (Eds.), Behavioral science in pediatric medicine. Philadelphia: Saunders, 1971.	Convenience	No	39	No	Statistical significance; single subject; differences study by study; brief differences	Not considered
Epstein, E. P., Harrington, N. D., Meagher, J. A., Rowlands, E. L., & Simons, R. K. Chemotherapy and the hyperkinetic child. Journal of Education, 1968, 151, 47-60.	Convenience	No	1	No	Differences study	Not considered

[a-f] Key for interpreting each column appears at the end.

Review References	Type of Sample [a]	Method of Selection Specified? [b]	Number of Primary Studies Cited [c]	Previous reviews cited, critiqued, and expanded? [d]	Outcomes of individual studies reported in terms of [e]	How were concomitant variables considered? [f]
Feighner, A. C., & Feighner, J. P. Multimodality treatment of the hyperkinetic child. American Journal of Psychiatry, 1974, 131, 459-463.	Convenience	No	9	No	Brief differences	Not considered
Fischer, K. C., & Wilson, W. P. Methylphenidate and the hyperkinetic state. Diseases of the Nervous System, 1971, 32, 695-698.	Convenience	No	8	No	Differences study by study	Not considered
Fish, B. Drug use in psychiatric disorders of children. American Journal of Psychiatry, 1968, 124, 31-36.	Convenience	No	7	No	Statistical significance; brief differences	Not considered
Freeman, R. D. Drug effects on learning in children: A selective review of the past thirty years. The Journal of Special Education, 1966, 1, 17-44.	Convenience	No	26	No	Statistical significance; differences study by study; brief differences	Not considered
Freeman, R. D. Review of medicine in special education. Journal of Special Education, 1970, 4, 377-384.	Convenience	No	9	No	Differences study by study; brief differences	Not considered
Glennon, C. A., & Nason, D. E. Managing the behavior of the hyperactive child: What research says. Reading Teacher, 1974, 27, 815-824.	Convenience	No	3	No	Statistical significance; differences study by study; brief differences	Not considered
Grant, D. R. Psychopharmacology in childhood emotional and mental disorders. Journal of Pediatrics, 1962, 61, 626-637.	Most research	No	62	No	Statistical significance; differences study by study; brief differences	Not considered
Grinspoon, L., & Singer, S. B. Amphetamines in the treatment of hyperkinetic children. Harvard Educational Review, 1973, 43, 515-555.	Convenience	No	15	No	Statistical significance; single subject; differences study by study; brief differences	Not considered
Havighurst, R. J. Choosing a middle path for the use of drugs with hyperactive children. School Review, 1976, 85, 61-77.	Convenience	No	2	No	Single subject; differences study by study	Not considered
Hirst, I. Effects of the psychoactive drug: Methylphenidate (Ritalin) on classroom disorders: Hyperactivity, emotional disturbance, and learning disorders. Paper presented at the Annual International Convention, The Concil for Exceptional Children, Chicago, April 1976.	Convenience	No	3	No	Differences study by study	Not considered

[a-f] Key for interpreting each column appears at the end.

7ɔ

Review References	Type of Sample[a]	Method of Selection Specified?[b]	Number of Primary Studies Cited[c]	Previous reviews cited, critiqued, and expanded?[d]	Outcomes of individual studies reported in terms of[e]	How were concomitant variables considered?[f]
Keogh, B. Hyperactivity and learning disorders: Review and speculation. Exceptional Children, 1971, 38, 101-109.	Convenience	No	5	No	Differences study by study	Not considered
Kornetsky, C. Psychoactive drugs in the immature organism. Psychopharmacologia, 1970, 17, 105-136.	Convenience	No	16	No	Statistical significance; differences study by study	Not considered
Lambert, N. M., Windmiller, M., Sandoval, J., & Moore, B. Hyperactive children and the efficacy of psychoactive drugs as a treatment intervention. American Journal of Orthopsychiatry, 1976, 46, 335-352.	Convenience	No	37	No	Statistical significance; differences study by study; brief differences	Not considered
Laufer, M. W., Denhoff, E., & Solomons, G. Hyperkinetic impulse disorder in children's behavior problems. Psychosomatic Medicine, 1957, 19, 38-49.	Convenience	No	4	No	Differences study by study; brief differences	Not considered
Lesser, L. L. Hyperkinesis in children. Clinical Pediatrics, 1970, 9, 548-552.	Convenience	No	2	No	Brief differences	Not considered
Lipman, R. S. NIMH-DRB support of research in minimal brain dysfunction and other disorders of childhood. Psychopharmacology Bulletin, 1973, 1-8.	Convenience	No	13	No	Differences study by study; brief differences	Not considered
Lipton, M. et al. Report to the Nutrition Foundation. New York: The Nutrition Foundation, Inc., 1975.	Convenience	No	2	No	Differences study by study	Not considered
McDonald, J. E. Pharmacologic treatment and behavior therapy: Allies in the management of hyperactive children. Psychology in the Schools, 1978, 15, 270-274.	Convenience	No	11	No	Statistical significance; single subject	Not considered
Millichap, J. G., & Fowler, G. W. Treatment of "minimal brain dysfunction" syndromes. Pediatric Clinics of North America, 14(4), 1967, 767-777.	Convenience	No	36	No	Percent improved	Logical for subset
Mira, M., & Reece, C. A. Medical management of the hyperactive child. In M. J. Fine (Ed.), Principles and techniques of intervention with hyperactive children. Springfield, Illinois: Charles C. Thomas, 1977.	Convenience	No	17	No	Statistical significance; differences study by study; brief differences	Not considered

[a-f]Key for interpreting each column appears at the end.

Review References	Type of Sample[a]	Method of Selection Specified?[b]	Number of Primary Studies Cited[c]	Previous reviews cited, critiqued, and expanded?[d]	Outcomes of individual studies reported in terms of[e]	How were concomitant variables[f] considered?
O'Leary, K. D. Pills or skills for hyperactive children. Presidential address to Clinical Division, Experimental Behavioral Science, American Psychological Association, Toronto, Canada, August 1970.	Convenience	No	29	No	Statistical significance; differences study by study; brief differences	Not considered
Patterson, G. R. Behavioral intervention procedures in the classroom and in the home. In A. E. Bergin & S. L. Garfield (Eds.), Handbook of Psychotherapy and Behavior Change. New York: Wiley & Sons, 1971.	Methodological superiority	No	33	No	Single subject; differences study by study; brief differences	Not considered
Prout, H. T. Behavioral intervention with hyperactive children: A review. Journal of Learning Disorders, 1977, 10(3), 141-142.	Convenience	No	14	Yes	Statistical significance, time series; differences study by study	Not considered
Ross, D. M., & Ross, S. D. Hyperactivity: Research, theory, and action. New York: Wiley, Interscience Publication, 1976.	Convenience	No	120	No	Statistical significance; single subject; differences study by study; brief differences	Not considered
Safer, D. J. Drugs for problem school children. The Journal of School Health, 1971, 41, 491-495.	Convenience	No	3	No	Statistical significance; differences study by study; brief differences	Not considered
Schrager, J., Harrison, S., McDermott, J., Wilson, P., & Lindy, J. The hyperkinetic child; an overview of the issues. Ann Arbor, Michigan: The University of Michigan Medical Center, Department of Psychiatry, 1965 (estimated from latest references in bibliography).	Convenience	No	3	No	Brief differences	Not considered
Sieben, R. L. Controversial medical treatment of learning disabilities. Academic Therapy, 1977, 13, 133-147.	Convenience	No	4	No	Single subject; differences study by study	Not considered
Silver, L. B. Acceptable and controversial approaches to treating the child with learning disabilities. Pediatrics, 1975, 55, 406-415.	Convenience	No	15	No	Differences study by study; brief differences	Not considered
Sprague, R. L., & Werry, J. S. Methodology of psychopharmacological studies with the retarded. In N. R. Ellis (Ed.), International review of research in mental retardation (Vol. 5). New York: Academic Press, 1971, 147-219.	Most research	No	184	Yes	Statistical significance; single subject; differences study by study; brief differences	Logical for subset

a-fKey for interpreting each column appears at the end.

