A Meta-Analysis of the Academic Status of Students with Emotional/Behavioral Disturbance

Robert Reid and **Jorge E. Gonzalez,** *Texas A & M University*
Philip D. Nordness, *Western Illinois University*
Alexandra Trout, *University of Iowa*
Michael H. Epstein, *University of Nebraska*

Emotional/behavioral disturbance (EBD) is characterized by a range of behaviors that adversely affect a child's academic performance and cannot be explained by other sensory or health impairments. Although research has clearly demonstrated that children and youth with EBD tend to exhibit high rates of problem behavior, research on the characteristics of their academic performance has been less clear. This article reports the results of a meta-analysis of the academic status of students with EBD. The overall effect size was −.64, which indicated that students with EBD had significant deficits in academic achievement. An examination of moderators (subject area, setting, and age) indicated that students with EBD performed at a significantly lower level than did students without disabilities across academic subjects and settings; greater deficits were not observed in older students with EBD (i.e., those more than 12 years old), as compared to younger students. Implications and areas for future research are discussed.

Emotional/behavioral disturbance (EBD) is characterized by a range of problems that adversely affect a child's academic performance and cannot be explained by intellectual, sensory, or other health factors (Individuals with Disabilities Education Act, 1997). These behaviors can include a number of internalizing and externalizing characteristics that inhibit a child's ability to build and maintain successful social relationships with peers, teachers, and adults. Children and adolescents with EBD are overwhelmingly male, behaviorally disruptive, noncompliant, verbally abusive, and aggressive. Because their behaviors are so disruptive and irritating, these children often arouse negative feelings in others, alienating schoolmates and adults and ultimately robbing these children of the benefits of learning opportunities (Kauffman, 2001). Inevitably, these behaviors significantly impair a child's ability to succeed in school and in society.

Although research has unequivocally demonstrated that children and youth with EBD exhibit high rates of problem behavior, research on the characteristics of their academic performance remains uncertain (Kauffman, Cullinan, & Epstein, 1987). Data on identification, academic outcomes, graduation rates, absenteeism, employment status, and criminality among children and youth with EBD has suggested that educating them is a complex, confusing, and often daunting task for educators, related services personnel, and family members alike (Smith & Coutinho, 1997). More specifically, the preponderance of studies on academic performance have indicated that students with EBD perform 1 to 2 years below grade level (Trout, Nordness, Pierce, & Epstein, 2003), with academic difficulties emerging at an early age and persisting throughout their schooling (Coutinho, 1986; Rosenblatt & Rosenblatt, 1999; Wagner, 1995). This should be expected, because current criteria for identification of EBD require that students show a deficit in academic achievement. Compared to peers from other high-incidence disability groups, children with EBD evince lower reading and math scores, lower graduation rates, and higher rates of course failure and grade retention, and are less likely to attend postsecondary school (Kauffman, 2001; Wagner, 1995). In 1998–1999, 50.6% of students age 14 and older with EBD had dropped out of high school (U.S. Department of Education, 2001). Furthermore, it has been estimated that 70% of students with EBD will be arrested within 3 years of leaving school, continuing a pervasive pattern of failure that becomes difficult to correct (U.S. Department of Health and Human Services, 1999).

Low academic performance and maladaptive behavior patterns are highly related. Evidence suggests that a reciprocal relationship between school failure and social failure emerges early in life (Brier, 1995; Kauffman, 2001), although the mechanism of the relationship is unknown. The prevalence of academic difficulties among children with EBD is uncertain. Ruhl and Berlinghoff (1992) suggested that between 33% and 81% of children with behavioral disorders have academic difficulties. A causal relationship between behavioral problems and academic underachievement, however, has yet to be determined (Hinshaw, 1992). Nonetheless, researchers have demonstrated that academic failure is one of the most powerful

Address: Robert Reid, 202L Barkley, University of Nebraska, Lincoln, NE 68583-0732

predictors of problem behavior and social failure (Maguin & Loeber, 1996; Morrison & D'Incau, 1997). Conversely, researchers have also demonstrated that academic success is associated with a decrease in problem behavior (Gottfredson, Gottfredson, & Skroban, 1996).

To further our understanding of the academic status of students with EBD, researchers have traditionally relied on conventional narrative reviews of the literature (e.g., Epstein, Kinder, & Bursuck, 1989; Trout et al., 2003). These reviews have found that students with EBD perform significantly below norms on standardized achievement tests (e.g., Coutinho, 1986) and achieve lower scores in mathematics than in reading (e.g., Schroeder, 1965; Stone & Rowley, 1964) and that older children with EBD are farther behind their same-age peers than are younger children with EBD (Couthino, 1986). Although these qualitative reviews have presented compelling evidence of the academic deficits experienced by children with EBD, they do not provide precise quantitative estimates of the magnitude of these deficits.

A quantitative review of the research on the academic performance of students with EBD compared to same-age, nondisabled peers or norm groups would extend our knowledge in this important area. Previous investigations have focused on grade-equivalent scores (e.g., Rosenblatt & Rosenblatt, 1999; Scruggs & Matropieri, 1986); however, these scores do not provide information on whether performance is within average limits, allow for normative comparisons, or allow for comparisons across studies. One statistical method for analyzing a body of research to compare group performance is meta-analysis. Meta-analytic reviews go beyond traditional narrative reviews in the sense that they are more systematic, more explicit, and rely on quantitative methods of analysis (Rosenthal, 1984). Because of these features, meta-analytic reviews are considered to provide more thorough, comprehensive, and precise summative evaluations that entail greater objectivity than narrative reviews (Rosenthal, 1984). The purpose of this study was to use a meta-analytic approach to quantitatively estimate the magnitude of difference in academic performance between students with EBD and their same-age, nondisabled peers or norm groups. The use of meta-analysis allows for a precise estimate of achievement difference, which allows for normative comparisons, provides a metric that is consistent across age levels, and allows for comparisons across studies. Moreover, meta-analysis is consistent with the American Psychological Association's guidelines calling for the use of effect sizes, which allows for an evaluation of the practical significance of differences (as opposed to statistical significance).

