Referral Rates for Intervention or Assessment:
A Meta-Analysis of Racial Differences

John L. Hosp, *University of Utah*
Daniel J. Reschly, *Vanderbilt University*

This study synthesized the literature on the rates of referral for intervention or assessment of students from three racial groups: Caucasian, African American, and Hispanic. Ten studies, published or unpublished, that presented frequency counts for the population and the referred sample for at least one school district yielded 44 comparisons. For comparisons between African American and Caucasian students, the mean risk ratios comparing the referral rates were significantly different from zero. No significant differences were found between the referral rates of Hispanic students and Caucasian students. Implications, limitations, and directions for future research are presented.

Determining whether students of different racial or ethnic groups are disproportionately identified for special education services is an issue that has been debated for more than 30 years (see Heller, Holtzman, & Messick, 1982; Reschly, 1997). Research on this issue has provided valuable data documenting the existence and details of the problem, but in general this research has not identified specific strategies to adequately address the issue. In order to develop strategies and materials to reverse and prevent disproportionate representation, however, all stages in the process of special education eligibility must be examined in order to determine when and where disproportionate representation occurs. The methods of determining disproportionate representation in special education programs must be examined and extended to the stage of referral, which is often considered the most important step in the special education eligibility process (Ysseldyke & Algozzine, 1983).

Disproportionate Representation in Special Education

In 1968, Lloyd Dunn published an article noting the overrepresentation of minority students and students from low-SES backgrounds in special day classes for mild mental retardation (MMR). This was one of the earliest and most well-known articles citing disproportionate representation in special education. It is often cited as a prophetic publication, one that occurred at the right time with the right message to serve as a catalyst for opening a proverbial floodgate, triggering much debate and additional research. In the article, Dunn (1968) stated that 60% to 80% of students in special day classes for MMR were from "low-status backgrounds" such as minority ethnic groups, nonstandard-English–speaking group, and non–middle class environments.

A few years later, Jane Mercer (1973) documented disproportionate representation of minority students in special education classes in Riverside, California, using data collected in the early to mid-1960s. She determined that the percentages of different groups placed in mental retardation (MR) classes were discrepant. The percentage of Mexican American students placed in MR classes was four times their percentage in the population. For African American students, that figure was three times their percentage in the population.

Jeremy Finn (1982) was the first researcher to examine national disproportionate representation trends by using the data from the *Elementary and Secondary School Civil Rights Compliance Report* from the Office for Civil Rights (OCR; U.S. Department of Education, Office of Civil Rights, 1994). He found that African American students were overrepresented in classes for MR and emotional disturbance (ED), Native American students were overrepresented in learning disability (LD) classes, and Asian American students were underrepresented in almost every category. Contrary to Mercer's (1973) findings, Hispanic and Caucasian students were found to be identified at similar rates. This discrepancy in results may have been prompted by the *Diana v. State Board of Education* (1970) and *Guadalupe Organization v. Tempe Elementary School District No. 3* (1972) court decisions regarding the use of nonverbal intelligence tests in the determination of eligibility.

Since Finn, use of the OCR *Compliance Report* to examine disproportionate representation has been common. Chinn and Hughes (1987) compared the results from the 1978, 1980, 1982, and 1984 OCR surveys. They found that the trends documented by Finn continued through the 1980 to 1984 surveys: African American students were overrepresented in the MR

Address: John L. Hosp, Dept. of Special Education, University of Utah, 1705 E. Campus Center Dr., Rm. 213, Salt Lake City, UT 84112; e-mail: john.hosp@ed.utah.edu

and ED categories, Native American students were overrepresented in the LD category, and there was widespread underrepresentation of Asian American students. An additional comparison that Chinn and Hughes made was that of representation in classes for students with gifts and talents (GT). In this category, Asian American students were overrepresented, whereas African American, Hispanic, and Native American students were underrepresented.

The national patterns noted by Finn (1982) and Chinn and Hughes (1987) have been robust and steady over time. More recent OCR surveys in 1986 (Reschly & Wilson, 1990), 1990 (MacMillan & Reschly, 1998; Reschly, 1997), and 1992 (Oswald, Coutinho, Best, & Singh, 1999) have been used to verify these patterns. In each updated OCR survey, African American students were overrepresented in the MR and ED classifications, Native American students were overrepresented in the LD category, and Asian American students were underrepresented in almost every category, with speech–language impairment (SLI) generally being the exception (Chinn & Hughes, 1987).

Why Disproportionate Representation Is Problematic

Disproportionate representation is problematic if the program in question is ineffective or stigmatizing or if the process of identification and placement is not applied equally to different groups of students (Heller et al., 1982). MacMillan and Reschly (1998) pointed out that minority representation in programs such as Head Start, Follow Through, and Title I is similarly disproportionate to representation in special education, yet it has not been criticized as much. In addition, minority representation in GT programs has been criticized (Chinn & Hughes, 1987; Harry & Anderson, 1994; Patton, 1997), but for *under*representation, not overrepresentation. This is most likely due to the view that GT programs are enriching and generally positive experiences. So why is disproportionate representation in special education criticized so severely? There are three main reasons: *labeling effects, segregation of placement,* and *presumed ineffectiveness of special education.*

The influence of negative labels is believed to have deleterious effects on the treatment of, placement of, and outcomes for students (Patton, 1998). Mehan, Hertweck, and Miehls (1986) found that teachers focus on the negative behaviors of students who are considered to have behavior problems, even if the behaviors are not significantly different from those of other students in the same classroom. In the Mehan et al. study, students were divided into referred and nonreferred groups. Every student in each class exhibited behaviors that the teachers, when asked to watch a video recording of their classrooms, had identified as negative. Interestingly, the teachers identified 46.5% of the negative behaviors of the referred students but only 13.9% of the negative behaviors of the nonreferred group. The authors concluded that the teachers were attempting to confirm their decisions regarding referral (i.e., their perceptions of the student).

The second problematic area is placement. The least restrictive environment (LRE) mandate of the Individuals with Disabilities Education Act (IDEA) states that students with disabilities must be educated with nondisabled children "to the maximum extent appropriate" ("Assistance to States for the Education of Children with Disabilities," 1999, p. 12457) and that they may be removed from the general education environment only if they cannot be satisfactorily educated with the use of supplementary aids and services. Placements that allow maximum access to the general education curriculum and contact with nondisabled peers (i.e., the general education classroom) are preferred unless the IEP team decides that an alternative placement is required. Serwatka, Deering, and Grant (1995) noted that African American students are more frequently placed in more segregated settings than are Caucasian students. This is prominent in higher incidence categories such as MMR (Reschly, 1988). Disproportionate placement of African American students in more segregated settings also occurs in lower incidence categories such as physical impairment and visual impairment (Serwatka, Dove, & Hodge, 1986), even though representation rates are not disproportionate in these categories.

Many reasons are given for this disproportionate placement: lack of a consistent identification process, bias in assessment instruments, overrepresentation of African American students in categories (i.e., MR and ED) that have been traditionally served in more segregated settings, and problems related to other variables such as poverty (Serwatka et al., 1995). The most commonly cited factor, however, is cultural differences. Many researchers have argued that there is a specific African American behavioral style (Hale-Benson, 1986; Hilliard, 1992). Today's U.S. teachers are primarily Caucasian (90%) and female (66%; U. S. Department of Education, Office of Educational Research and Improvement, 1998), which can generate conflict between the teacher's cultural behavioral expectations and non-Caucasian male students' behaviors. Because teacher tolerance is the prime indicator for identification of behavior problems (Gerber & Semmel, 1984) and because teachers are generally less understanding (i.e., tolerant) of behaviors that are not part of their cultural experience and expectations (Lambert, Puig, Rowan, Lyubansky, & Winfrey, 1998; McIntyre, 1992), students who come from a culture other than the teacher's are at greater risk for being identified as having a behavior problem and thus being placed outside the general education classroom.

A third reason disproportionate representation in special education is thought to be problematic is because special education is often presumed to be ineffective (Dunn, 1968; Harry & Anderson, 1994). Research into the effectiveness of special education has been inconclusive and occasionally contradictory (Carlberg & Kavale, 1980; Wang & Baker, 1985–1986; both as cited in Artiles & Trent, 1994). There is also evidence

that different programs and placements may be differentially effective (Kavale & Forness, 1999).