ERIC

81

Review References	Type of Sample[a]	Method of Selection Specified?[b]	Number of Primary Studies Cited[c]	Previous reviews cited, critiqued, and expanded?[d]	Outcomes of individual studies reported in terms of[e]	How were concomitant variables considered?[f]
Spring, C., & Sandoval, J. Food additives and hyperkinesis: A critical evaluation of the evidence. Journal of Learning Disabilities, 1976, 9, 560-569.	Most research	No	6	No	Statistical significance; differences study by study; brief differences	Not considered
Sroufe, L. A., & Stewart, M. A. Treating problem children with stimulant drugs. New England Journal of Medicine, 1973, 289, 407-413.	Convenience	No	30	No	Differences study by study; brief differences	Not considered
Swift, M. S., & Spivack, G. Therapeutic teaching: A review of teaching methods for behaviorally troubled children. The Journal of Special Education, 1974, 8, 259-289.	Convenience	No	28	No	Differences study by study; brief differences	Not considered
Taylor, E. Food additives, allergy, and hyperkinesis. Journal of Child Psychology, 1979, 20, 357-363.	Convenience	No	8	No	Statistical significance; differences study by study; brief differences	Not considered
Tobiessen, J., & Karowe, H. E. A role for the school in the pharmacological treatment of hyperkinetic children. Psychology in the Schools, 1969, 6, 340-346.	Convenience	No	10	No	Statistical significance; differences study by study; brief differences	Not considered
Weiss, G., & Hechtman, L. The hyperactive child syndrome. Science, 1979, 205, 1348-1354.	Convenience	No	23	No	Statistical significance; differences study by study; brief differences	Not considered
Werry, J. S. Developmental hyperactivity. Pediatric Clinics of North America, 1968, 15, 581-599.	Convenience	No	13	No	Differences study by study; brief differences	Not considered
Williams, J. I., & Cram, D. M. Diet in the management of hyperkinesis. Canadian Psychiatric Association Journal, 1978, 23, 241-248.	Convenience	No	10	No	Differences study by study	Not considered
Wolraich, M. L. Behavior modification therapy in hyperactive children. Clinical Pediatrics, 1979, 18, 563-570.	Convenience	No	7	No	Brief differences	Not considered
Wolraich, M. L. Stimulant drug therapy in hyperactive children: Research and clinical implications. Pediatrics, 1977, 60, 512-518.	Convenience	No	62	Yes	Differences study by study; brief differences	Logical for subset

a-f Key for interpreting each column appears at the end.

Review References	Type of Sample[a]	Method of Selection Specified?[b]	Number of Primary Studies Cited[c]	Previous reviews cited, critiqued, and expanded?[d]	Outcomes of individual studies reported in terms of[e]	How were concomitant variables considered?[f]
Wunderlich, R. C. Treatment of the hyperactive child. Academic Therapy, 1973, 8, 375-390.	Convenience	No	0	No	Not considered	Logical for subset
Zentall, S. S. Environmental stimulation model. Exceptional Children, 1977, 43, 502-510.	Convenience	No	26	No	Brief differences	Logical for subset
Zentall, S. Optimal stimulation as theoretical basis of hyperactivity. American Journal of Orthopsychiatry, 1975, 45, 549-563.	Convenience	No	32	No	Differences study by study; brief differences	Logical for subset

[a]The type of sample was coded as convenience, methodologically superior, representative, or comprehensive. Decisions about the type of sample included in each review were somewhat subjective. If the review was based on a limited number of studies and gave no procedures for the selection techniques used to assure representativeness, it was assumed that the sample was a convenience sample. If procedures had been used to assure a representative or comprehensive sample, we assumed the author would have mentioned them. Samples coded as methodologically superior were described as such by the author(s).

[b]To be coded "yes," the specific procedures used to identify and select articles for the review had to be described. It was not enough to say, "articles were limited to research on home-based programs," since the procedures for identifying and selecting home-based articles are not specified.

[c]Numbers cited in this column represent only those empirical research articles used to support a contention about the effectiveness of a treatment for hyperactivity. All articles listed in the article's bibliography or reference section are not listed.

[d]To be coded "yes," a review article had to cite previous review articles and critique them and explain how the current review would differ from or expand on previous reviews--all three were necessary.

[e]The way in which the outcome of each study was reported in the review was coded as Effect Size (i.e., any kind of measure which could be compared across studies like Glass's ES, eta [n] squared, r^2, or omega [ ] squared); Statistically Significant (i.e., the statistical significance whether in favor or against the particular treatment reported for each study; Differences (i.e., studies were considered individually and differences found were reported but without reporting an ES measure or statistical significance); Brief Differences (i.e., differences found by studies were reported in groups instead of study by study); or Single Subject Designs. Entries in the column represent the most frequent way(s) of reporting outcomes of individual studies for a particular review.

[f]The way in which the review considered how study characteristics covaried with outcome was coded as Systematically (data based) (i.e., the covariation was examined where possible for all studies in sample); Logical for Subset (i.e., covariation was discussed for substantial number of studies in review but not using data-based approach); or Not considered.

Appendix 2

Coding Sheet for Efficacy of Drug

Treatments for Hyperactivity

Coder: _____

☐ Question for G^2, explain on back

META-ANALYSIS OF HYPERACTIVITY

Coding Instrument

_____ (Authors)   _____ (Title)

I

(1-4) . . . . . _____ 1) Study ID #

(5,6) . . . . . _____ 2) Year

(7-10) . . . . . _____ 3) Source

(11) . . . . . _____ 4) Supported by commercial company

   (1 = yes, complete; 2 = yes, partial; 3 = no; 4 = not stated)

(12) . . . . . _____ 5) Dissertation

   (1 = dissertation article w/ ES estimated; 0 = all others)

(13) . . . . . _____ 6) Side Effect (1 = yes; 0 = no)

   _____
   (specify)

Check List

_____ 1. references needed from this article checked
_____ 2. computations for ES shown
_____ 3. every blank marked
_____ 4. test names listed for ES's
_____ 5. references of test articles to Bev
_____ 6. comments about conventions
_____ 7. disagreement about conventions
_____ 8. "coded" written on article
_____ 9. references and additions for future mini meta-analyses

---

II DESCRIPTION OF SAMPLE

	ES6	ES5	ES4	ES3	ES2	ES1	
14-16)							**1. MEAN AGE (MONTHS)** — Experimental (- = not given)
17-19)							Control (- = not given)
							2. MEAN IQ (1 = 130+, 2 = 71-129, 3 = 70-)
(20)							Experimental (- = not given) _____ (test)
(21)							Control (- = not given) _____ (test)
							3. SIZE
22-25)							Experimental (- = not given)
26-29)							Control (- = not given)
							4. SES (1 = high, 2 = low, 3 = middle, 4 = mixed, - = not given)
(30)							Experimental _____
(31)							Control _____ } (how SES determined)
							5. SEVERITY OF HYPERACTIVITY (0 = none, 1 = mild, 2 = moderate, 3 = severe, 4 = extreme, 5 = mixed, - = unable to tell)
(32)							Experimental
(33)							Control
							6. DIAGNOSIS
(34)							Experimental
(35)							Control

1 = normal
2 = hyperactive/hyperkinetic
3 = MBD
4 = LD
5 = ED
6 = attention deficit disorder
7 = other _____
          (specify)
- = unable to tell

(1)

85

	ES6	ES5	ES4	ES3	ES2	ES1

7. % HANDICAPPED (code , -9 = some, but exact not known, - = not given)