This study focused on the following. First, we hoped to determine whether there were differences in academic status across age, gender, race, and socioeconomic status. Age is an important factor because academic difficulties of children with EBD are thought to increase over time (e.g., Coutinho, 1986). Gender, also important, is an understudied area. Despite the fact that girls constitute approximately 20% of children with EBD (U.S. Department of Education, 2001), little is known about possible gender differences in academic status. Similarly, achievement differences among racial/ethnic groups (Meece & Kurtz-Costes, 2001) and among individuals from different socioeconomic levels (Dixon-Floyd & Johnson, 1998) have been well documented and may occur within EBD groups. Second, we examined the possibility of differences across academic subject areas. Differences in achievement across academic subject areas would have direct implications for instruction of students with EBD. Third, we investigated placement setting as a possible moderator of academic status. Little is known about the effects, if any, of setting on academic status of students with EBD. Given the increasing trend towards inclusion of students with EBD in less restrictive settings, comparisons of academic outcomes across setting is especially relevant. Finally, we investigated whether method of identification (e.g., school identified, *DSM* [*Diagnostic and Statistical Manual of Mental Disorders;* American Psychiatric Association, 2000] identified) had effects on academic status. Students identified via different criteria might differ in severity. For example, children identified via *DSM* criteria might exhibit more serious behavior problems and be placed in much more restrictive settings. Thus, different identification criteria could result in qualitatively different subgroups of children who differ in academic achievement.

Method

Study Selection Criteria

In this meta-analysis, we used a five-step literature search strategy to identify relevant studies published between 1961 and 2000. First, computer searches of PsychINFO and ERIC were conducted. The literature search terms used were *emotional disturbance, emotional and behavioral disorders, behavior disorder, conduct disorder, behavior and academic status, academic achievement, reading, math, arithmetic, written expression,* and *academics.* Second, manual searches of all issues of the *Journal of Emotional and Behavioral Disorders* and *Behavioral Disorders* were conducted. Third, we conducted ancestral searches of all identified articles. Fourth, articles previously reported in reviews of the literature were examined. Finally, researchers who had written on the academic status of children with EBD were contacted via e-mail and asked for assistance in locating further articles. This search strategy yielded a total of 205 articles potentially relevant to the meta-analysis.

Two graduate students screened the Method and Results sections of all 205 articles, to ensure that the articles included data from students with EBD and had reported academic achievement as a dependent variable. In the cases ($n = 6$) where a disagreement occurred, a third reader independently assessed the article and a decision was made to include or exclude the article. This screening resulted in the exclusion of 104 articles that did not specifically relate to the academic status of students with EBD or that did not provide data (e.g., practitioner pieces). Thus, 101 articles met the initial search criteria for inclusion in the review.

Inclusion Criteria

Inclusion criteria were chosen to ensure that the studies sampled accurately reflected the literature on the academic status of children and youth with EBD. Three graduate students read all 101 articles obtained during the preliminary search. For articles to be included in the literature review, each of the following conditions must have been met:

- The study was published in a peer-reviewed journal between 1961 and 2000. Only peer-reviewed studies were considered, as a means of overall quality control (Weisz, Weiss, Han, Granger, & Morton, 1995).
- Participants in the study were identified as EBD by one of the following methods: (a) identified through the school/Individuals with Disabilities Education Act (IDEA), (b) identified through *DSM-IV* (or prior versions) with conduct disorder or co-occurrence of EBD and another disability (e.g., attention-deficit/hyperactivity disorder), (c) currently being served in programs for severe behavior problems (e.g., psychiatric hospital, university clinic), or (d) identified as functioning in the clinical range through the use of behavior rating scales.
- The study included a mean score and standard deviation from a standardized test on at least one academic content area.
- The study sample consisted of children and youth between 5 and 21 years of age.

After applying these criteria, an additional 75 studies were excluded, leaving 26 studies that met the inclusion criteria for this meta-analysis. A number of factors resulted in the exclusion of the 75 studies:

1. Thirty-three studies were eliminated either because they were small sample size studies or because only grade-equivalent scores were used as the dependent measure. In the cases where grade-equivalent scores were provided, we were unable to derive a score suitable for meta-analysis, because scores were reported as aggregates across age ranges (e.g., a mean grade-equivalent score was reported for an age range).
2. Seventeen studies failed to provide data regarding specific academic characteristics of the students (e.g., academic status was determined by assessing school competence on behavior checklists).
3. Six studies presented subtest scores from IQ measures.
4. Eight studies used Individualized Education Program (IEP) goals and objectives as the dependent variable or reported only narrative data (e.g., poor performing, below grade level).
5. Nine studies did not disaggregate the scores of students with EBD from their peers with other disabilities.
6. Two studies reported only correlation between academic measures.

Data Analysis Strategy

Statistical Independence. We made several decisions during the literature review process based on commonly used meta-analytic literature review guidelines to reduce statistical dependencies (Cooper & Hedges, 1996). Because samples are often reported across more than one study, we reviewed each of the studies to ensure the independence of the samples. One study (Glassberg, Hooper, & Mattison, 1999) was eliminated because it used a sample that had been previously reported. In the end, each of the studies represented an independent sample.

Homogeneity Statistics. Hedges and Olkin's (1985) categorical fixed-effects model was used to evaluate which study characteristics, if any, were moderators of effect sizes. Using this approach, we selected a priori variables that were thought to be significant moderators of effect sizes and grouped the studies respectively (Durlak & Wells, 1998).