Methods for Determining Disproportionate Representation

Although there is little disagreement that disproportionate representation can be problematic, several methods have been used to demonstrate differential representation of various groups in special education. Determining the extent of disproportionate representation is important to establishing how large a problem the issue is (or isn't). The first approach used, in Dunn's (1968) classic article, was his "best judgment." Dunn's figure of 60% to 80% of special day-class students being from minority or low-SES backgrounds was deduced from his years of experience and observation. One person's judgment (even if supported by other people's judgment) is not a reliable method for estimating the scope of a problem. This is especially true when the stakes, such as funding and civil rights issues, are as high as in the area of disproportionate representation.

Another approach that has been used is comparing the percentage of students from different racial groups within a certain category or placement. This has been termed *percent of category by group,* or the *composition index* (MacMillan & Reschly, 1998; Reschly, 1997). Chinn and Hughes (1987) and Mercer (1973) used this method. For example, Chinn and Hughes reported that in 1978, 45.3% of the MMR population was African American, whereas only 20.1% of the total population was. Although this method appears to present the figures in a straightforward manner, it has been criticized as actually inflating the appearance of disproportion. Reschly (1997) cited the example that 85% of elementary teachers are women, yet the overwhelming majority of women are not elementary teachers. In other words, even though the composition index is high, the percentage of women in the group is low and may provide a more accurate picture of the issue.

Percentage of group in a category or placement, also often termed the *risk index,* is another method for presenting overrepresentation results (Reschly, 1997). It is calculated by dividing the number of students of a certain group (e.g., African American) in a certain category or placement (e.g., LD) by the total number of students in that group. An example of this approach can be found in MacMillan and Reschly (1998). In 1990, according to the OCR *Compliance Report,* nearly 35% of students classified as MMR were African American, but only 2% of African American students were classified as MMR. Obviously, different methods of calculation provide different pictures of the severity of the same issue (see MacMillan & Reschly for a more thorough comparison of the composition index and the risk index).

Rather than rely solely on the risk index, some researchers have explored other approaches for comparing groups. The first alternate approach was used by Finn (1982), who calculated the odds of a student of a certain group (e.g., African American) being identified in a certain category (e.g., MR). In other words, Finn took the number of students of that group in that category and divided that number by the number of students in that group *not* in that category. The resulting number is then divided by the odds of students of all other groups being identified for the same category. This measure is one form of a statistic known as the *odds ratio* (Finn, 1982; Lipsey & Wilson, 2000). Although this form of odds ratio provides a consistent measure of the likelihood of a certain event, it fails to include a measure of the risk index. It also provides only a binary comparison of one group (i.e., members vs. nonmembers), rather than a direct comparison of groups.

In order to remedy these shortcomings of the odds ratio, other researchers (Fleiss, 1994; Kleinbaum, Kupper, Muller, & Nizam, 1998) have proposed a method known as the *rate ratio,* which is also referred to as *relative risk.* Relative risk compares the risk index for one group to the risk index for another group (or the total population; see Coutinho & Oswald, 1998, for an example). In addition to addressing the problems of the risk index and direct comparison of groups, current advances in meta-analytic statistical approaches allow for statistical significance testing of the findings. Not only does this type of measure address the issues that have been of concern for the past 30 years, but it also allows for quantitative analysis and interpretation in addition to the previously used, more qualitative judgments. For example, the relative risk of a school district that has African American and Caucasian special education identification rates of 14% and 12%, respectively, is 1.17. A rate ratio of 1.00 would indicate that the individual risk for one group is identical to the risk index for another.

An issue that is still pertinent to both the odds ratio and the rate ratio is determining the referent group to be used (i.e., the denominator used in the equation). There are three main methods of calculating this denominator:

1. use the odds or rate for all students not in the target groups,
2. use the odds or rate for all students in the population of interest, or
3. use a consistent group for comparison.

The first method was used by Finn (1982). As already stated, this method does not include a direct comparison of groups because the composition of the denominator changes for each target group (e.g., African American, Hispanic). The influence of large groups (e.g., Caucasians) on the denominator will be greater than that of smaller groups (e.g., Native American).

The second method addresses one shortcoming of the first—it provides a common denominator. Using the odds or rate for the total population of interest as the denominator ensures that it will be consistent across comparisons. An additional problem is that the numerator of the equation (i.e., the target group) is included in the denominator. As the size of

the target group increases, the dependency of these data increases. This means that the magnitude of the ratio may, in part, depend on the size of the target group.

The third method of calculating the denominator is to use a consistent group as the comparison, for example, using Caucasian students as the comparison group for all target groups (e.g., African American and Hispanic students). This provides direct comparison of groups, offers a common denominator, and eliminates dependency in the formula. One assumption of this method, however, is that the comparison group is accurate or "true." When using Caucasian students as the comparison group (as is generally the case, by reasoning that they are the largest race/ethnicity group in the United States), the implicit assumption is that the odds or rate of identification for Caucasian students is appropriate or accurate. When using this method of denominator, it may be important to not assume direction of causation or effect. A discrepancy between the identification rates for African American and Caucasian students may be an indication of *over*representation of one group or *under*representation of the other. It should be considered only as the first step in the investigation. Further steps examining the effect of intervening variables must also be conducted.

Application of Relative Risk to National Databases

The disproportionate representation of racial groups in different categories of special education eligibility has often been established using large databases such as the OCR *Compliance Report*. This database was first established in 1968 as a means of monitoring demographic patterns in school districts that previously had civil rights violations or who applied for funds under the Emergency School Aid Act, as well as providing a sample representative of other districts across the country on 13 variables important to OCR (Finn, 1982).

Another large national database is that maintained by the Office of Special Education Programs (OSEP) for the 1998–1999 school year (U.S. Department of Education, 2000). Information from this database also indicates overall patterns of disproportionate representation. One can use the "All Disabilities" eligibility frequencies for students ages 6 to 21 years from Table AA3 (p. A-5) and U.S. student population race/ethnicity frequencies from the *Digest of Education Statistics* (available online at http://nces.ed.gov/pubs2001/digest/dt039.html) in the equation for a rate ratio. The result indicates that African American students were found eligible for special education at a rate 1.16 times that for Caucasian students (13.97% vs. 11.99%). Although this number may not appear extraordinary in isolation, when applied to an aggregation of 1 million students (approximately 1 out of every 50 U.S. students), 160,000 more African American students than Caucasian students are expected to be placed in special education programs. For Hispanic students, the rate was .92 when compared to Caucasian students. This indicates that Hispanic students are identified as eligible for special education services at a lower rate than are Caucasian students (i.e., 80,000 fewer Hispanic students than Caucasian students per million). These figures, however, are given for the overall rate of eligibility. Because representation is differentially disproportionate in different categories, the rate ratios for individual categories of eligibility can be even more alarming.

The OCR *Compliance Report* also provides data on eligibility rates for different racial groups in the high-incidence categories of special education eligibility. For the category of MR, which has been the primary focus of much of the litigation and research on disproportionate representation (for a review, see Reschly, 1988), African American students were found to be eligible at 2.23 times the rate of Caucasian students, whereas Hispanic students were identified at .65 times the rate of Caucasian students. In the category of emotional disturbance (ED), African American students were classified at 1.68 times the rate of Caucasian students, while Hispanic students were classified at .65 times the rate of Caucasian students. In contrast to these large differences, the rates for the category of learning disability (LD) are more similar. African American students were classified at 1.11 times the rate for Caucasian students, and Hispanic students were classified at 1.08 times the rate.

Although the reliability and validity of the large-scale databases has been questioned (see Finn, 1982; Reschly, 1997), the conclusions drawn from the information in them are generally agreed to be accurate. What is missing from this discussion, however, is an analysis of the variables that produce or are related to the overrepresentation patterns. For example, a number of variables—such as overall achievement rates, teachers' misperception of behavior, and test biases—may be involved. For a variety of reasons, referral rates are one of the most important variables related to understanding overrepresentation.