A. MULTIPLE

35-38)   Experimental

39-41)   Control

B. DEAF

42-44)   Experimental

45-47)   Control

C. BLIND

48-50)   Experimental

51-53)   Control

D. MR

54-56)   Experimental

57-58)   Control

E. LEARNING/PERCEPTION/MINOR NEUROLOGICAL PROBLEM (LD, MBD, soft neurological signs, minor EEG abnormalities)

59-61)   Experimental

62-64)   Control

F. GROSS NEUROLOGICAL PROBLEMS (obvious physical trauma, gross EEG abnormalities)

65-67)   Experimental

68-70)   Control

G. ED (neurosis/psychosis)

71-73)   Experimental

74-76)   Control

8. % INSTITUTIONALIZED (code %, -9 = some, - = not given)

77-79)   Experimental

(1-3)   Control

9. % MINORITY (black, hispanic, and/or immigrant) (code %, -9 = some, - = not given)

(4-6)   Experimental

(7-9)   Control

10. % MALE (code %, -9 = some, - = not given)

(10-12)   Experimental

(13-15)   Control

11. % SUBJECTS ON ACTIVE DRUGS (check time from list) (code %, -9 = some, - = not given)

(16-18)   Experimental

(19-21)   Control

12. % ALLERGIES (code %, -9 = some - = not given)

(22-24)   Experimental _____ specify allergence substance

(25-27)   Control _____

13. % DELINQUENT (code %, -9 = some, - = not given)

(28-30)   Experimental

(31-33)   Control

(2)

86

```
III  CLASSIFICATION OF CHILDREN
          AS HYPERACTIVE
```

	ES6	ES5	ES4	ES3	ES2	ES1
34-36)						
37-39)						
40-42)						
43-45)						
46-48)						
49-51)						
52-54)						
55-57)						
58-60)						
61-63)						
64-66)						
67-69)						
(70)						

Code four digit # for all selection methods used

1. Experimental Group

a) .... specify any "other" _____

b) .... specify any "other" _____

c) .... specify any "other" _____

d) .... specify any "other" _____

2. Control Group

a) .... specify any "other" _____

b) .... specify any "other" _____

c) .... specify any "other" _____

d) .... specify any "other" _____

3. General Quality of Procedures Used to Classify Children as Hyperactive

1. good: objective measures by 2 or more people or in 2 or more settings

2. fair: objective measure used by only 1 person in 1 setting - may or may not use additional subjective measures.

3. poor: only subjective measures used

4. unable to tell

Specify names of any instruments used to classify children as hyperactive

1. _____

2. _____

3. _____

4. _____

1st Digit - Basis for Selection

1 = General Hyperactivity

2 = Activity Level

3 = Attention/Vigilance

4 = Aggression

5 = Impulsivity

6 = Impossible to Determine

7 = Other

8 = Children were not hyperactive

2nd Digit - Instrument

1 = Opinion

2 = Rating

3 = Systematic Observation

4 = Actometer

5 = Perceptual-Motor Test

6 = Experimental Task

7 = Impossible to Determine

8 = Other

3rd Digit - Evaluator

1 = Teacher

2 = Parent

3 = Physician

4 = Researcher

5 = Observer

6 = Clinician/Psychologist

7 = Caretaker

8 = Impossible to Tell

9 = Other

4th Digit - Setting

1 = School

2 = Home

3 = Clinic/Doctor's Office

4 = Experimental Setting

5 = Day Care

6 = Institution

7 = Composite

8 = Impossible to Determine

9 = Other

	ES6	ES5	ES4	ES3	ES2	ES1

```
IV  TREATMENT
```

1) Treatment type (2 = only, 1 = 1 of several, 0 = no)

(71)   Drug

(72)   Behavioral

(73)   Diet

(74)   Biofeedback

(75)   Comparison of Groups

2) Control Group

(76)   A) Selection

    1 = random
    2 = convenience
    3 = matched
    4 = own
    5 = crossover
    6 = impossible to determine
    7 = children in control group nonhyperactive
    8 = other _____

(77)   B) Treatment

    1. None
    2. Placebo
(78)   Quality
         4 = present, can't tell
         3 = good
         2 = poor
         1 =
         0 = absent
    3. Drug (specify) _____
    4. Behavioral (specify)_____
    5. Diet (specify) _____
    6. Biofeedback (specify) _____
    7. Impossible to determine
    8. Other (specify) _____

(1-4)   3) Duration of treatment in days("-" = impossible to determine; -1 = N/A)

(5-8)   4) Days after treatment dependent variable observed (- = not given; -1 = N/A)

(9)   5) Reliability of treatment implementation

    1 complete implementation
    2 minor problems
    3 moderate problems
    4 major problems
    5 impossible to determine
    6. no treatment given--comparison of groups

(10)   6) Confidence with which IV, #5 was coded

    1 data based
    2 guess based on data
    3 convention
    4 impossible to determine
    5. not applicable

	ES6	ES5	ES4	ES3	ES2	ES1

V DESIGN

(11) 1) Type

    1 = random assignment
    2 = non-random but matched
    3 = convenience
    4 = pre-post no control
    5 = single subject/case study
    6 = crossover
    7 = other _____
               (specify)

2) Blinding (1 = yes, 2 = probably, 3 = no, 4 = can't tell)

(12) Subject

(13) Treatment Implementor

(14) Data Gatherer

VI THREATS TO VALIDITY

1. Threats to validity

(15) A) Maturation

(16) B) History

(17) C) Testing

(18) D) Instrumentation

(19) E) Statistical Regression

(20) F) Selection Bias

(21) G) Experimental Mortality

(22) H) Novelty and Disruption

(23) I) Experimenter Effect

(24) J) Inappropriate Statistical Procedures

(25) K) Other _____
              (specify)

(26) 2. General index of validity
    (1 = high ←→ 5 = low; code from conventions)

0 = not a plausible threat to the stud internal validity

1 = potential minor problem in attribu the observed effect to the treatme by itself, not likely to account fc substantial portion of observed re:

2 = plausible alternative explanation could account for substantial amou of the observed results

3 = plausible alternative explanation by itself could explain most or al the observed results

ES=1 _____

ES=2 _____

ES=3 _____

ES=4 _____

ES=5 _____

ES=6 _____

Outcome Used (Test Name)

ID#

	ES6	ES5	ES4	ES3	ES2	ES1
7-28)						
(29)						
0-31)						
(32)						
(33)						
(34)						
(35)						

VII OUTCOME

1) Type of measure

1 = General hyperactivity
2 = Cognitive performance
3 = Attention and vigilance
4 = Physiological
5 = Affective
6 = Activity Level
7 = Aggression
8 = Impulsivity
9 = Achievement
10 = Other _____
(specify)

2) Instrument

1 = Opinion
2 = Rating
3 = Systematic observation
4 = Actometer
5 = Standardized Test _____
(specify)
6 = Experimental Task
7 = Impossible to determine
8 = Composite
9 = Other _____
(specify)

3) Data Collector

1 = Teacher   7 = Clinician/Psychologist
2 = Parent   8 = Caretaker
3 = Physician   9 = Subject
4 = Researcher   10 = Composite
5 = Observer   11 = Other _____
6 = Counselor

4) Setting where measured

1 = School   6 = Institution
2 = Home   7 = Composite
3 = Clinic/Doctor's Office   8 = Impossible to determine
4 = Day Care   9 = Other _____
5 = Experimental Setting

5) Reliability (- = impossible to determine, 1=1.0 - .80, 2 = .80 - .60, 3 = 60-)

6) Treatment Implementor was Data Collector (1 = yes, 2 = no, - = impossible to determine)

7) Setting was 1 = structured, 2 = unstructured, 3 = mixed, - = impossible to determine

(6)

	ES6	ES5	ES4	ES3	ES2	ES1	
(35-36)							
(39-40)							
(41)							
(42-44)							
(45)							