To determine whether the whole sample of effect sizes were homogeneous or heterogeneous, the homogeneity statistic Q_T was calculated (Hedges & Olkin, 1985). A significant Q_T would indicate that the effect sizes are not homogeneous across all the studies. This finding would suggest that the overall sample of effect sizes can be further partitioned into smaller groups according to a priori categorical dimensions. Afterwards, an omnibus between-class fit statistic (Q_B) and an omnibus within-class statistic (Q_W) was applied to the groupings (see Note).

The Q_B statistic tests whether the average effect sizes of each of the groupings are significantly different from each other, whereas the Q_W statistic tests for homogeneity of the effect sizes within each class. The best-case scenario is to obtain a Q_B statistic that is significant and a Q_W statistic that is not significant (Durlak & Wells, 1998). In the event that the Q_W statistic is statistically significant (indicating that the effect sizes within classes are not homogeneous), Hedges and Olkin (1985) suggested that it may be informative to partition the within-class fit statistic Q_W into individual study characteristics. In so doing, subgroupings with especially poor fits (i.e., effect sizes that are heterogeneous) are revealed, as is information regarding the possibility of other potentially more informative classification dimensions.

Estimation of Grand Effect Size. MetaWin 2.1 (Rosenberg, Adams, & Gurevitch, 2000) was used to assess the grand effect. MetaWin 2.1 allows researchers to summarize the results of multiple studies using meta-analytic procedures. In this meta-analysis, we included a number of studies that compared a group of children with EBD to the normative sample of a standardized instrument. In these cases, we set the control group size equal to the size of the EBD group to ensure equal weighting. Because most studies compared the EBD group to a normative group, it was decided that a pooled standard deviation would be appropriate. A pooled standard deviation also provides a better estimate of the population standard deviation (Rosenthal, 1996). Moreover, each ES was weighted by the inverse of its variance, a procedure that gives more weight to effect sizes emerging from studies with larger samples (Durlak & Wells, 1998; Hedges & Olkin, 1985). Each effect size comparison between the EBD group and a normative or comparison group was expressed in terms of the standardized difference between mean scores divided by the average population standard deviation (Rosenthal, 1995). The formula used to derive effect sizes was defined as follows:

$$ES = (M_{\mathrm{ED}} - M_{\mathrm{C}})/SD_{\mathrm{pooled}}$$

Thus, a negative ES indicated that the EBD group was lower in achievement than was the control group or normative sample.

Moderator Effect Sizes. Each study was coded according to the domains of interest to assess whether there were meaningful effect size differences across these dimensions. We examined the following variables for potential moderators: (a) subject area (i.e., reading, math, spelling, or written expression), (b) academic setting (i.e., not reported, general education, resource room, self-contained room, separate school, residential treatment, home school, university/clinic, other, or combined settings), (c) age (i.e., under 12 years or 12 or more years), and (d) method of identification (i.e., school identified, *DSM* identified, or rating scale identified).

Coding Agreement. Three graduate students reviewed and coded each of the final 25 studies. Interrater agreement was assessed for 32% (n = 8) of the articles. To assess interrater agreement, two reviewers read and coded the same articles; their responses for each item were then compared, to determine agreement. In instances where coders disagreed, a third reviewer read and coded the article to determine congruence. Percentage of agreement was determined by dividing the number of agreements by the number of agreements plus disagreements and then multiplying by 100. Agreement ranged from 85% to 100%, with a mean of 98%.

Supplemental Data Analysis

Confidence Intervals. We calculated bootstrap confidence intervals around the grand mean and subordinate effects sizes. The bootstrap method was selected for the following reasons. First, resampling methods such as bootstrap can be used to estimate confidence intervals because they generate their own distributions. Using this method allowed freedom from the distribution assumptions of parametric tests (Rosenberg, Adams, & Gurevitch, 2000). Second, the bootstrap method is more conservative. This is important because of the use of normative comparisons, which may overstate the difference between groups; thus, a conservative method seemed appropriate. Third, because of the limited number of studies, there was a possibility of a violation of distributional assumptions. The use of bootstrap, which is a nonparametric test, avoids this problem. Interpretation of confidence intervals is straightforward. In the case of individual comparisons, means whose bootstrap confidence intervals do not contain zero differ significantly from zero at the .05 level (Cooper & Hedges, 1996; Durlak & Wells, 1998). This indicates that the average effect size of the control (or normative) group differs significantly from that of the EBD group. In the case of comparisons across multiple areas, means whose confidence intervals do not overlap indicate that these effect size means differ from each other significantly.

Results

Table 1 shows studies included in this analysis. The 25 studies included 2,486 participants with EBD. The mean age and IQ for the EBD sample was reported in 17 studies. The weighted mean age across those studies was 11.22 years, and the weighted mean IQ was 94.89. The gender of the EBD participants was reported on 1,809 (82%) of the participants in 15 (71%) studies. Of those participants, 80% were boys and 20% were girls. Demographic information pertaining to race was available for 1,330 (61%) of the participants with EBD and was reported in 10 (48%) studies. Of these participants, 69% were Caucasian, 27% were African American, 3% were Hispanic, and 1% were of mixed backgrounds. Because we were unable to disaggregate the academic results based on race and gender, we could not analyze these factors. Due to the possibility that there were systematic differences between effect sizes derived from control groups (n = 49) versus those derived from normative comparisons (n = 52), we compared the results of the effect sizes derived from normative comparisons with those derived from control groups. Differences were not significant, so the two groups were combined for further analysis. We also tested for the possibility that there would be systematic differences across tests (e.g., effect sizes from the *Wide Range Achievement Test* might be higher than those of the *Woodcock-Johnson III Tests of Achievement*). With two exceptions, the