Importance of Referral

Referral for assessment or intervention has been cited as one of the most important predictors of future special education eligibility (Artiles & Trent, 1994; Mehan et al., 1986; Ysseldyke & Algozzine, 1983) because most students referred for consideration of special education are eventually placed in special education programs. This phenomenon is open to differing interpretations. Some researchers have suggested that teachers are accurate judges regarding which students need help. For example, Gresham, Reschly, and Carey (1987) found that teachers' judgments of student achievement (as measured by the *Teacher Rating of Academic Performance*) were as accurate as measures of intelligence (*Wechsler Intelligence Scale for Children–Revised;* Wechsler, 1974) and achievement (*Peabody Individual Achievement Test;* Dunn & Markwardt, 1970). One potential limitation of this study, however, was the use of a reevaluation sample (i.e., students in the LD group had already been in special education for nearly 3 years). To address

this limitation, Gresham, MacMillan, and Bocian (1997) conducted a similar study using students being initially referred for evaluation. Their results corroborated the accuracy of teacher judgments. More recently, Hecht and Greenfield (2001) found that teachers' ratings of student reading performance predicted future performance with an accuracy similar to that of reading-related tests.

A competing suggestion for interpretation of the strong relationship between referral and eligibility is the presence of *confirmatory bias* (O'Reilly, Northcraft, & Sabers, 1989), a term that refers to the tendency of an evaluator or team to draw conclusions that align with preliminary hypotheses despite the presence of contradictory evidence or a lack of sufficient confirmatory evidence (Darley & Gross, 1983). Using vignettes, O'Reilly et al. demonstrated that school psychologists are likely to agree with a reason for referral despite contradictory evidence. Kastner and Gottlieb (1991) demonstrated that the reason for referral is a strong predictor of classification category. This strong correlation (in the face of conflicting data) may be a result of (a) *expectancy effects* (i.e., a tendency to find that for which one is looking or for expectations to be "reified"; Rosenthal, 1991) or (b) *illusory correlation* (i.e., an assumption of the existence of an explanatory variable [e.g., neurological deficits causing poor academic performance] based on chance co-occurrence or an unmeasured variable; Gnys, Willis, & Faust, 1995).

In addition, if different groups are referred at different rates, any differences between the groups within the referred sample may be due to a larger proportion of one group being included in the sample. For example, if 23% of African American students in a district were referred but only 9% of Caucasian students were, differences between these two referral groups may be attributable to differential or biased selection criteria, differences in the underlying distribution of the trait for the two groups, differences regarding related variables, or other causes. Difference in referral rates has implications for (a) whether the same criteria for referral are being used for all students and (b) whether other variables are involved in differences in groups' performance.

The relationship between eligibility and referral suggests that a better understanding of disproportionate representation in special education categorization requires investigation of factors affecting referral rates and processes. There are two primary reasons that less attention has been directed toward referral than toward eligibility. First, most referral bias research is analogue, using reactions to vignettes in order to control for the variables of race (Zucker & Prieto, 1977), gender (Gregory, 1977), or SES (Lanier, 1975) while keeping other variables, such as achievement, consistent. Although vignettes and simulated cases have been instrumental in adding to the knowledge base, the generalization of the results of studies employing them has been questioned (Bahr, Fuchs, Stecker, & Fuchs, 1991; Shinn, Tindal, & Spira, 1987).

The second reason that there is less research on referral than on eligibility is that most studies comparing eligibility rates of various groups use large national databases. Referral data (disaggregated by racial group or gender) are not collected for such large databases as OCR's *Compliance Report* or the U.S. Department of Education's *Annual Report to Congress on the Implementation of IDEA*. In order to compare eligibility rates to referral rates, they should be in the same form (i.e., a rate ratio) and drawn from similar populations. Few studies have actually been conducted that examined referral rates by comparing them to population rates. In addition, all of these studies have been conducted within individual school districts. The only source of data in a consistent form on similar populations for a large-scale analysis thus is a synthesis of the results from individual studies.

Purpose of the Study

The present study synthesized the results of individual studies in order to compare referral rates to population rates of students from different racial backgrounds. Quantitative synthesis of this research may allow for a better understanding of overall referral rates and the processes involved. This will provide a basis for future research aimed at identifying or ruling out various factors of risk of bias. Our hypothesis was that the referral rates of the different groups are similar to the eligibility rates that have been reported. In other words, African American students are referred at a higher rate than are Caucasian students, and Caucasian and Hispanic students are referred at similar overall rates.

Method

Criteria for Eligibility of Studies

Studies were selected on the basis of criteria intended to provide a comprehensive view of the samples used in research on overrepresentation. The eligibility criteria were as follows.

Distinguishing features meant that a study had to report the frequencies or proportions of racial groups within a sample of students referred for assessment or intervention as well as the population from which the sample was drawn. The sample could be (a) the entire set of referrals in a given school or district for a given time period or (b) a random sample of the entire set. Eligible studies included referrals for psychoeducational assessment, prereferral assistance, or special education eligibility. Office referrals, disciplinary referrals, and referrals solely for related services such as counseling, health, or speech services were not eligible. Eligible studies had to use school populations for the sample, not populations in hospitals, university clinics, and correctional institutions.

Research respondents meant a study had to involve students in elementary, middle, and/or secondary settings (i.e., students ages 5–21 years without a high school diploma or equivalent). Public or private school settings were eligible, provided

they served students in this range. Preschool, early childhood, or postsecondary schools were not eligible. A study had to also include multiple racial groups (e.g., African Americans and Caucasians).

Research methods meant a study had to include quantitative demographic information regarding the racial composition of a specific referred sample and the population from which it was drawn. Vignettes, case studies, opinion papers, and literature reviews were not eligible.

Cultural and linguistic range meant a study had to be conducted in the United States. Choosing only one country was important because systems of and criteria for referral for evaluation and intervention vary from country to country. Because we sought to study the process in the United States, studies conducted in other countries were excluded.

Time frame meant the data for the studies had to have been collected after 1975, which was the year that the Education for All Handicapped Children Act of 1975 (Pub. L. 94-142) was passed. This law set the stage for special and remedial education as it is practiced today; hence, any data gathered after this date were within the context of the Act.

Publication type meant published and unpublished studies were eligible, regardless of source. This included, but was not limited to, books, book chapters, refereed journal articles, nonrefereed journal articles, dissertations, technical reports, conference papers, and unpublished manuscripts.

Identification and Retrieval of Eligible Studies

The entire population of studies, both published and unpublished, that met the eligibility criteria was targeted for retrieval. The data for the current meta-analysis were identified using a variety of sources and three general methods. The first method involved keyword bibliographic searches of computerized databases, including *Educational Resources Information Center* (ERIC), *PsycInfo, Social Sciences Abstracts, Education Abstracts,* and *Dissertation Abstracts International*. The second method was manual searches of popular refereed journals in the fields of special education and school psychology. The journals we searched were *Exceptional Children* (1975–2000), *Journal of Negro Education* (1975–2000), *Journal of School Psychology* (1980–2000), *Psychology in the Schools* (1975–2000), *School Psychology Quarterly* (1980–2000), and *School Psychology Review* (1975–2000). The third method was to examine the references of retrieved studies for additional potentially eligible studies.

The initial search identified more than 1,500 citations that potentially could contain the desired information. After reviewing the abstracts, we determined that 121 of these documents met all of the inclusion criteria, and we targeted these for retrieval. All 121 were retrieved through a library search or via interlibrary loan.

Only 9 of the 121 retrieved studies contained the data required for inclusion. One additional report, however, presented New York City population proportions but not the frequencies (Gottlieb & Alter, 1994). The total enrollment for the 32 New York City Community School Districts (CSD) was presented; however, the enrollment for each CSD was not. Using the *School Based Budget Reports* (Crew, 1999), we were able to determine the proportion of enrollment of the 32 CSDs within the Board of Education. These proportions were then applied to the total reported enrollment to derive the frequencies of students of each racial group. This creates a degree of uncertainty because the budget report data were from the 1997–1998 school year, whereas the Gottlieb and Alter data were from 1992–1993. These proportions may vary significantly over time and thus must be considered an approximation. This allowed for 10 usable studies yielding 44 comparisons.

Coding System

The first author coded the data. Each study was coded twice, with a minimum of 2 weeks between codings to ensure consistency of the coding procedure. Intracoder agreement was 100%. This high degree of consistency probably was aided by the low amount of necessary inference in the coding process. In addition, the second author coded 12 of the 44 comparisons (27%), and intercoder agreement was 99.3%.