VIII CONCLUSIONS

1. Effect Size

2. Data from which ES was calculated

    1 = means and control groups SD
    (code scale of means from list)
    2 = means and pooled SD
    (code scale of means from list)
    3 = means and published test SD
    (code scale of means from list)
    4 = t ratio/F ratio from one-way ANOVA
    5 = t ratio from matched pairs
    t, test or F ratio from repeated measures or other complex ANOVA design
    6 = non-parametric test statistic except Chi squared
    7 = probability estimate for t test
    or one-way ANOVA
    8 = S of V Table from n-way ANOVA
    9 = S of V Table from ANCOVA,
    repeated measures, or other complex ANOVA design
    10 = Regression lines
    11 = Proportions ("probit" transformations)
    12 = Chi square
    13 = Other _____
    (specify)

3. Scale of Mean Difference
   (if #2 coded 1, 2, or 3, code from 1-6 below)
   1 = final status measure
   2 = raw gain scores
   3 = residual gain scores
   4 = covariance adjusted scores
   5 = impossible to determine
   6 = other (specify) _____
   7 = if #2 was coded 4-13

4. Statistical Significance (code p value, - = not given)

5. Author's Conclusions

   0 = not considered
   1 = intervention appears to work
   2 = impossible to determine
   3 = intervention appears not to work

IX DRUG STUDIES

	ES1	ES2	ES3	ES4	ES5	ES6
(1-4)						
(5)						
(6,7)						
(8,9)						
(10)						
(11-14)						
(15-18)						
(19)						
(20-24)						
(25)						
(26-29)						
(30-33)						
(34)						
(35-39)						

Begin fourth (4th) card

Study #

Card #

1. Type of drug used for experimental group (code from list)

2. Type of drug used for control group (code from list, 0 = no drug; 50 = placebo)

3. Basis of dosage for experimental group (1 = varied by Ss weight; 2 = standard across Ss; 3 = clinical decision; - = can't tell)

4. Beginning daily average for experimental group (0 = varied; - = not given; dosage always in mg/kg, use decimal point with form --.-)

5. Ending daily average for experimental group (0 = varied; - = not given; dosage always in mg/kg, use decimal point with form --.-)

6. Number of times drug was administered/day for experimental group (0 = varied; - = not given)

7. Hours after drug ingested, dependent variable was measured for experimental group (- = not able to determine; "-9.0" = follow-up study; "-1.0" = data collected continuously; use form ---.-)

8. Basis of dosage for control group (1 = varied by Ss weight; 2 = standard across Ss; 3 = clinical decision; - = can't tell)

9. Beginning daily average for control group (0 = varied; - = not given; dosage always in mg/kg, use decimal point with form --.-)

10. Ending daily average for control group (0 = varied; - = not given; dosage always in mg/kg, use decimal point with form --.-)

11. Number of times drug was administered/day for control group (0 = varied; - = not given)

12. Hours after drug ingested, dependent variable was measured for control group (- = not able to determine; "-9.0" = follow-up study; "-1.0" = data collected continuously; use form ---.-)

DRUG TYPE

Major Tranquilizers

Phenothiazines

1 Chlorpromazine (Thorazine, Largactil)
2 Thioridazine (Mellaril)
3 Procholoperazine (Compazine, Stemetil)
4 Trifluoperazine (Stelazine)
5 Promazine (Sparine)
6 Perphenazine (Trilafon)
7 Triflupromazine (Vesprin)
8 Mepazine (Pacatal)
9 Acetophenazine (Tindal)
10 Fluphenazine (Prolixin, Permitil)
11 Promethazine (Phenergan)

Rauwolfia Alkaloids

12 Reserpine (Serpasil)

Chlorprothixene

13 Chlorprothixene (Taractan)

Minor Tranquilizers

Diphenylmethane Derivatives

14 Diphenhydramine (Benadryl)
15 Hydroxyzine (Atarax, Vistaril)
16 Captodiame, captodiamine (Suvren, Covatin)
17 Azacyclonol (Frenquel)

Substituted Propanediol Derivatives

18 Mephenesin (Tolserol)
19 Meprobamate (Miltown, Equanil, Trelmar)

Benzodiazepines

20 Chlordiazepoxide (Librium)
21 Diazepam (Valium)
22 Oxazepam
23 Nitrazepam

Butyrophenone Derivatives

24 Haloperidol
25 Fluperidol

Miscellaneous

26 Phenaglycodol (Ultran)
27 Benactyzine (Suavitil, Deprol)

Stimulants and Antidepressants

Amphetamines

28 Amphetamine (Benzedrine)
29 Dextro-amphetamine (Dexedrine)
30 Deanol (Deaner)
31 Methylphenidate (Ritalin)
32 Pipradrol (Meratran)
33 Ipronaizid (Marsilid)
34 Imipramine (Tofranil)
35 Nialamide (Niamid)
36 Phenelzine (Nardil)
37 Phenylisopropylhydrazine (Catron)

Compounds Part of Normal Metabolic Processes

38 Hormones
39 Vitamins
40 Glutamic Acid

Anticonvulsants

Miscellaneous

41 Neostigmine
42 Celastrus Paniculata Seeds
43 Pure Caffeine
44 Coffee

Biochemical

45 Desoxyribonucleic Acid (DNA)
46 Ribonucleic Acid (RNA)
47 Puromycin
48 Magnesium Pemoline
49 Other _____
   (specify)
50 Placebo

92

COMPUTATION OF EFFECT SIZES

_____
(Article ID =)

ES#1	ES#4
ES#2	ES#5
ES#3	ES#6

7

FUTURE NEED IN META-ANALYSES

1. ATTITUDES OF CHILDREN, PARENTS, TEACHERS TOWARDS DRUG MEDICATION

8. RELATIONSHIP BETWEEN HYPERACTIVITY AND LATER DELINQUENCY

2. SIDE EFFECTS OF DRUGS

9. DEFINITION OF PREVALENCE OF HYPERACTIVITY, DRUG USAGE, ETC.

3. CHANGES IN CNS (EVOKED POTENTIAL, ETC.) OR PHYSIOLOGICAL FUNCTIONS AS A RESULT OF DRUG USAGE

10. ASSESSMENT DEVICES AND INSTRUMENTS

4. DO DRUGS AFFECT HYPERACTIVITY DIFFERENTIALLY IN STRUCTURED (FORMAL) AND UNSTRUCTURED (INFORMAL) SITUATIONS?

11. EFFECT OF DRUGS ON MOTHER/CHILD AND/OR PEER INTERACTIONS

5. DO CHILDREN WHO HAVE BEEN MEDICATED FOR HYPER- ACTIVITY HAVE GREATER DRUG DEPENDENCY PROBLEMS LATER IN LIFE?

12. AGREEMENT BETWEEN OBSERVATIONAL DATA AND RATINGS; OR TEACHER RATINGS AND PARENT RATINGS

6. IS THERE EVIDENCE THAT CHILDREN "GROW OUT" OF HYPERACTIVITY AT PUBERTY?

13. PROBLEMS AND GUIDELINES FOR COLLECTING OBSERVATIONAL DATA ON HYPERACTIVE CHILDREN

7. DIFFERENCE IN THE RESPONSE OF HYPERACTIVE AND AGGRESSIVE CHILDREN TO DRUGS?

14. RELATIONSHIP OF HYPERACTIVITY TO THE AMOUNT OF DISTRACTIBILITY IN THE SETTING?

COMMENTS ON CODING CONVENTIONS

Article ID#

Notes on Clarification and Expansion of Conventions (note page nos. from article)

Notes on Disagreements with Conventions (note page nos. from article)

Appendix 3

Coding Conventions for Efficacy of Drug

Treatments for Hyperactivity

HYPERACTIVITY META-ANALYSIS CONVENTIONS

Contained in this document are the conventions or basic rules for coding the hyperactivity research articles. Additional examples of how these basic rules have been applied are contained in the conventions notebook. While coding articles, these rules should be used to make most decisions. If an item is impossible to code using these rules, the item should generally be coded "-". Occasionally however, educated guesses are possible. For example, if the study were done at a "boy's" reform school, item II-10 (% males) should be coded 100; and item II-13 (% delinquent) should be coded 100 even though this information was not specifically given. Another example would be if a parent completes a rating scale on the child's level of activity but the setting is not mentioned in this case #VII and (setting) should be coded "home." When guesses are made include a brief explanation on the "comments or conventions" page so the example can be incorporated into the conventions notebook. Guesses should be the exception rather than the rule and should only be made when you are confident about the accuracy.