TABLE 1. Studies Included in Analysis

Authors	Participants/ identification	Student characteristics	Setting	Academic areas assessed	Dependent measures
Anderson, Kutash, & Duchnowski (2001)	*N* = 42, 36 boys, 6 girls school identified	Age = 5.90 yrs, IQ = 95.3	Special education resource, separate day school or clinic	Reading, arithmetic	KTEA, PIAT, WJ
Browne, Stotsky, & Eichorn (1977)	*N* = 510, 428 boys, 82 girls *DSM* identified	Age = 12.50 yrs, IQ = 92.0	Self-contained, separate day school, residential	Reading, arithmetic, and written expression	WRAT
Coutinho (1986)	*N* = 45 rating scale identified	Age = third grade to ninth grade at time of testing, IQ not reported	NA	Reading and vocabulary	SAT, ITBS
Duncan, Forness, & Hartsough (1995)	*N* = 85, 66 boys, 19 girls *DSM* identified	Age = 12.80 yrs, IQ = 94.0	Separate day school	Reading, arithmetic, and written expression	WJ
Foley & Epstein (1992)	*N* = 86, 78 boys, 8 girls school identified	Age = 13.30 yrs, IQ = 98.4	Self-contained	Reading, arithmetic, and written expression	QSAT
Forness, Youpa, Hanna, Cantwell, & Swanson (1992)	*N* = 41, 41 boys *DSM* identified	Age = 9.10 yrs, IQ = 105.0	Clinical outpatient	Reading and arithmetic	WRMT, PIAT, KMDT, *Gray Oral*
Glassberg (1994)	*N* = 252, 217 boys, 38 girls school identified	Age = 10.00 yrs, IQ = 100.0	Resource room, self-contained, separate day school	Reading, arithmetic, and written expression	WJ
Gresham, MacMillan, Bocian, Ward, & Forness (1998)	*N* = 130, group only for gender rating scale identified	Age = third grade, IQ = not reported	General education	Reading and arithmetic	CTBS, MAT, ITBS
Grizenko, Papineau, & Sayegh (1993)	*N* = 30, 27 boys, 3 girls *DSM* identified	Age = 9.00 yrs, IQ = 100	Separate day school	Reading, arithmetic, and written expression	WRAT
Grizenko & Sayegh (1990)	*N* = 23 *DSM* identified	Age = 9.00 yrs, IQ = not reported	Residential	Reading, arithmetic, and spelling	WRAT
Hodges & Plow (1990)	*N* = 76, 50 boys, 26 girls *DSM* identified	Age = 10.00 yrs, IQ = 97.5	University, psychiatric facility, hospital	Reading, arithmetic, and written expression	WJ
Lee, Elliott, & Barbour (1994)	*N* = 58, all boys school identified	Age = 14.40 yrs, IQ = 93.8	General education, residential	Reading, arithmetic, and written expression	WJ, PIAT, WRAT
Lopez, Forness, MacMillian, Bocian, & Gresham (1996)	*N* = 48, group only for gender school identified	Age = 7–12 yrs, IQ = not reported	General education	Reading, arithmetic, and written expression	WRAT
Mattison et al. (1986)	*N* = 109 school identified	Age = 11.40 yrs, IQ = 95	NA	Reading	WRAT

(table continues)

(Table 1 continued)

Authors	Participants/ identification	Student characteristics	Setting	Academic areas assessed	Dependent measures
Mattison, Spitznagel, & Felix (1998)	*N* = 111 school identified	Age = 12.10 yrs, IQ = 93	Self-contained, special school	Reading	WRAT
Meadows, Neel, Scott, & Parker (1994)	*N* = 19, all boys school identified	Age = 13.40 yrs, IQ = 90	Self-contained	Reading, arithmetic, and written expression	WJ
Naylor, Staskowski, Kenney, & King (1994)	*N* = 54, 20 boys, 34 girls *DSM* identified	Age = 14.00 yrs, IQ = not reported	University, psychiatric facility, hospital	Reading, arithmetic, and written expression	WJ
O'Brien et al. (1992)	*N* = 35, 29 boys, 6 girls school identified	Age = 9.30 yrs, IQ = 98.8	University, psychiatric facility, hospital	Reading, arithmetic, and written expression	WRAT
Rennick, Klinge, Hart, & Lennox (1978)	*N* = 14, gender not reported *DSM* identified	Age = 8.93 yrs, IQ = 95.6	Not identified	Reading, arithmetic, written expression, and other	PIAT
Richmond & Blagg (1985)	*N* = 30, gender not reported school identified	Age = 7–9, IQ = not reported	Resource room	Reading and arithmetic	WRAT
Rosenblatt & Rosenblatt (1999)	*N* = 61, 52 boys, 9 girls school identified	Age = 11.80 yrs, IQ = not reported	Separate day school	Reading, arithmetic, and written expression	WRAT
Sinclair, Forness, & Alexson (1985)	*N* = 350, 254 boys, 96 girls *DSM* identified	Age = 10.20 yrs, IQ = 94.0	General education, resource room, self-contained, separate day school, university, psychiatric facility, hospital	Reading and arithmetic	PIAT WRAT
Tramontana, Hooper, Curley, & Nardolillo (1990)	*N* = 50, 40 boys, 10 girls *DSM* identified	Age = 12.60 yrs, IQ = 94.3	University, psychiatric facility, hospital	Reading, arithmetic, and other	WJ
Wilson, Cone, Bradley, & Reese (1986)	*N* = 140, gender ratio not reported school identified	Age = 11.89 yrs IQ = 93.2	Separate day school	Reading, arithmetic, and written expression	WRAT
Zimet & Farley (1993)	N = 87, 61 boys, 26 girls *DSM* identified	Age = 9.33 yrs,	University, psychiatric facility, hospital	Reading, arithmetic, and written expression	WRAT