Data were coded in two general categories of variables: study level and effect-size level. At the study level, four variables were coded: publication year, publication type (e.g., journal article, dissertation), state in which the study was conducted, and length of the study. At the effect-size level, 30 variables were coded. Twelve of these variables were not included in any study, and 3 were consistent across all studies. These 15 variables therefore were eliminated from subsequent analyses. The 15 remaining variables included school level (e.g., elementary, middle), total size of the referred sample, and which racial group was predominant in the population. Also coded were the frequencies and/or proportions of African American (AA), Caucasian (C), and Hispanic (H) students in the referred sample and in the overall population. Eligible studies and their characteristics are listed in Table 1.

Data Analysis

The desired analysis was a comparison of the rates at which each group was referred (i.e., the percentage of the population of the group that is referred). Once the data from the included studies were coded, effect sizes were calculated. Due to the nature of the data (i.e., frequency counts across two factors) and to compare the results to studies of eligibility rates, the rate ratio was used as the effect-size statistic (hereafter, the terms *effect size* and *rate ratio* are used interchangeably). The rate ratio is used for comparing two groups across levels of a dichotomous variable (Fleiss, 1994). For this meta-analysis, the referred sample and the population served as the levels of the dichotomous variable (the groups being represented by racial groups). The equation used to calculate the rate ratio for each two-group comparison thus was as follows:

$$ES = \text{Rate Ratio} = \frac{\dfrac{\text{frequency of Group A in referred sample}}{\text{frequency of Group A in population}}}{\dfrac{\text{frequency of Group B in referred sample}}{\text{frequency of Group B in population}}}$$

This equation yielded a statistic comparing the referral rates of two groups.

Because the rate ratio does not allow comparison across three levels of a variable, separate rate ratios were calculated comparing African American and Caucasian students (AA to C) and Hispanic to Caucasian students (H to C). Caucasian students were used as the comparison group because they represent the largest group nationwide. This means that most studies were expected to yield two effect sizes, one for each comparison, to be used in separate analyses. Unfortunately, only three studies included frequency counts for Hispanic students (Argulewicz & Sanchez, 1983; Farrell, Olson, Malloy, & Boykin, 1983; Gottlieb & Alter, 1994). Therefore, a total of 44 effects sizes were included in the AA to C comparison, whereas 35 were included in the H to C comparison.

Before the comparisons could be made, however, some adjustments were required. First, because the rate ratio is statistically inconvenient (i.e., centered on 1, with a minimum value of 0 and no maximum value; Lipsey & Wilson, 2000), the natural log of each effect size was used in all analyses. The logged rate ratio has an approximately normal distribution ($M = 0$, $SD = 1.83$), which allows it to be presented in a form similar to other measures of effect size and facilitates the calculation of the standard error (Lipsey & Wilson, 2000).

Second, greater weight was given to effect sizes based upon the inverse of the sampling error variance to account for the various sample sizes that the effect sizes were based upon. For the rate ratio, the equation for the sampling error was as follows:

$$SE_n = \sqrt{\frac{1 - P_A}{\#\text{ Group A in population } (P_A)} + \frac{1 - P_B}{\#\text{ Group B in population } (P_B)}}$$

where P = the risk index for a group (i.e., frequency of group in category/frequency of group in population. The weight ascribed to each effect size is then 1 divided by this number (the sampling error) squared (Lipsey & Wilson, 2000).

The first step in the analysis was to plot both sets of effect sizes using histograms and box plots to determine if there were any outliers. Once the distributions were determined to be approximately normal, the mean and the confidence interval (CI) of each effect-size distribution were calculated. A test of the homogeneity of each distribution was also conducted. If the homogeneity results are significant (i.e., Q [the homogeneity statistic] exceeds the upper tail critical value of the chi-square distribution at k –1 degrees of freedom), there is significant variability in the distribution of effect sizes. Descriptive study characteristics (e.g., predominant group in the district, state in which the district is located) can then be used to explain portions of this variance.

Results

Table 2 presents the characteristics of the comparisons from the studies in the meta-analysis. As can be inferred, the report (Gottlieb & Alter, 1994) that yielded 32 of the 44 comparisons (73%) provided consistent data across all of its comparisons. The variability among the comparisons from the Gottlieb and Alter report was tested and found to be significant for both the AA to C ($Q = 216.74$, $p < .00005$) and the H to C ($Q = 207.37$, $p < .00005$) comparisons. This is potentially problematic because to partial out any variability among the comparisons of that study, a variable is needed to discriminate among them. If a variable is constant across the 32 subsamples, it cannot be used to partial out that variability. The predominant racial group variable (i.e., which racial group constituted the largest percentage of the population) was the only variable coded that varied among the comparisons in the Gottlieb and Alter study; therefore, it was used as the explanatory variable in the analyses.

Also noteworthy is that although the logged rate ratio was used in all analyses, the scores presented in this and all subsequent sections have been transformed (i.e., unlogged) and are presented in their original form (i.e., the rate ratio; see above). This was done to facilitate discussion of the findings. The rate ratio is interpreted as the referral rate of the first group as compared to the second group. For example, in the African American to Caucasian comparison, a rate ratio of 1.5 would indicate that African Americans were referred at 1.5 times the rate of Caucasians (i.e., given their proportion of the population, African Americans were one-and-one-half times more likely to be referred).

African American to Caucasian Comparison

Visual inspection of the plots was used to determine that the effect sizes were approximately normally distributed. The 44 subsamples with African American and Caucasian frequencies yielded a fixed-effects model mean effect size of 1.58 (confidence interval from 1.53 to 1.63). An analysis of homogeneity, however, yielded a Q of 628.71 ($p < .0001$). An analog to the analysis of variance (Lipsey & Wilson, 2000) was used to determine if predominant racial group in the population accounted for the within-group variability. Because the results of this analysis revealed that predominant racial group did not sufficiently account for the variability within groups (within-$Q = 519.05$, $p < .0001$), a random-effects model was used. This type of model accounts for the unexplained variance by inserting a constant (an estimation of the common error) into the regression equations (Raudenbush, 1994). Because the unexplained variance is accounted for statistically, this method is more conservative than a fixed-effects model, which does not address the unexplained variance (Berlin, Laird, Sacks, & Chalmers, 1989). This allows evaluation of the effect that can be attributed primarily to the grouping variable

TABLE 1. Eligible Studies and Characteristics

Study	Publication type	State	Length of study (mos.)	Grades	Sample size	Predominant group	AA to C RR	H to C RR
Andrews, Wisniewski, & Mulick, 1997	Journal article	OH	12	K–12	135	Caucasian	2.69	—
		OH	12	K–12	69	Caucasian	2.24	—
Argulewicz & Sanchez, 1983	Journal article	AZ	24	K–8	318	Caucasian	.69	2.01
Diment, 1998	Dissertation	MN	12	K–5	209	African American	3.19	—
Farrell, Olson, Malloy, & Boykin, 1983	Journal article	WI	12	K–12	2,799	Caucasian	1.34	.66
		WI	12	K–12	1,973	African American	1.02	.97
Gottlieb & Alter, 1994	Technical report	NY	12	K–8	534	Hispanic	2.17	2.01
		NY	12	K–8	749	Caucasian	1.63	1.65
		NY	12	K–8	772	African American	1.02	1.04
		NY	12	K–8	696	Hispanic	2.17	1.96
		NY	12	K–8	290	African American	.43	.56
		NY	12	K–8	693	Hispanic	.94	.59
		NY	12	K–8	751	Hispanic	.31	.25
		NY	12	K–8	539	Hispanic	1.63	1.14
		NY	12	K–8	815	Hispanic	.54	.50
		NY	12	K–8	1,511	Hispanic	1.54	1.25
		NY	12	K–8	1,145	African American	1.09	1.11
		NY	12	K–8	942	Hispanic	.87	.81
		NY	12	K–8	527	African American	.44	.49
		NY	12	K–8	829	Hispanic	2.06	1.41
		NY	12	K–8	872	Hispanic	2.65	1.85
		NY	12	K–8	350	African American	.34	.35
		NY	12	K–8	561	African American	.38	.20
		NY	12	K–8	474	African American	2.42	2.76
		NY	12	K–8	878	African American	.96	.97
		NY	12	K–8	764	Caucasian	.98	.89
		NY	12	K–8	1,027	Caucasian	1.90	1.58
		NY	12	K–8	1,134	African American	1.58	1.29
		NY	12	K–8	467	African American	3.44	2.17
		NY	12	K–8	787	Hispanic	1.30	1.05
		NY	12	K–8	623	Caucasian	1.49	1.61
		NY	12	K–8	441	Caucasian	1.29	1.25
		NY	12	K–8	1,178	African American	1.17	.89
		NY	12	K–8	921	African American	1.69	1.36
		NY	12	K–8	687	African American	1.48	.98
		NY	12	K–8	944	Hispanic	2.24	1.40
		NY	12	K–8	1,577	Caucasian	1.48	1.47
		NY	12	K–8	678	Hispanic	.61	.58
Lietz & Gregory, 1978	Journal article	WI	11	K–5	69	Caucasian	1.02	—
Masica, 1996	Dissertation	OH	12	K–12	275	Caucasian	.82	—
Speece & Cooper, 1990	Journal article	MD	12	1	63	Caucasian	.93	—
Tyler & Manzo, 1999	Technical report	NC	12	K–12	1,228	Caucasian	2.40	—
		NC	12	K–12	3,098	Caucasian	2.55	—
Wisniewski, Andrews, & Mulick, 1995	Journal article	OH	12	K–12	312	Caucasian	1.17	—