I INTRODUCTION

1. Study ID# - taken from photocopy of article

2. Year - year of publication

3. Source - where published, coded from list of journals

4. Supported by commercial company

    1 = Article stated source(s) of support and the study was completely supported by a commercial company.

    2 = Article stated the source of support and the study was partially supported by a commercial company. Code 2 if a representative from a commercial company served as a consultant or if the company donated materials and/or equipment.

    3 = Article stated source(s) of support and it was not a commercial company.

    4 = Article did not state source(s) of support.

5. Dissertation

    1 = Article was based on a dissertation and the ES had to be estimated using procedures other than.

    0 = Article was a dissertation, or an article based on a dissertation where the ES did not have to be estimated, or an article not based on a dissertation.

6. Side Effect - Article reported side effects of treatment such as weight loss, growth suppression, insomnia, changes in heart rate, or other physiological or psychological functions not directly related to the manifestations of hyperactivity.

    1 = Yes

    2 = No

    - Specify side effects in blank next to item on coding sheet.

II DESCRIPTION OF SAMPLES[1, 2]

1. Mean Age

    - Report in months

    - If rounding is necessary, .5 or greater round up, below .5 round down.

    - If range only is reported, use midpoint as best estimate of mean, e.g., subjects were 9-10 years old, X = .5 years = 114 months.

2. Mean IQ

    1 = 130 or more
    2 = 71 - 129
    3 = 70 or less
    4 = not reported

    - If range only is reported, use midpoint as best estimate, e.g., subjects' IQs ranged from 90-110, X = 100.

    - Specify test used in blank next to item on coding sheet.

3. Size of Sample - Number of subjects at time data was analyzed.

4. Socioeconomic Status (SES) - Specify how SES was determined on coding sheet. Examples II: Low SES would be Title I recipients, or low income subjects. Middle SES would be blue collar, or lower management families, high SES would be children of university professors, doctors, or upper management. Code as 4=mixed if the group contains a mixture of SES. If article states that subjects were low middle or high without determining how it was determined, use authors statement.

---

[1] For all items in Section II, assume subject mortality is proportioned unless otherwise stated. In other words, compute the percentages in each group at the beginning and don't change the percentage as a result of subject mortality unless the article specifically states how many were lost from each group.

[2] If the article states that some of the sample belonged in categories 7-13 but does not specify the exact percentage, code the relevant category "-9" to indicate some.

5. Severity of Hyperactivity

    1 = None - none of the subjects were hyperactive.

    2 = Mild - all subjects in regular education classes.

    3 = Moderate - one or more subjects receiving special education services outside of the regular classroom but no more than half of subjects in self-contained classroom.

    4 = Severe - more than half of subjects in self-contained classroom.

    5 = Extreme - half or more of subjects institutionalized because of hyperactivity related concerns.

    6 = Unable to tell - article gives insufficient information for determining severity of hyperactivity.

6. Diagnosis - What the author(s) most often call the condition.

7. % Handicapped - Code the % of children in the sample in categories A-G below. If the sample has "some" MR children but doesn't say how many, code MR "-9". Use the same rule for other handicapping conditions.

    A. Multiple - Children having two or more handicaps.

    B. Deaf

    C. Blind

    D. MR - Mentally retarded. Subjects' IQs are below 70.

    E. Learning/Perception/Minor Neurological Problems - Children referred as LD (learning disability), MBD (minimal brain dysfunction), that exhibiting soft neurological signs (e.g., low scores on perceptual-motor tests, coordination problems, etc.) or EEG abnormalities the author(s) refer to as minor.

    F. Gross Neurological Problems - Obvious physical trauma, or EEG abnormalities. Count in this category children suffering from seizures and for convulsions.

    G. ED - Emotionally disturbed, children referred to as neurotic or psychotic.

8. % Institutionalized - Subjects are full-time residents of an institution. On this item subjects in an institution would be counted as institutionalized whether or not their institutionalization had anything to do with Hyperactivity.

9. % Minority

   - Include Black, Hispanic, Native American, and immigrant subjects.

   - Do not include american-oriental subjects.

10. % Male

11. Subjects on Active Drug Other Than the Drug Being Investigated

   - Subjects were on an active drug other than the drug investigated at a time that would have effected the ES for the drug being investigated. For example, if the design required using the time pretest/baseline measures in computing an ES and subjects were on an active drug during baseline, this should be coded.

   - Check drug list for time to become active/inactive in system.

12. Allergies - Author(s) state that subjects were allergic to some substance.

13. Delinquent - Author(s) state that children were delinquent or had been in trouble with the law.

III CLASSIFICATION OF CHILDREN AS HYPERACTIVE

This section describes the basis by which the child was classified as hyperactive. It should not be confused with other criteria used in selecting the sample such as IQ, EEG abnormalities, sex, age, etc. These other characteristics are coded in Section II. This section deals only with the information which was used to decode whether the child was hyperactive. For each group (experimental & control) code separately each source of information collected. For example, a study might have used a parent's opinion about general hyperactivity at home (this should be coded 1122 under "a.") and also a classroom observation of activity level collected by the researcher (this should be coded 2341 in b) and a test of impulsivity such as the Matching Familiar Forms Test (this should be coded 5644 in c). For both experimental and control groups code as many separate methods as were used to classify the children as hyperactive and list in a-d. If the "control" group was non-hyperactive the basis by which children were classified as hyperactive is irrelevant- therefore code "a" under control group as 8000. If article states only that children were referred by parents or teachers as being hyperactive, code basis = 1 (general hyperactivity) and instrument = 1 (opinion).

Basis for Selection - Code 4 digits indicating basis, instrument, evaluator, and setting.

   1 = General Hyperactivity - Basis for selection was a composite measure of hyperactivity in which characteristics such as activity, attention, agression, or impulsivity could not, or were not, separated. Or the article states that children were referred by parents because of problems with hyperactivity.

   2 = Activity - Index reflecting general motor activity.

   3 = Attention/Vigilance - An index which requires the ability to sustain attention/vigilance as a primary focus.

   4 = Aggression - Index reflecting a subjects tendencies or actual behaviors which are intended to destroy or cause injury.

   5 = Impulsivity - Any measure which yields an indication of whether a child makes decisions too rapidly, feels to pause to consider possible alternatives, feels to reflect on possible consequences of a decision, and/or seizes on the first response that comes to mind.

   6 = Other - Any other basis on which subjects are diagnosed as hyperactive. Specify basis on coding sheet.

Instrument

   1 = Opinion - Global impression.

   2 = Rating - Placement on a scale/rating form via opinion or recall.

   3 = Systematic Observation - Systematic recording of one or more aspects of the child's behavior, e.g., frequency, intensity, duration, etc.

4 = Actometer - Any instrument used to automatically record movement.

5 = Perceptual-motor Test - Standardized test designed to measure a child's ability to coordinate sensory information and movement.

6 = Experimental Task - Any task designed specifically to serve as an index of hyperactivity or some aspect of it. Specify task on coding sheet.

7 = Impossible to Determine - Cannot be determined from article what instrument(s) were used to assess hyperactivity.

8 = Other - any other instrument used to assess hyperactivity. Specify instrument on coding sheet.

Evaluator

1 = Teacher - Person with primary responsibility for providing the child with instruction.

2 = Parent - or legal guardian.

3 = Physician - MD.

4 = Researcher - Any person instrumental in the conceptualization and/or design of the study. Code physicians, psychologists, teachers, etc. who were on the research team in this category.

5 = Observer - Anyone not listed in another category making systematic observations of the subjects' behavior.

6 = Clinician/Psychologist - Any person whose background is primarily psychology and whose role is consistent with such background.