Note. KTEA = *Kaufman Test of Educational Achievement* (Kaufman & Kaufman, 1998); PIAT = *Peabody Individual Achievement Test–Revised* (Markwardt, 1998); WJ = *Woodcock-Johnson III Tests of Achievement–Revised* (Woodcock et al., 2001); *DSM = Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2000); WRAT = *Wide Range Achievement Test* (Wilkinson, 1993); SAT = *Stanford Achievement Test* (1995); ITBS = *Iowa Test of Basic Skills* (Hoover et al., 1996); QSAT = *Quick-Score Achievement Test* (Hammill et al., 1987); WRMT = *Woodcock Reading Mastery Tests* (Woodcock, 1998); KMDT = *Key Math: Diagnostic Arithmetic Test* (Connoly et al., 1971); *Gray Oral = Gray Oral Reading Tests–Fourth Edition* (Wiederholt & Bryant, 2001); CTBS = *Comprehensive Test of Basic Skills* (McGraw-Hill, 1990); MAT = *Metropolitan Achievement Test* (Harcourt Educational, 2001).

95% bootstrap confidence interval for all tests overlapped, indicating no significant differences across instruments. The exceptions were Anderson, Kutash, and Duchnowski (2001), with 2 effect sizes using the *Kauffman Test of Educational Achievement,* and Coutinho (1986) with 12 effect sizes using the *Stanford Achievement Test*. Because both exceptions were within single studies, there did not appear to be any significant effects due to instrumentation.

The 25-study database yielded 101 effect sizes. The weighted mean effect size for the sample was –0.6905 (*SD* =

.40), with a range of −3.371 to +0.503. According to Cohen (1988), an effect size of this magnitude denotes a moderate to large difference in the academic performance of students with EBD, compared to that of same-age peers without disabilities. Figure 1 illustrates the distribution of effect sizes. With the exception of one outlier, the distribution is essentially normal. Of the 101 effect sizes, 90 were negative, indicating that in 89% of the comparisons, students with EBD performed lower in academics than did nondisabled control or norm groups. The grand effect size was also significantly different from zero (95% bootstrap confidence interval = −.81 to −.57). Thus, overall academic performance was significantly lower for the EBD group. Finally, the overall homogeneity statistic was significant (Q_T = 704.96, $p < .05$) denoting the possibility of moderator variables.

Moderators

Table 2 provides data on the number of effect sizes for each moderator (*k*), weighted mean effect size for each moderator (d_+), confidence intervals, and Q_{wi} (which indicates whether or not effect sizes within group were homogenous). These analyses were conducted using procedures recommended by Hedges and Olkin (1985) and Rosenburg et al. (2000) and computed using MetaWin. According to Durlak and Wells (1998), a well specified fixed-effects model is reflected by a significant Q_B, but a nonsignificant Q_{wi}, for each class within moderator.

Academic Subjects. Academic subject areas possessed sufficient heterogeneity (Q_B = 46.64, $p < .05$) among effect sizes to justify subcategories (e.g., reading, math). Overall, students with EBD performed significantly lower than the nondisabled or norm groups across all subjects and all subject area categories were statistically different from zero. The largest effect sizes were obtained in mathematics (−.81) and spelling (−.81). The bootstrap confidence intervals for all of the academic subjects overlapped, denoting no statistical difference in the academic performance of students with EBD between and among subject areas. The within-group homogeneity statistics indicated, however, that the subject areas were not homogeneous. This finding indicated that heterogeneity greater than chance remained at the designated within-group subject category. Thus, subject matter as categorized did not explain all the variability in effect sizes.

Setting. As with the academic subjects, setting contained sufficient variability among effect sizes to justify subcategories (Q_B = 150.71, $p < .05$). All placement settings were statistically different from zero, which would suggest that students with EBD performed significantly lower than comparisons, regardless of setting. The largest effect sizes were demonstrated by students in residential facilities demonstrated the largest (−1.49), followed by those who did not report this information (−1.04). The smallest effect sizes were found in resource rooms (−.33). With the exception of residential and self-contained settings, the within-group homogeneity statis-