Note. AA = African American; C = Caucasian; H = Hispanic; RR = rate ratio.

TABLE 2. Characteristics of Included Subsamples

Variable	*n*	%
Publication type		
Journal article	8	18
Dissertation	2	5
Technical report	34	77
Publication year		
1978	1	2
1983	3	7
1990	1	2
1994	32	73
1995	1	2
1996	1	2
1997	2	5
1998	1	2
1999	2	5
State conducted in		
New York	32	73
Other than New York	12	27
Length of study (in months)		
11	41	93
24	1	2
Setting		
Elementary school	6	14
Elementary & middle school	33	75
All levels	5	11
Predominant racial group		
Caucasian	16	36
African American	15	34
Hispanic	13	30

Note. Due to rounding, some percentages may not add to 100.

of interest (in this case, predominant racial group) because the unexplained variability (i.e., the error) is controlled for.

With this accounted-for variability (within-$Q = 49.19$, $p = .18$), the between-groups effect was found to be nonsignificant (between-$Q = 1.42$, $p = .49$). The referral rate of African American students, as compared to Caucasian students, was not significantly influenced by which racial group was predominant in the population under study. Examination of the overall effect, however, revealed a significant discrepancy between the referral rates of African American and Caucasian students ($M = 1.32$, $p = .0005$).

Hispanic to Caucasian Comparison

Visual inspection of the plots was used to determine that the effect sizes were approximately normally distributed. The 35 subsamples with Hispanic and Caucasian frequencies yielded a fixed-effects model mean effect size of 1.05 (confidence interval from 1.02 to 1.09). An analysis of homogeneity, however, yielded a Q of 410.61 ($p < .0001$). Attempts at partialing out the Q by means of an analog to the ANOVA using predominant racial group was unsuccessful (within-$Q = 391.46$, $p < .0001$). As a result, a random-effects model was adopted that accounted for the within-group variability ($Q = 38.82$, $p = .19$) but yielded a nonsignificant between-group difference ($Q = 1.97$, $p = .37$). Examination of the overall effect size also revealed a nonsignificant difference ($M = 1.06$, $p = .49$). The referral rates for Caucasian and Hispanic students were found to be similar.

Discussion

The results indicated that the referral rates of racial groups vary significantly. The initial fixed-effects models indicated that the rate of referral was greater for both African Americans and Hispanics as compared to Caucasians. Because no intervening variables were identified that mediated these referral rate differences, the variability within each set of comparisons was statistically accounted for (i.e., random-effects models were used). Although the random-effects African American to Caucasian comparison was significant despite the larger confidence interval, the Hispanic to Caucasian comparison was not.

Compared to the national rates for eligibility stated previously, the referral rates identified by the present analysis indicate higher rates for both African American and Hispanic students. The OSEP-reported eligibility rate for African American students, compared to that for Caucasian students, was 1.18 versus a mean referral rate of 1.32. For Hispanic students compared to Caucasian students, the OSEP-reported eligibility rate of .89 was lower than the mean referral rate of 1.06. For example, these rates indicated that for every 100 Caucasian students who are referred for assessment or intervention, 132 African American students and 106 Hispanic students are referred. In contrast, for every 100 Caucasian students found to be eligible for special education, 118 African American students and 89 Hispanic students are found to be eligible. This indicates that (a) more African American students and a similar number of Hispanic students are referred for special education and (b) more African American students but fewer Hispanic students than Caucasian students are found to be eligible for special education.

Because the operational definition of referral for this analysis included referrals for assessment *or* intervention, students who ultimately were provided with services through means other than special education (e.g., Title I, Bilingual or English Language Learner [ELL] services) appear in the referral data but not in the special education classification data. Differences in referral and eligibility rates may be explained by disproportionate numbers of African American and Hispanic students served by Title I and ELL, respectively (MacMillan & Reschly, 1998). For example, consider the referral rate of 100 Caucasian students versus the referral rate of 132 African American students. If the base eligibility rate (from the referred sample) for Caucasian students were 60%, the corre-

sponding rate for African American students would be 60% times 1.18 (the ratio of African American eligibility rate compared to Caucasian rate), or 70.8% (94 of the 132 students). In addition to the 94 students found to be eligible for special education, 10 students might be provided services through Title I or a district-wide remedial reading program. Ineligibility for special education services does not automatically equate to inappropriate referral—especially when other outcomes (e.g., Title I) also may indicate appropriate referral. Without eligibility rates for Title I and other alternate remedial services disaggregated by race, an exact account of overall eligibility rates to compare to referral rates is not possible. The rest of this discussion thus primarily addresses potential sources of variability in rates of referral between different groups, acknowledging that referral, assessment, eligibility, and placement are interrelated phases of one process.

The most common reason to refer a student for assessment or intervention is low achievement (Gottlieb & Alter, 1994; Hyde, 1975; Kastner & Gottlieb, 1991). Because an adverse impact on educational performance is an important indicator of the need for special education services, this is not surprising. Referrals for low achievement most likely culminate in classification as MR or LD—depending upon performance on achievement and intelligence tests. Another common reason for referral, however, is disruptive or disturbing behavior (M. G. Anderson, 1994; Hyde, 1975). There is limited evidence that African American students may be referred more than Caucasian students for behavior issues (Del'Homme, Kasari, Forness, & Bagley, 1996). Clearly, more research is needed in this area; however, several claims have been made that this discrepancy in referral reason exists and may be rooted in cultural differences and/or perceptions of higher rates of behavioral difficulties (Harry & Anderson, 1994; Hilliard, 1992; Serwatka et al., 1995).

Researchers have demonstrated that teachers' perceptions of appropriate behaviors (Rong, 1996) and deviant behaviors (Puig et al., 1999; Weisz et al., 1988) are culturally defined. In addition, it has been demonstrated that teachers may perceive the same behavior differently in students of different ethnicity from themselves (Carlson & Stephens, 1986) or in students considered "bad" rather than "good" (Mehan et al., 1986). Given the fact that almost 90% of teachers nationwide are Caucasian (U.S. Department of Education, 1998), the possibility exists for cultural misunderstanding or mismatch of expectations that might affect referral rates of non-Caucasian students.

Another implication of the relation between referral and eligibility rates is support for the presence of confirmatory bias (Darley & Gross, 1983). When a student is referred for assessment of special education eligibility, a hypothesis is made that the child has a disability and needs special education services (Walker, 1976, as cited in O'Reilly et al., 1989). Several mechanisms for confirming this hypothesis have been demonstrated in the special education eligibility process. These include selective data gathering (Snyder & Gangestad, 1983), selective data use (Langer & Abelson, 1972), and ignorance of disconfirming data (Nisbett & Ross, 1980). Other factors that may contribute to confirmatory bias include the differential interpretation of behavior, depending on student ethnicity (Carlson & Stephens, 1986) and cultural expectations (Collier & Hoover, 1987; Harry & Anderson, 1994), as has already been mentioned. These may lead to the creation of additional a priori hypotheses for the evaluator or multidisciplinary team members. These hypotheses will then be subject to conscious or unconscious strategies of confirmation. Confirmatory bias, however, is only problematic in instances of inappropriate referral or negative outcomes.