7 = Caretaker - Any person whose role within a residential institution gives him/her primary responsibility for the care of the child.

8 = Impossible to Determine - Article does not state who evaluator was.

9 = Other - Any other person who evaluated the children to determine if they were hyperactive. Specify person on coding sheet.

Setting (Code setting where observations of child were made)

1 = School - Classroom or other area of school building.

2 = Home - Residence other than institution.

3 = Clinic/Doctor's office - Any medical facility.

4 = Experimental Setting - Any non-naturally occurring setting established specifically as an area/situation in which to observe/record hyperactivity.

5 = Day Care - Place other than educational facility or home where child is cared for during the day.

6 = Institution = Any residential facility for the treatment of handicapped or disturbed persons.

7 = Composite = A combination of settings where hyperactivity was assessed and it is impossible to separate the various settings.

8 = Impossible to Determine - Article does not state in what setting the children were assessed as hyperactive.

9 = Other - Any other setting in which children were assessed as hyperactive. Specify setting on coding sheet.

General Quality of Procedures Used to classify Children as Hyperactive

1 = Good: objective measures by 2 or more people or in 2 or more settings.

2 = Fair: objective measures used by only 1 person in 1 setting--may or may not have used additional subjective measures.

3 = Poor: only subjective measures used.

4 = Unable to tell: in those cases where it is impossible to determine from the article how children were classified as hyperactive, in coding the general quality of the procedures used to classify children as hyperactive opinions and ratings wil' generally be considered subjective measures and systematic observatin, actometers, other electro/mechanical recording devices, and perceptual motor and other experimental tasks will generally be considered objective. However, if a rating scale were done by blind observers and included good criteria, it would qualify as an objective measure. Or if an observation was done by non-blind observers and/or used vague criteria, it would be a subjecative measure. Using similuar rationale, other exceptions may be used. Be sure and not use exceptions for conventions book of examples.

IV TREATMENT

1. Treatment Type

   2 = This was the only treatment given the subjects, subjects may have received other treatments consecutively but not concurrently.

   1 = This was one of several concurrent treatments given the subjects.

   0 = This was not a treatment given the subjects.

   - Drug - Any treatment where a nonfood substance was administered to the subjects.

   - Behavioral - Any systematic manipulation of the subjects' environment or reward/punishment system.

   - Diet - Any treatment where a food substance was added to or deleted from the subjects' diet.

   - Biofeedback - Any treatment where the subjects are given feedback on some parameter of their physiology.

   - Comparison of Groups - Any study where the behavior of hyperactive children was compared to that of another group of children in the same environment but no treatment was involved.

2. Control Group

   A. Selection - How comparison group was selected.

   1 = Random - Randomly assigned to experimental and control groups from same initial group.

   2 = Convenience - Selected simply because subjects were available.

   3 = Matched - Selected because they matched experimental group on some parameter.

   4 = Own - Experimental group was observed pre-post treatment or under two conditions and served as its own control.

   5 = Crossover - Control group was formed by having a crossover design in which all subjects received both the experimental and the control (or placebo) conditions.

   6 = Impossible to Determine - Article does not state how the control group was selected.

   7 = Children in the "control" group were non-hyperactive.

   8 = Other - Any other basis on which the control group was selected. Specify basis on coding sheet.

   B. Treatment - Condition under which control group was observed.

   1 = None - Received no treatment.

   2 = Placebo - Received a placebo treatment.

   - Quality of Placebo

   4 = Present, Can't Tell - Placebo was given but article contained insufficient information for determining its quality.

   3 = Good - Precautions were taken to assure that the placebo could not be distinguished from treatment. For example, pills should be the same size, shape, color, and taste to qualify as a good placebo.

   2 = Poor - Placebo could be distinguished from the treatment. For example, pills should be the same size, shape, color, and taste to qualify as a good placebo.

   0 = Absent

   3 = Drug - Received a drug treatment, specify on coding sheet.

   4 = Behavioral - Received a behavioral treatment, specify on coding Sheet.

   5 = Diet - Received a diet treatment, specify on coding sheet.

   6 = Biofeedback - Received a biofeedback treatment, specify on coding sheet.

   7 = Impossible to Determine - Article does not specify treatment received by control group.

   8 = Other - Received any other treatment, specify on coding sheet. Specify treatment on coding sheet.

3. Duration of Treatment in Days - Number of Days from first day of treatment to last, include the minimum number of weekends that could have occurred in the time span. In cases where there is no treatment (e.g., comparison of groups) code this -1.

4. Days After Treatment the Dependent Variable was Observed - Time from end of treatment to when outcome measure was made. Code as "0" if Dependent Variable was measured when day treatment was still being administered. Code as -1 if no treatment was given.

5. Reliability of Treatment Implementation - Confidence one can have that treatment was implemented as described. The article may present data on the reliability of treatment implementation, for example, state that two observers watched a teacher implement a behavior program and that she was 90% reliable in implementing it. Alternatively, the article may present information that allows an estimate of the reliability of treatment

implementation, for example, a drug study might state that the parents of all children were given 30 pills to administer during the treatment period and that at the end, 15% of the pills were returned from which it could be estimated that the reliability of treatment implementation was 85%. If no data or basis for an estimate is given, reliability of treatment implementation may be estimated using the convention in 1-5 below.

1 = Complete - Reliability data is presented or can be estimated and is .95 or better.

2 = Minor - Reliability data is presented or can be estimated and is .80-.94, or no reliability data is given but treatment was implemented by a professional, teacher, researcher, physician, etc.

3 = Moderate - Reliability data is presented or can be estimated and is .70-.79 or no reliability data is presented but treatment was implemented by parents or paraprofessionals.

4 = Major - Reliability data is presented or can be estimated and is .69 or less.

5 = Impossible to Determine - No reliability data is presented and it is not clear from article who implemented treatment.

6 = No treatment given - In situations where groups are being compared (e.g., hyperactive children vs. non-hyperactive children) there is no treatment and consequently, no reliability of treatment implementation.

6. Confidence with which #5 was Coded

1 = Data Based - Article gives data-based information about how completely the treatment was implemented.

2 = Guess Based on Data - Article gives information from which a coefficient describing the reliability of treatment implementation can be estimated.

3 = Convention - The reliability of treatment implementation was estimated based on who implemented the treatment (see #5).

4 = Impossible to Determine - #5 was coded 5 because the article gave was insufficient information regarding how treatment implementation given.

5 = Not applicable - This item would be not applicable in cases where groups are being compared or where #IV-5 was left blank.

1. Type

1 = Random Assignment - Subjects are randomly assigned to groups.

2 = Non-Random but Matched - Not randomly assigned to groups but control subjects were matched to experimental subjects on some parameter.

3 = Convenience - Basis for selecting subjects was that they were available.

4 = Pre-post, No Control - Experimental group was observed under two or more conditions and no crossover design was employed.

5 = Single Subject - Data is presented as graphic display for individual subjects.

6 = Other - Any design. Specify design on coding sheet.

2. Blinding

1 = Yes - Individual definitely blind.

2 = Probably - Individual was not told the purpose of the study and/or what subjects were under what conditions but possibly could have figured it out.

3 = No - Individual definitely was not blind.

4 = Can't tell - Article did not give enough information to determine if individual was blind

- Subject - Individual(s) for whom treatment was implemented.

- Treatment Implementor - Individual(s) who implemented treatment.

- Data Gatherer - Individual(s) who collect data about Dependent Variable.

VI. THREATS TO VALIDITY

General Convention: Each of the "threats," listed below should be coded using the following conventions. Definitions and examples of the "threats" follow the general convention. Two are contained in conventions notebook.

0 = Not plausible threat to internal validity.

1 = Potential minor problem in attributing the observed effect to treatment; by itself, not likely to account for substantial amount of the observed results.

2 = Very plausible alternative explanation which could account for substantial amount of the observed results.