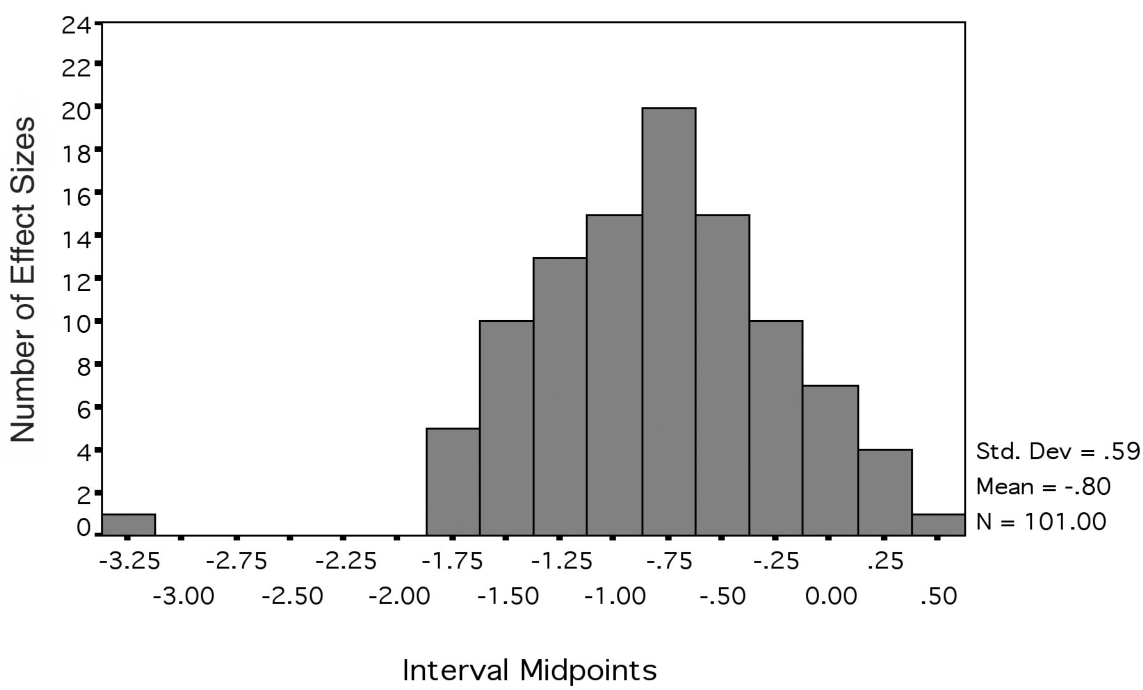

FIGURE 1. Distribution of effect sizes.

tics were significant, indicating that—with the exception of residential and self-contained settings—setting as categorized did not explain all the variability in effect sizes. However, the residential setting only encompassed one study, so this should be viewed with caution as the results are likely due to the fact that effect sizes were derived from a single sample.

Also, with the exception of the self-contained/university clinic comparison, bootstrap confidence intervals among the remaining placement comparisons overlapped, denoting no statistical difference between and among them. The bootstrap confidence interval for self-contained and residential students did not overlap with any other setting, indicating that students with EBD in self-contained classrooms and residential settings performed significantly lower in academics than did students in other settings.

Age. To test the effect of age on academic performance, the participants from each study were assigned to one of two groups, based on whether the reported mean age for each sample was 12 years and older or below 12 years. The two age groups demonstrated sufficient variability ($Q_B = 38.88$, $p < .05$) among effect sizes to justify the use of subcategories (i.e., 12 years or older, younger than 12 years). Both older and younger students with EBD were statistically significant from zero, which suggests that students with EBD performed significantly lower in academics than did the same age nondisabled control or norm groups. As in subject area and setting, both age categories possessed within-group effect size variability beyond chance, indicating that age as categorized did not explain all the variability in effect sizes. Bootstrap confidence intervals for both age groupings overlapped, indicating

TABLE 2. Summary of Effect Size Moderator Statistics

Characteristic	k	d_+	95% bootstrap CI	Q_{wi}	Omnibus
Subject area					
Reading	48	−0.61	−0.79 to −0.49	357.07*	
Math	30	−0.81	−0.99 to −0.57	167.13*	
Spelling	12	−0.81	−1.05 to −0.39	86.63*	
Written expression	11	−0.46	−0.79 to −0.24	47.48*	
					Q_w 658.32*
					Q_B 46.64*
Setting					
General education	10	−0.54	−0.80 to −0.29	48.80*	
Resource room	5	−0.33	−0.77 to −0.21	70.40*	
Self-contained	12	−0.83	−0.92 to −0.76	4.15	
Separate school	18	−0.62	−0.89 to −0.46	65.87*	
University clinic	22	−0.47	−0.66 to −0.28	92.47*	
Combined	8	−0.76	−1.04 to −0.44	160.34*	
Residential	3	−1.49	−1.64 to −1.29	.60	
Not reported	23	−1.04	−1.29 to −0.93	111.65*	
					Q_w 554.26*
					Q_B 150.71*
Age					
12 years or more	34	−0.82	−1.00 to −0.60	257.69*	
Less than 12 years	67	−0.60	−0.74 to −0.48	408.39*	
					Q_w 666.08*
					Q_B 38.88*
Method of identification					
School	42	−.47	−0.58 to −0.31	220.21*	
DSM	39	−.74	−0.95 to −0.57	279.89*	
Rating scale	20	−1.12	−1.48 to −0.94	63.98*	
					Q_w 564.08*
					Q_B 140.88*

Note. k = number of effect sizes; d_+ = mean effect sizes; CI = confidence interval; Q_{wi} = omnibus within-class fit statistic; Q_w = omnibus within-class fit statistic; Q_B = omnibus between-class fit statistic; *DSM* = *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2000).

no statistical difference between ages in the performance of students with EBD.

Method of Identification. To test for effects due to method of identification, we employed four categories based on identification methods in the studies: school identified, *DSM*-identified, rating scale–identified, described as receiving services for EBD. The diagnostic groups demonstrated sufficient variability ($Q_B = 140.88$, $p < .05$) among effect sizes to justify the use of subcategories. Students in all three categories were statistically significant from zero, which suggests that students with EBD in all categories performed significantly lower in academics than did the same age nondisabled control or norm groups. All three categories possessed within-group effect size variability beyond chance, indicating that identification method as categorized did not explain all the variability in effect sizes. Bootstrap confidence intervals for school-identified and *DSM*-identified groupings overlapped, indicating no statistical difference in the performance of these students. Rating scale–identified students differed significantly from school-identified, but not *DSM*-identified, students. However, because the rating scale–identified groups were included in only two studies, some caution is in order when generalizing these results.

Discussion

The purpose of this investigation was to quantitatively synthesize the research on the academic status of students with EBD, to determine the magnitude of difference in performance when compared to students without disabilities. In this section we discuss participant characteristics and problems with generalization, the overall level of academic deficits of children with EBD, academic deficits across subject areas and settings, and achievement deficits across age groups.