When the assessment process confirms an accurate hypothesis (e.g., a student's achievement is significantly behind his or her ability, and he or she is therefore is determined to have a learning disability), no bias can be identified because the "biased" decision and the "correct" decision are the same. When a student is inappropriately referred, the biased decision and the correct decision are at odds and confirmatory bias *might* exist. In the context of disproportionate representation, if students of different racial groups are inappropriately referred at similar rates, then confirmatory bias is a generic problem to address. If students of different racial groups are inappropriately referred at different rates, however, there is inequality of process or treatment (Reschly, 1997), and addressing confirmatory bias may be considered in terms of a civil rights issue as well as a generic problem.

The issue of appropriate versus inappropriate referral revolves around the outcomes of that referral. As stated previously, referral outcomes other than special education eligibility or ineligibility are possible. Students may receive services through a Section 504 plan, Title I, a district-wide intervention program (e.g., *Reading Recovery*), or supplementary services (e.g., individual or group counseling). Provided the student's needs have been accurately identified and are addressed through these services, the outcome of the referral should be viewed as positive. As stated previously, disproportionate representation in special education is often viewed as negative because special education is seen as ineffective, or worse, detrimental to students' development. Conversely, if special education is viewed as a positive service that addresses each student's individual needs, representation in special education should be proportional to the needs of different groups, not their percentage of the population. Because African American and Hispanic students have been demonstrated to perform more poorly than their Caucasian peers on academic measures (Meece & Kurtz-Costes, 2001) and often come from more economically disadvantaged backgrounds (Kennedy, Jung, & Orland, 1986), some critics have argued that their rates of referral and eligibility for special education should be higher than the same rates for Caucasian students (Forness & Knitzer, 1992; Kauffman, 1999; Oswald, 1995). This position is also based on one of the following two premises:

1. The assessment measures and selection criteria do not bias the outcome in negative ways.
2. Bias is unimportant because it is better to provide special education to a child who does not need it rather than not provide services to a child who needs them.

This meta-analysis provides support for the variability of referral rates among different groups of students. It also highlights the importance of referral in future classification in special education categories. These results do need to be considered, however, in light of some limitations.

Limitations

The number of districts represented in the present study is small compared to the total number in the United States. The degree to which a sample is representative of the population being generalized to is negatively correlated to sample size (Glass & Hopkins, 1996). Of the more than 16,000 school districts in the United States today, only 42 were represented in this study (two districts were represented twice). With such a small sample, the potential error in sampling is great. The sampling of these districts is not random either because each consented to some degree to participation in a research study. In addition, the districts included in the sample were (a) exclusively from predominantly urban or suburban areas, (b) all of medium to large size—only 3 having fewer than 10,000 students, and (c) all public school districts—no private or parochial schools were represented. These factors suggest that generalization from this sample to all school districts in the United States is questionable.

It is also possible that the variable(s) that could account for the variability in the AA to C comparison were not coded for this meta-analysis. Inclusion of additional descriptor variables would be necessary for a more comprehensive synthesis, but this is related to the number of eligible studies that are available. If the sample of coded comparisons was larger and/or more diverse (i.e., coming from more studies), it would have been possible to analyze more complex models. For example, the effect that publication year or state the district is in has on referral rates could have been examined. In addition, if the studies had reported the reason for referral more consistently, this variable might have been found to differentially affect the referral rates of racial groups.

A third limitation is that most of the effect sizes were derived from a single research study (Gottlieb & Alter, 1994). If other variables may account for some of the variability among the effect sizes (e.g., publication year, publication type, school level), a larger, more diverse group of studies would be necessary to make the appropriate comparisons. Despite these limitations, however, there are clear implications for future research and practice.

Implications for Future Research

Although more research has focused on disproportionality in classification as disabled and placement in special education than on referral rates, the patterns identified in this meta-analysis are somewhat consistent with others reported in the literature. In order to make comparisons between referral and eligibility rates on a similar scale, a vast amount of additional research is needed to be able to compare groups of similar size. This could be achieved in two ways. First, the amount of available studies reporting referred sample and population frequencies could be increased. If researchers adopted inclusion of these data as standard practice in articles on disproportionate representation in the referral process, a substantial body of literature would be developed over time. This literature could then be examined not just for overall trends but also for intervening variables that might moderate the effects of disproportionate representation in referrals.

Second, a large national database could be developed that included frequencies of referral and population for different racial groups. The best-case scenario would be that the OCR *Compliance Report* or the U.S. Department of Education's *Annual Report to Congress* could be restructured to include these data. The same sample thus would be used for the referral and eligibility data, and a direct comparison would not require any generalization between samples.

Another direction for future research is to disaggregate the data at a more specific level than nationally. Finn (1982) made comparisons not only at the national level but also by disaggregating the data by state. In general, he noted that disproportion was not an issue in the northeastern, midwest, or western states. The states that had the most severe disproportionality were in the southeastern region of the United States, the region where most of the difficulty in developing integrated school systems in the 1960s occurred (Spring, 1997).

Although Finn (1982) compared data at the state level and found some interesting trends, he built a good case for disaggregating at the district level as well. First, he argued, in data aggregated at the state level, large districts may overshadow or obscure smaller districts. This could occur because a larger district contains a larger proportion of the state's population. It is conceivable that size of the district may affect trends in disproportionality. Second, districts with no students in a certain program will inflate the state's population numbers, reducing the appearance of disproportionality. Finn cited the example of the 1978 OCR survey, which showed only 12 states in which all of the districts reported having students with MMR. By aggregating at the district level, analyses could be done to look at the effect of including or excluding these districts. Third, eligibility and placement, although regulated at the state level, are district-by-district processes. Just as Mercer, Jordan, Allsopp, and Mercer (1996) have shown that there is great variability among states in eligibility criteria, there is also variability within states on the criteria that dis-

tricts utilize. Aggregation of data at the district level would allow a closer investigation of this phenomenon and provide a better indicator as to which district-level variables may intervene in this area.

A third direction for future research is to examine other variables that may affect disproportionate representation. Oswald et al. (1999) took preliminary steps in this direction. Using the 1992 OCR database, they examined predictors of categorization as MR or ED. Their results, aggregated at the district level, indicate that environmental factors (e.g., median housing income, percentage of children in poverty) predict ED rate ($R^2 = .117$) and MR rate ($R^2 = .356$). The addition of race variables (percentage of African American enrollment and whether a student was African American or not) added significantly to the models (both ps $< .0001$). Another interesting finding is that the environmental/racial model was a better predictor for African American students than other students. Oswald et al. concluded that race-related factors influence the identification of African American students more than non–African American students.

As suggested by Oswald et al. (1999), proportion of minority students in the district is an important variable to consider. Although no relationship was found in the present study, this could be a result of the small sample (see above). However, several other researchers (Finn, 1982; Gottlieb & Alter, 1994; Harry, 1992) have noted this phenomenon and have documented, to varying degrees, that disproportionate representation is lower at the extremes. In other words, when the minority population in a school is a small percentage of the overall population (less than 10%) or a very large percentage (over 90%), the eligibility rates for different minority groups are more similar than in schools with minority percentages between 10 and 90. This trend does not hold true for small districts (those with total enrollment below 3000), however. In these districts, disproportion increases as minority percentage increases.

Another variable to examine is the relation between SES and disproportionate representation. Ethnicity has been thought of as being a proxy for SES (MacMillan & Reschly, 1998). This is illustrated by the impact of poverty on a child's life regardless of his or her race (Hodgkinson, 1995). As was shown by Oswald et al. (1999), environmental factors provide a strong predictor of identification. When Oswald et al. examined the specific interaction between poverty variables in a district and race, distinct patterns emerged. For MR, African American students were overrepresented at all levels of each poverty variable (median household income, percentage of children in poverty, mean housing value). For ED, however, as indicators of poverty in a district decreased (i.e., SES increased), overrepresentation of African American students increased. This distinction between MR and ED rates lends support to (a) the differentiation between academic and behavioral reasons for referral and (b) cultural tolerance as a contributing factor in behavioral referral. SES also needs to be considered at different levels of aggregation or influence. Anderson, Hollinger, and Conaty (1993) demonstrated that a school's poverty level and individual poverty status both have significant and distinct influences on a student's academic performance. Further research in this area is certainly warranted.