3 = Very plausible alternative explanation which by itself could explain most or all of the observed results.

Maturation - Definition - Biological, physiological, or psychological "processes within the respondents varying systematically with the passage of time" but not as the result of specific events (including the experimental treatment) external to the respondents, e.g., growing older, more tired, better coordinated, etc. Suppose an experimenter claimed that a series of prescribed play activities were effective in promoting bladder control in infants; as evidence he showed that 2% of the 15-month old infants starting his experiment had control, and 75% of these infants achieved control 9 months later. His claim is questionable since the normal infant develops bladder control during this period naturally.

2. Testing

The effects of taking a test on the outcomes of subsequent administration of the same or a highly related test. Taking some cognitive-ability tests may increase your score by several points on a second administration of the same test or a parallel form of it. For example this would be a threat if children were tested repeatedly with the same test instrument and no control group was included in the design.

3. Instrumentation

Changes in the instruments (tests, judges, experiment are observed may produce changes in the scores over time which are mistaken as treatment effects. For example, judges observing and rating some performance may be more lenient from time 1 to time 2.

4. Statistical Regression - The inevitable tendency of persons whose scores are extreme (high above or far below the mean) on Measurement A to be less extreme (less high above or less far below the mean) on Measurement B. This phenomenon of "regression toward the "mean" will be observed whenever Measurements A and B are not perfectly correlated, which for all practical purposes is always. For example, this will be a threat if children in the experimental group were selected on the basis of an extreme score which was used simultaneously as a pretest and there was not a control group or the control group was not selected on the basis of the same extreme scores.

5. Selection Bias - Children in the experimental and control group were selected on different bases. Definition - All of those factors which conspire to make the experimental and the control groups unequal at the outset of an experiment in ways which cannot be properly taken into account in the analysis of the data. For example, selection might invalidate a comparison of curricula A and B if older, more experienced teachers were selected to teach the more difficult curriculum. It appears that in almost all instances the only feasible way to completely guard against selection bias is by employing the random assignment of persons or classrooms to treatments and then using statistical analyses of the final data which are based on the randomization procedure. Quasi experimental designs will almost always have some selection bias. Designs in which subjects serve as their own controls (pre-post, crossover) will usually not have selection bias, but often have other problems.

6. Experimental Mortality - The differential loss or "dropping out" of persons from two or more groups being compared in an experiment. If attrition is greater under curriculum A than curriculum B, a comparison of A and B at the end of one school year might be biased in that the students completing A would be brighter--on the average--than those completing B. This is true simply because the slower students were fatalities under curriculum A.

7. Novelty and Disruption - Measurement of the children's behavior was made in an environment that was new to them and it is plausible that the newness of the environment was responsible for different scores and no control group was included in the design of the study.

8. Experimenter Effect - Attitudes of experimenter regarding expected research results are known to treatment implementor, data collector, or subject.

2. General Index of Validity.

GENERAL INDEX OF VALIDITY

RATINGS: 1	2	3	4	5
• well executed true experimental designs • well executed double blind crossover designs with order effects balanced and sufficient time for previous drugs to become inactive	• true experimental designs with minor problems (1-3 "1" ratings) • well executed quasi experimental designs (no "1" except for selection • well executed single subject • crossover designs with minor problems Only "1" ratings, no more than 3 points	• quasi experimental designs with minor problems (1-3 "1" ratings or 1 "2" rating) • well executed pre post designs (no "1" besides selection, maturation, history • single subject with minor problems • true experimental with moderate problems (2-4 "1" ratings or 1-3 "2" ratings) Only "1" or "2" ratings, no more than 6 points	• pre post designs with minor to moderate problems (2-4 "1" ratings or 1-2 "2" ratings) • quasi experimental with moderate problems (6 or more points, with at least 2 "2" ratings) • true experimental with major problems • single subject with moderate problems	• any design with one or more "3" ratings • pre post designs with major problems (6+ points with at least 2 "2" ratings) • single subject/case studies with major problems

VII OUTCOME

1. Type of Measure

    1 = General - Index reflecting general hyperactive behavior.

    2 = Cognitive Performance - Any general or specific measure of cognitive ability such as might be obtained through an IQ test or one of Piagets conservation problems.

    3 = Attention/Vigilance - Index reflecting ability to sustain attention/vigilance.

    4 = Physiological - Any measure of a physiological parameter.

    5 = Affective - Any general or specific measure of the perception of oneself or others, or the ability to relate to others such as might be obtained by a self-concept or sociometric test respectively.

    6 = Activity - Index reflecting general motor activity.

    7 = Aggression - Index reflecting behavior(s) defined as indicative of aggression.

    8 = Impulsivity - Index reflecting behavior(s) defined as indicative of impulsivity.

    9 = Achievement - Index reflecting learning in one or more areas.

    10 = Other - Any other outcome measure. Specify measure on coding sheet.

2. Instrument

    1 = Opinion = Global impression.

    2 = Rating - Placement on a scale via opinion.

    3 = Systematic Observation - Systematic recording of one or more aspects of the child's behavior, e.g., frequency, intensity, duration, etc.

    4 = Actometer - Any instrument used to automatically record behavior.

    5 = Standardized Test - Test for which norms have been collected. Specify test on coding sheet.

    6 = Experimental Task - Any task designed specifically to serve as an index of hyperactivity or some aspect of it. Specify task on coding sheet.

    7 = Impossible to Determine - Article does not state what instrument was used to measure outcome.

    8 = Composite - Any combination of instruments used to measure outcome and yielding one score.

    9 = Other - Any other instrument used to measure outcome.

3. Data Collector

    1 = Teacher - Persons with primary responsibility for providing the child with instruction.

    2 = Parent - or legal guardian.

    3 = Physician - MD

    4 = Researcher - Any person instrumental in the conceptualization and/or design of the study.

    5 = Observer - Anyone not listed in another category making systematic observations of the subjects' behavior.

    6 = Counselor - Any person whose background is primarily counseling and guidance and whose role within a school or other institution is consistent with such background.

    7 = Clinician/Psychologist - Any person whose background is primarily psychology and whose role is consistent with such background.

    8 = Caretaker - Any person whose role within a residential institution gives him/her primary responsibility for the care of the child.

    9 = Subject - Individual receiving treatment or being used for comparison to determine treatment effects.

    10 = Composite - Any combination of data collectors whose measurements yield one score.

    11 = Other - Any other person who evaluated the children to determine if they were hyperactive. Specify person on coding sheet.

4. Setting (Code setting where observations of child were made)

    1 = School - Classroom or other area of school building.

    2 = Home - Residence other than institution.

    3 = Clinic/Doctor's office - Any medical facility.

    4 = Day Care - Place other than educational or home where child is cared for during the day.

111

112

94

5 = Experimental Setting - Any setting established specifically as an area/situation in which to observe/record hyperactivity.

6 = Institution = Any residential facility.

7 = Composite ='Any combination of settings where hyperactivity was assessed.

8 = Impossible to Determine - Article does not state in what setting the children were assessed as hyperactive.

9 = Other - Any other setting in which children were assessed as hyperactive. Specify setting on coding sheet.

5. Reliability of Outcome Instrument

    0 = not given
    1 = 1.0 - .80
    2 = .80 - .60
    3 = .60 or below

    In cases where the reliability of the instrument is not specified in the article, but is known from the published norms for subjects which are similar to the subjects in the study, use the published reliability. If published reliability and reliability reported in study conflict, use the reliability reported in the study, unless there is some reason to doubt it.

6. Treatment Implementor was Data Collector

    1 = Yes

    2 = No

    3 = Impossible to determine

7. Setting where Outcome Measure was Taken was

    1 = Structured - Any setting where requirements are made on the child, for example, a classroom where he/she must stay in his/her seat and do assigned work.

    2 = Unstructured - Any setting where no requirements are made on the child, for example, a playground where the child can do anything he wants.

    3 = Mixed - Code 3 when the article states that outcome was measured in both structured and unstructured settings and measurements were combined into one index.