Participant Characteristics

Problems with thorough reporting of characteristics of samples have been noted in previous reviews of academic status of children with EBD (e.g., Mooney et al., 2003; Trout et al., 2003). Although more than 2,000 students were represented in this study, it is difficult to assess how representative these students were of the total EBD population. Race, gender, and SES can be potentially important moderators; however, the extent to which these data were reported was less than optimal. Nearly 30% of studies did not provide information on the gender of participants, and no study provided disaggregated data on female participants. As a result, we could not assess whether gender was a significant moderator of achievement. This is a serious gap in our knowledge base, because at present, approximately 20% of children with EBD are girls (U.S. Department of Education, 2001). Data on race and ethnicity were reported in fewer than half the studies. Again, this represents a serious omission. Minority groups make up more than 25% of the U.S. population (U.S. Census Bureau, n.d.). African Americans account for slightly more than 12% of the U.S. population, and between 1990 and 2000, the percentage of Hispanics in the United States doubled (U.S. Census Bureau, n.d.). These groups also constitute a significant proportion of the EBD population. African American and Hispanic groups constitute 27.3% and 8.9% of the EBD population in the United States, respectively (U.S. Department of Education, 2001). The lack of information on SES is also troubling. Knowledge of SES is important because low SES is a well-recognized risk factor for EBD (Cullinan, 2002). Another problem lies in the nature of the EBD groups sampled. Most of the studies appear to have reported convenience samples. None of the studies used epidemiological methods. Thus, we cannot be certain to what extent the students included in this analysis are representative of the current EBD population.

Overall Achievement

The findings demonstrate that there is a moderate to large (−.69) overall difference in the academic performance of students with EBD, compared to students without disabilities. This is consistent with previous research which has noted persistent problems with low achievement (e.g., Kauffman, 2001; Kauffman et al., 1987). The use of effect sizes allow for an estimate of the proportion of children in the EBD group who score below the mean of contrast group. In this case, 75% of students in the EBD group scored below the mean of the contrast group. The overall mean achievement level of the EBD group is at the 25th percentile.

Interpretation of the overall degree of academic deficits among children with EBD should be done in light of two factors. First, and most important, we have no information on the exact timing of the academic assessments or on instructional programming. There could well be significant differences in the degree of academic deficits between children whose data were taken at intake and those who had been previously identified and provided with effective instructional programming for a period of time. Hopefully, the degree of deficit would be less if an effective instructional program was in place. Second, we have no information on age of identification. It is possible that there would be significant differences between children who were identified and treated early.

Academic Subject Area

Students with EBD performed significantly below their peers in all academic subject areas. Though problems with reading and literacy have been the focus of public concerns and research, the two areas with the greatest absolute deficits were math and spelling; however, differences across the academic areas were not significant. This was somewhat surprising. One possible explanation for the lack of difference is that it reflects the increased emphasis now placed on reading and written expression. This does not appear to be the case, because studies reporting reading data were evenly distributed over the time span reviewed, so factors related to increased or improved reading instruction are unlikely to have resulted in the observed differences. It is also possible that placement differences or age differences may have affected results. Because of the possibility that differences across academic subject areas could result from systematic differences across placement and subject area (e.g., if there were more math effect sizes from self-contained settings or from older students), we conducted a post hoc chi-square analysis. Results of this analysis were not significant for academic subject by setting, $\chi^2(21, N = 80) = 30.97$, $p = .07$. Thus, differences in settings would not account for the lack of differences across subject areas. Age by subject area results were significant, $\chi^2(3, N = 101) = 8.12$, $p = .043$. However, this was due to a disparity within one area, spelling, for which younger children had 11 effect sizes, compared to a single effect size for older children. Another possible explanation is that inadequate math skills are more influenced by limited classroom exposure than reading (Forness, Bennett, & Tose, 1983; Naylor, Staskowski, Kenney, & King, 1994).

The size of the deficit in spelling was somewhat surprising because research has suggested that there is a strong relation between reading and spelling; correlations between reading and spelling achievement range from .50 to .90 (Graham, Harris, & Chorzempa, 2002). However, Treiman (1993) noted that some children who experience difficulty spelling have little difficulty with reading. Age of participants is also a potential factor. Of the 12 spelling effect sizes, 11 were from the group of children younger than 12 years. There is some evidence that spelling develops relatively slowly in low-achieving students (O'Malley, Francis, Foorman, Fletcher, & Swank, 2002). Additionally, deficits in spelling may be explained by limited short-term memory span common in children with attention and behavior problems (Levy & Hobbes, 1989). However, the fact that the Q statistic was significant indicates that subject areas alone do not explain all the observed variability.

It is interesting to note that the observed deficits in achievement in the present study were consistent with those of reported by Kavale and Nye (1985–1986) for children with learning disabilities (LD). Kavale and Nye found overall differences in academic achievement of .683 between a control group and children with LD. Across subject areas, differences between the children with LD and the control group for comparable academic subject areas were as follows: reading, .757; math, .607; spelling, .726. This suggests that the major difference between LD and EBD groups may lie in reading and that there may be little difference between LD and EBD in terms of math and spelling achievement. This is consistent with current conceptualization of LD—that reading is the core deficit area for children with LD, due to an underlying deficit in phonological processing (Adams, 1996; National Research Council, 1998). Thus, children with EBD may have a slightly different profile from children with LD in that they are somewhat less likely to demonstrate underachievement in reading.

Setting Differences

Across general education, resource room, self-contained, and special school settings, there was no significant difference in academic performance. This finding was surprising. One might expect that students placed in general education would have higher academic performance than students in resource rooms or more restrictive setting. However, this does not appear to be the case. We would caution, however, that the lack of difference may be due to the variability demonstrated within each setting.