Some areas that have not been studied to date also need further examination. The Oswald et al. (1999) study focused on district-level variables (because that was the level of aggregation of the data). An interesting comparison would be a hierarchical model of schools within districts and the addition of school-level variables. Also, because special education is intended to be an individual-level decision process, various student and teacher variables such as interaction patterns, behavioral variables, and demographic variables may provide insight into the issue and potential strategies for addressing the problem of disproportionate representation.

Conclusions

Comparisons of the rate of referral for intervention or assessment of various racial and ethnic groups are practically absent from the research literature. Of the studies that have focused on referred samples, most do not report the proportions of each group in the population from which the sample was derived. In addition, most studies that have focused on potential bias in referral processes have relied on vignettes rather than actual referrals. Both of these methods, while providing much information and insight, are limited in their generalizability. Studies not reporting population data do not take into consideration possible differential processes for referral for various groups. Studies using vignettes may be capable of detecting overt cases of bias in attitudes, but they do not examine the actual practice of referral.

This meta-analysis provides preliminary data regarding referral rates as compared to population proportions. The results reflect what has been previously reported in narrative literature reviews and analyses of the OCR database. Nevertheless, definite answers are precluded by the paucity of research reporting the frequency counts necessary to quantitatively synthesize findings. Future research on bias in referred samples should include population demographics such as the frequency of the various groups being compared. Only then will the field be able to assess whether research on bias in referrals provides an accurate, overall picture of actual practice and determine where bias resides in the process and how it can be eliminated.

REFERENCES

References marked with an asterisk (*) indicate studies included in the meta-analysis.

Anderson, J., Hollinger, D., & Conaty, J. (Spring *1993).* Re-examining the relationship between school poverty and student achievement. *ERS Spectrum, 11*(1), 21–31.

Anderson, M. G. (1994). Perceptions about behavioral disorders in African American cultures and communities. In R. Peterson & S. Ishii-Jordan (Eds.), *Multicultural issues in the education of students with behavioral disorders* (pp. 93–104). Cambridge, MA: Brookline Books.

*Andrews, T. J., Wisniewski, J. J., & Mulick, J. A. (1997). Variables influ-

encing teachers' decisions to refer children for school psychological assessment services. *Psychology in the Schools, 34,* 239–244.

*Argulewicz, E., & Sanchez, D. (1983). The special education evaluation process as a moderator of false positives. *Exceptional Children, 49,* 452–454.

Artiles, A. L., & Trent, S. C. (1994). Overrepresentation of minority students in special education: A continuing debate. *The Journal of Special Education, 27,* 410–437.

Assistance to states for the education of children with disabilities: Final regulations, 64 Fed. Reg. 12, 405–412, 672 (March 12, 1999).

Bahr, M. W., Fuchs, D., Stecker, P. M., & Fuchs, L. S. (1991). Are teachers' perceptions of difficult-to-teach students racially biased? *School Psychology Review, 20,* 599–608.

Berlin, J. A., Laird, N. M., Sacks, H. S., & Chalmers, T. C. (1989). A comparison of statistical methods for combining event rates from clinical trials. *Statistics in Medicine, 8,* 141–151.

Carlson, P. E., & Stephens, T. M. (1986). Cultural bias and identification of behaviorally disordered children. *Behavioral Disorders, 11,* 191–199.

Chinn, P. C., & Hughes, S. (1987). Representation of minority students in special education classes. *Remedial and Special Education, 8*(4), 41–46.

Collier, C., & Hoover, J. J. (1987). Sociocultural considerations when referring minority children for learning disabilities. *Learning Disabilities Focus, 3*(1), 39–45.

Coutinho, M. J., & Oswald, D. P. (1998). Ethnicity and special education research: Identifying questions and methods. *Behavioral Disorders, 24,* 66–73.

Crew, R. F. (1999). *School based budget reports: Fiscal year 1998-1999.* New York : Board of Education of the City of New York.

Darley, J. M., & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology, 44,* 249–256.

Del'Homme, M., Kasari, C., Forness, S. R., & Bagley, R. (1996). Prereferral intervention and students at-risk for emotional or behavioral disorders. *Education and Treatment of Children, 19*(3), 272–285.

Diana v. State Board of Education, No. C-7037 RFP, 1970 U. S. Dist. Court, Northern District of California, consent decree.

*Diment, K. A. (1998). Identification of seriously emotionally disturbed students: An ethnic comparison of maladaptive behaviors at both point of referral and eligibility (Doctoral dissertation, University of Minnesota, 1998). *Dissertation Abstracts International, 59,* 2341.

Dunn, L. M. (1968). Special education for the mildly mentally retarded: Is much of it justifiable? *Exceptional Children, 23,* 5–21.

Dunn, L. M., & Markwardt, F. C., Jr. (1970). *Peabody individual achievement test.* Circle Pines, MN: American Guidance Service.

*Farrell, W., Olson, J., Malloy, W., & Boykin, W. (1983). Discrimination in educational placement and referral. *Integrated Education, 21,* 120–123.

Finn, J. D. (1982). Patterns in special education placement as revealed by the OCR survey. In K. A. Heller, W. Holtzman, & S. Messick (Ed.), *Placing children in special education: A strategy for equity* (pp. 322–381). Washington, DC: National Academy Press.

Fleiss, J. L. (1994). Measures of effect size for categorical data. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 245–260). New York: Russell Sage Foundation.

Forness, S. R., & Knitzer, J. (1992). A new proposed definition and terminology to replace "serious emotional disturbance" in Individuals with Disabilities Education Act. *School Psychology Review, 21,* 12–20.

Gerber, M. M., & Semmel, M. I. (1984). Teacher as imperfect test: Reconceptualizing the referral process. *Educational Psychologist, 19*(3), 137–148.

Glass, G. V., & Hopkins, K. D. (1996). *Statistical methods in education and psychology* (3rd ed.). Boston: Allyn & Bacon.

Gnys, J. A., Willis, W. G., & Faust, D. (1995). School psychologists' diagnosis of learning disabilities: A study of illusory correlation. *Journal of School Psychology, 33*(1), 59–73.

*Gottlieb, J., & Alter, M. (1994). *An analysis of referrals, placement, and progress of children with disabilities who attend New York City public schools.* New York: New York University. (ERIC Document Reproduction Service No. 414372)

Gregory, M. (1977). Sex bias in school referrals. *Journal of School Psychology, 15,* 5–8.

Gresham, F. M., MacMillan, D. L., & Bocian, K. M. (1997). Teachers as "tests": Differential validity of teacher judgments in identifying students at-risk for learning difficulties. *School Psychology Review, 26,* 47–60.

Gresham, F. M., Reschly, D. J., & Carey, M. P. (1987). Teachers as "tests": Classification accuracy and concurrent validation in the identification of learning disabled children. *School Psychology Review, 16,* 543–553.

Guadalupe Organization v. Tempe Elementary School District No. 3, No. 71-435, 1973. U. S. Dist. Court, District of Arizona, consent decree.

Hale-Benson, J. (1986). *Black children: Their roots, culture, and learning styles.* Baltimore: Johns Hopkins University Press.

Harry, B. (1992). *Cultural diversity, families, and the special education system: Communication for empowerment.* New York: Teachers College Press.

Harry, B., & Anderson, M. G. (1994). The disproportionate placement of African American males in special education programs: A critique of the process. *Journal of Negro Education, 63,* 602–619.

Hecht, S. A., & Greenfield, D. B. (2001). Comparing the predictive validity of first grade teacher ratings and reading-related tests on third grade levels of reading skills in young children exposed to poverty. *School Psychology Review, 30,* 50–69.

Heller, K. A., Holtzman, W. H., & Messick, S. (Eds.). (1982). *Placing children in special education: A strategy for equity.* Washington, DC: National Academy Press.

Hilliard, A. G. (1992). Behavioral style, culture, and teaching and learning. *Journal of Negro Education, 61,* 370–377.

Hodgkinson, H. L. (1995). What should we call people? Race, class, and the Census for 2000. *Phi Delta Kappan, 77,* 173–179.

Hyde, E. M. (1975). School psychologist referrals in an inner-city school. *Psychology in the Schools, 12,* 412–420.