    4 = Impossible to determine - Article does not state in what setting outcome measures were made.

VIII CONCLUSION

1. Effect Size (ES) = $\dfrac{\bar{X}_e - \bar{X}_c}{SD_c}$ or best estimate. Express as positive or negative relative to the desired effect of treatment. For example, if the treatment decreases hyperactivity, calculations from the formula may leave a negative effect size which should, however, be expressed positively because the desired effect of treatment is to decrease hyperactivity. Alternatively, if the treatment increases drowsiness, calculations from the formula may leave a positive effect size which should, however, be expressed negatively because this is not a desirable treatment effect. Show formula used and calculation on coding sheet.

    In calculating effect sizes when $\bar{X}$'s and SD's are not given, the estimates must sometimes be made. The following conventions have been adopted for some of the most frequently required estimates.

    Correlations: Between two standardized psycho-educational tests
    between two rating scales

    Reliabilities: Rating scales - .60

2. Data ES was Calculated From

    1 = Means and control group SD - Article gave means for the experimental and control groups and a standard deviation for the control group from which ES was calculated.

    2 = Means and pooled SD - Article gave means for the experimental and control groups and a pooled standard deviation from which the ES was calculated.

    3 = Means and published test SD - Article gave means for the experimental and control groups and the standard deviation was known for the published test used as an outcome measure. ES was calculated from these data.

    4 = t ratio/F ratio from one-way ANOVA - Article gave a t or F for one way ANOVA value from which ES was calculated.

    5 = t ratio from matched pairs t test or F ratio from repeated measures or other complex ANOVA design.

    6 = Non parametric test statistic except chi squared.

    7 = Probability estimate for t test or one-way ANOVA - Article gave a p-value from which a t or F from a one-way ANOVA was calculated and ES was calculated using these estimates.

    8 = Source of variance estimate for n-way ANOVA - Article gave a source of variance table for n-way ANOVA from which ES was calculated.

    9 = Source of variance table from ANCOVA, repeated measures, etc.

ERIC

10 = Regression lines

11 = Proportions

12 = Chi square

13 = Other - Any other basis on which ES was calculated. Specify basis on coding sheet.

3. Scale of Means Difference

1 = Final status measure - Raw or standard scores were used to calculate means.

2 = Raw gain score - Difference between pretest and posttest scores were used to calculate means.

3 = Residual gain score = Pretest and posttest scores were correlated, the correlation was used to predict posttest score from pretest score, and the difference between the predicted and the obtained posttest scores were used to calculate means.

4 = Covariance adjusted scores - Outcome scores were correlated with scores on a covariate and adjusted to represent the outcome scores that would have been obtained if all subjects had obtained the same score on the covariate and used to calculate means.

4. Statistical Significance

3 digit p value = provide p value for test of statistical significance. If intervention reduced hyperactive condition or associated variable, p value should be low (e.g., .010, .030, .070). If intervention increased hyperactive condition, p value should be high (e.g., .940, .980, .990). In latter case, the obtained p value from data showing a mean difference favoring the control group will be subtracted from 1.00.

- = Not given - Article does not present information on statistical significance of treatment.

5. Author's Conclusion

0 = not considered - author(s) make no statement regarding clinical significance of treatment.

1 = intervention appears to work - author(s) conclude that treatment works. Those cases where the author concludes that the intervention works but only for certain subsets will usually be accounted for by the different ES categories. If this does not account for it, code it "1" anyway.

2 = Impossible to determine - author(s conclude that effect of treatment can't be evaluated for some reason.

3.= intervention doesn't work - author(s) conclude that treatment doesn't work.

Appendix 4

Procedures for Contacting Authors

for Additional Information

REQUEST FOR ADDITIONAL INFORMATION

Objectives: 1) Where it is impossible to calculate an ES from the journal article, TO OBTAIN INFORMATION NECESSARY TO CALCULATE AN ES.

2) Where it is possible to calculate an ES but the ES must be estimated using "cookbook" procedures, TO OBTAIN INFORMATION ON MEANS AND STANDARD DEVIATIONS SO THAT THE ADEQUACY OF "COOKBOOK" PROCEDURES CAN BE EXAMINED.

Procedures: After reading the article, if either of the above objectives are relevant, fill out the information on the attached form and give the form to Marilyn. The number of dependent variables listed on the form will be determined by the number of ES you can differentiate from the article. In other words, an article may describe a study which collected data and reported F ratios for three dependent variables for both brain injury and non-brain injured hyperactive children. From this article, you would probably want the following six ES's.

1) Disruptive behavior - brain injured children
2) California Achievement Test - reading subtest - brain injured children
3) California Achievement Test - math subtest - brain injured children
4) Disruptive behavior - non-brain injured children
5) California Achievement Test - reading subtest - non-brain injured children
6) California Achievement Test - math subtest - non-brain injured children

Do not request information for subgroups within the study that the author did not consider in the original article. In the above example, the three dependent variables should not be broken down by brain injured and non-brain injured unless the author in the original article considered differences between brain injured and non-brain injured children.

Finding Addresses: If address is provided in article and publication is since January, 1978, use it. If address provided, but article before January, 1978, check procedures noted below and use most recent address. If no better address is available, use the one in the article. If no address is noted in the article, check recent APA Directory, AERA directory, or Directory to Faculty of American Colleges and Universities. If none of these work, give up.

Date   ( ) Data no longer available
             ( ) I would rather not respond

①

Dear _____②_____

    I am directing a project funded by the National Institute of Education to integrate the previous research which has examined the effectiveness of various types of intervention for hyperactivity. We have obtained your article _____③_____ which appeared in _____④_____.
                              (was presented at)
The article discusses some interesting research which we would like to include in our integration effort. However, we did not find all of the information we needed. Would you be willing to send additional information regarding the means ($\bar{X}$) and standard deviations (SD) for the following dependent variables reported in your study?

⑤   **Dependent Variables** $\bar{X}$ SD

1. . . . .
2. . . . .
3. . . . .
.
.
n

    Enclosed is an addressed envelop which requires no postage. If it would be easier for you, the information could be filled out directly on this form and returned without any need of a separate cover letter. If the relevant data are no longer available, or if you would rather not respond to this request, please check the appropriate box in the upper right hand corner of this letter and return it so that we will not bother you with follow-up requests.
⑥
    Thank you for your consideration of this request. I look forward to hearing from you in the near future.

                         Sincerely,

                         Karl R. White, Ph.D.
                         Director, Planning & Evaluation

①

Dear _____②_____

    I am directing a project funded by the National Institute of Education to integrate the previous research which has examined the effectiveness of various interventions for hyperactive children. In our search we came across a reference to your article _____3_____ which was (presented) at _____4_____. (written)
I would like to obtain a copy of that article. If it has subsequently been published in complete form, could you provide me with the reference? If not, would it be possible for you to send me a copy? I would be happy to reimburse you for the costs of reproduction and mailing.

    If the article is not available, I would appreciate it if you would return this letter with a note to that effect so that I will not trouble you with follow-up requests. Thank you for your consideration of this request.

                    Sincerely,

                    Karl R. White, Ph.D.
                    Director, Planning & Evaluation

Date

Dear _____ ② _____

Recently I wrote to you requesting some additional information about the research you reported in an article entitled _____ ③ _____ . Attached is a copy of that letter in case the original went astray in the postal system. If you have been too busy to respond, or the first letter arrived at an inconvenient time--I understand, since I have frequently been in similar situations myself.

The additional information which we requested from you is very important to the success of our project. Although you are only one of dozens of researchers from whom we have requested information, every bit of information contributes valuable information to the total picture. I would appreciate it if you could take a few minutes in the near future to respond. If you have some of the information readily available, but have delayed trying to get additional information which is more difficult to access, it may be better to send what you have now, with the remainder to follow if and when you can get it.

Thank you for your consideration of this request.

Sincerely,

Karl R. White, Ph.D.
Director, Planning & Evaluation