Individuals with Disabilities Education Act of 1990, 20 U.S.C. § 1400 *et seq.*

Kastner, J., & Gottlieb, J. (1991). Classification of children in special education: Importance of pre-assessment information. *Psychology in the Schools, 28,* 19–27.

Kauffman, J. M. (1999). How we prevent the prevention of emotional and behavioral disorders. *Exceptional Children, 65,* 448–468.

Kavale, K. A., & Forness, S. R. (1999). Effectiveness of special education. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (3rd ed., pp. 984–1024). New York: Wiley.

Kennedy, M. M., Jung, R. K., & Orland, M. E. (1986). *Poverty, achievement, and the distribution of compensatory education services.* Washington, DC: U.S. Department of Education. (ERIC Document Reproduction Service No. ED 293902)

Kleinbaum, D. G., Kupper, L. L., Muller, K. E., & Nizam, A. (1998). *Applied regression analysis and other multivariable methods* (3rd ed.). Pacific Grove, CA: Duxbury Press.

Lambert, M. C., Puig, M., Rowan, G. T., Lyubansky, M., & Winfrey, T. (1998). *Adult perspectives on behavior and emotional problems in African American children.* Manuscript submitted for publication.

Langer, E. J., & Abelson, R. P. (1972). A patient by any other name . . . : Clinical group differences in labeling bias. *Journal of Consulting and Clinical Psychology, 42,* 4–9.

Lanier, J. E. (1975). *Teachers' attitudes toward students' race, socio-economic status, sex, and classroom behavior on their referral to an educable mentally retarded program* (EMR). Unpublished doctoral dissertation, University of Florida, Gainesville.

*Lietz, J., & Gregory, M. (1978). Pupil race and sex as determinants of office and exceptional educational referrals. *Education Research Quarterly, 3,* 61–66.

Lipsey, M. W., & Wilson, D. B. (2000). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

MacMillan, D. L., & Reschly, D. J. (1998). Overrepresentation of minority students: The case for greater specificity of the variables examined. *The Journal of Special Education, 32,* 15–24.

*Masica, K. (1996). The relationships between gender, race, behavior, ability-

achievement discrepancies, and eligibility for learning disabilities services among urban elementary students (Doctoral dissertation, Ohio State University, 1996). *Dissertation Abstracts International, 57,* 4261.

McIntyre, T. (1992). A primer on cultural diversity for educators. *Multicultural Forum, 1*(1), 6 & 13.

Meece, J. L., & Kurtz-Costes, B. (2001). Introduction: The schooling of ethnic minority children and youth. *Educational Psychologist, 36,* 1–7.

Mehan, H., Hertweck, A., & Miehls, J. L. (1986). *Handicapping the handicapped.* Stanford, CA: Stanford University Press.

Mercer, C. D., Jordan, L., Allsopp, D. H., & Mercer, A. R. (1996). Learning disabilities definitions and criteria used by state education departments. *Learning Disability Quarterly, 19,* 217–232.

Mercer, J. R. (1973). *Labeling the mentally retarded.* Berkeley: University of California Press.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice Hall.

O'Reilly, C., Northcraft, G. B., & Sabers, D. (1989). The confirmation bias in special education eligibility decisions. *School Psychology Review, 18,* 126–135.

Oswald, D. P. (1995). Identification of special education students with mental retardation: Correlates of state child-count data. *Journal of Child and Family Studies, 4,* 389–397.

Oswald, D. P., Coutinho, M. J., Best, A. M., & Singh, N. N. (1999). Ethnic representation in special education: The influence of school-related economic and demographic variables. *The Journal of Special Education, 32,* 194–206.

Patton, J. M. (1997). Disproportionate representation in gifted programs: Best practices for meeting this challenge. In A. J. Artiles & G. Zamora-Duran (Eds.), *Reducing disproportionate representation of culturally diverse students in special and gifted education* (pp. 59–86). Reston, VA: The Council for Exceptional Children.

Patton, J. M. (1998). The disproportionate representation of African Americans in special education: Looking behind the curtain for understanding and solutions. *The Journal of Special Education, 32,* 25–31.

Puig, M., Lambert, M. C., Rowan, G. T., Winfrey, T., Lyubansky, M., Hannah, S. D., et al. (1999). Behavioral and emotional problems among Jamaican and African American children, ages 6 to 11: Teacher reports versus direct observations. *Journal of Emotional and Behavioral Disorders, 7,* 240–250.

Raudenbush, S. W. (1994). Random effects models. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 301–322). New York: Russel Sage Foundation.

Reschly, D. J. (1988). Minority mild mental retardation overrepresentation: Legal issues, research findings, and reform trends. In M. Wang, M. Reynolds, & H. Walberg (Eds.), *Handbook of special education: Research and practice* (Vol. 2, pp. 23–41). Oxford, England: Pergamon.

Reschly, D. J. (1997). *Disproportionate minority representation in general and special education: Patterns, issues, and alternatives.* Des Moines: Iowa Department of Education.

Reschly, D. J., & Wilson, M. S. (1990). Cognitive processing vs. traditional intelligence: Diagnostic utility, intervention implications, and treatment validity. *School Psychology Review, 19,* 443–458.

Rong, X. L. (1996). Effects of race and gender on teachers' perceptions of the social behavior of elementary students. *Urban Education, 31,* 261-290.

Rosenthal, R. (1991). Teacher expectancy effects: A brief update 25 years after the Pygmalion experiment. *Journal of Research in Education, 1,* 3–12.

Serwatka, T. S., Deering, S., & Grant, P. (1995). Disproportionate representation of African Americans in emotionally handicapped classes. *Journal of Black Studies, 25,* 492–506.

Serwatka, T. S., Dove, T., & Hodge, W. (1986). Black students in special education: Issues and implications for community involvement. *Negro Education Review, 37,* 17–26.

Shinn, M. R., Tindal, G. A., & Spira, D. A. (1987). Special education referrals as an index of teacher tolerance: Are teachers imperfect tests? *Exceptional Children, 54,* 32–40.

Snyder, M., & Gangestad, S. (1983). Hypothesis-testing processes. In J. Harvey, W. Ickes, & P. Kidd (Eds.), *New directions in attribution research* (pp. 43–65). Hillsdale, NJ: Erlbaum.

*Speece, D., & Cooper, D. (1990). Ontogeny of school failure: Classification of first-grade children. *American Educational Research Journal, 27,* 119–140.

Spring, J. H. (1997). *The American school 1642-1996* (4th ed.). New York: McGraw-Hill.

*Tyler, D., & Manzo, D. (1999). *Intervention/prevention grant evaluation 1996-1998* (Report No. E&R 99.16). Raleigh, NC: Wake County Public Schools System. (ERIC Document Reproduction Service No. ED 431811)

U.S. Department of Education. (2000). *Twenty-second annual report to Congress on the implementation of the Individuals with Disabilities Education Act.* Washington, DC: Author.

U.S. Department of Education, Office of Civil Rights. (1994). *Elementary and secondary school civil rights compliance report.* Washington, DC: Author.

U.S. Department of Education, Office of Educational Research and Improvement. (1998). *The schools and staffing survey (SASS) and teacher followup survey (TFS) CD-ROM: Electronic codebook and public-use data for three cycles of SASS and TFS* [CD-ROM data file]. Washington, DC: National Center for Educational Statistics. (NCES 98-312)

Wechsler, D. (1974). *Wechsler intelligence scale for children–Revised.* New York: Psychological Corp.

Weisz, J. R., Suwanlert, S., Chaiyasit, W., Weiss, B., Achenbach, T. M., & Trevathan, D. (1988). Epidemiology of behavioral and emotional problems among Thai and American children: Teacher reports for ages 6-11. *Journal of Child Psychology and Psychiatry, 30,* 471–484.

*Wisniewski, J. J., Andrews, T. J., & Mulick, J. A. (1995). Objective and subjective factors in the disproportionate referral of children for academic problems. *Journal of Consulting and Clinical Psychology, 63,* 1032–1036.

Ysseldyke, J. E., & Algozzine, B. (1983). LD or not LD: That's not the question! J*ournal of Learning Disabilities, 16,* 29-31.

Zucker, S. H., & Prieto, A. G. (1977). Ethnicity and teacher bias in educational decisions. *Instructional Psychology, 4,* 2–5.