For students in self-contained and residential settings, the Q_W statistic was not significant, suggesting that no other moderators were necessary to explain the observed variability. We would caution that although the data suggest homogeneous effect sizes, we cannot determine exactly why this occurred. There are three plausible explanations for the observed result. First, the residential setting comprised only one study with a small sample ($n = 23$). Thus, for this group, the results may be a statistical artifact. Second, it is possible that students in self-contained settings and residential settings are homogeneous. We would expect that students in self-contained and residential settings would generally demonstrate more serious behavior problems. The onset, frequency, and persistence of behavior problems in children is related to poor academic performance (McEvoy & Welker, 2000). Thus, it is possible that the students in these settings are a more homogeneous group in terms of academic deficits. This could be seen as supporting the need for self-contained placement options. A third explanation is that the lack of significant variation among self-contained and residential groups is due to the setting. The students in these settings may not receive consistent, effective academic instruction, perhaps because of a focus on behavior management strategies. From this perspective, academic deficits are the result of placement in a setting that produces lower academic performance. Again, we caution that the lack of information on programming precludes any firm conclusions. It is also important to note that some effect sizes reflected multiple settings that included self-contained groups and that there were no significant differences across general education, resource, and self-contained settings.

Age Differences

The difference in academic achievement across age groups was not statistically significant, and the overlap of confidence intervals across the groups was pronounced. This was surprising, as previous research has suggested that underachievement tends to increase with age (Coutinho, 1986). And, once again, the significant Q_W statistic indicates that age alone is insufficient to explain observed variability. There is little longitudinal data on the academic status of children with EBD. Only one study reported longitudinal data using standard scores (Anderson et al., 2001), and this study did not find significant decreases over a 4-year time span. One explanation for the lack of difference across age groups could be due to the fact that many previous studies have reported deficits in terms of grade-equivalent scores. Coutinho's study is often cited as evidence that academic deficits increase over time. However, Coutinho based this conclusion on the increase in grade-equivalent differences over time and did not actually test for differences. The problems inherent in grade-equivalent scores are well documented (e.g., Linn & Gronlund, 2000). Grade-equivalent scores are less reliable with older children, are not directly comparable across age groups, and can decline independent of any real change in academic status (Venn, 2004). For example, a student who is 1 year below grade level at age 6 and a student who is 3 years below grade level at age 16 may have equal deficits in terms of their rank in the population. This could explain why our results differed from Coutinho's. It is important to note that in our study, groups' age scores were based on the mean reported age and that the comparisons made are not within-group. Rather, they are across groups that may differ across potentially important variables (e.g., SES, gender) that were not assessed in this analysis.

Method of Identification

Students across all three methods of identification demonstrated significant deficits in achievement. There was a significant difference between school-identified and rating scale–identified students, which suggests that academic status differs across diagnostic methods. This should be viewed with caution, however, because only two studies used rating scale identification. There was no significant difference across the two largest groups—school-identified and *DSM*-identified students. This was surprising considering that many of the *DSM*-identified groups were in hospitals or psychiatric facilities. One might expect that students in hospitals or psychiatric facilities would demonstrate concomitant levels of impairment in academics. Several factors temper interpretation of this result, however. First, the confidence interval overlap across the two groups was minimal. Second, we used very conservative methods to assess significance. Finally, the within-group Q statistic was significant for all three groups, indicating that identification method does not explain all the variability in effect sizes.

Limitations

Despite an exhaustive literature search, we were able to locate only 25 studies that met criteria for the meta-analysis. A more refined picture may be possible with more studies. The fact that we were able to locate only 25 studies demonstrates the need for more research in this area. Including studies that used curricular-based measures would have increased the number of studies slightly; however, our concern was overall academic status. We did not think it appropriate to equate the standardized and curriculum-based measures. Given the small number of studies, several caveats common to meta-analytic studies should be considered when interpreting the results. First, only one within-group Q_W statistic was significant; thus, additional moderators or different categorizations or combinations of moderators are necessary to explain the variability present. Second, given the ratio of effect sizes to studies (i.e., 101:25), statistical dependencies could have accounted in part for the findings, because multiple effect sizes were calculated for the same participants. Third, because each effect size was treated as an individual unit, the results may be skewed, as studies with larger sizes were given greater weight. Fourth, due to the small number of effect sizes for some moderators (e.g., spelling or resource room setting), we need to interpret these results with great caution. Fifth, the participants included were likely highly heterogeneous, due to such factors as variation in definitions of EBD across states. This likely served to increase variance across studies. Finally, all the studies reported were cross-sectional "snapshots" of academic performance. This study cannot address within-student changes over time.

Future Research

Our findings highlight several areas in need of future research. First, the fact that we were able to locate only 25 studies with 101 effect sizes that provided academic data indicates that additional research aimed at academic achievement among children with EBD is necessary. A similar meta-analysis by Kavale and Nye (1985–1986) assessing the achievement of students with LD generated 260 effect sizes. Second, inadequate demographic reporting limited our understanding of the academic characteristics of students with EBD. This problem appears to be endemic. For example, in a review of 66 studies assessing the academic status of students with EBD, Trout and colleagues (2003) found that only 44% of the studies included demographic information pertaining to race, and few of those studies reported students' academic performance across specific subcategories (i.e., gender, race, and age groups). Similar problems were noted by Mooney et al. (2003). Given the concerns with overidentification of minorities and the increasing identification of girls with EBD, researchers need to provide detailed demographic information and disaggregated data that will allow analysis of subgroups. Third, researchers have practiced an overreliance on measuring academic performance by age- and grade-equivalent scores (e.g., Rosenblatt & Rosenblatt, 1999; Scruggs & Matropieri, 1986). Although age- and grade-equivalent scores may be easier to interpret, they do not provide information on whether the child's performance is within average limits, allow for normative comparisons, or allow for comparisons across studies. Finally, we looked at academic performance in a broad context. Future research is needed to examine the academic performance of students with EBD within specific academic skill sets, such as reading comprehension and mathematical reasoning.

Implications

The findings from our study suggest that students with EBD are in need of additional effective academic instruction across all subject areas, settings, and age groups. From a research perspective, these findings emphasize the need for additional research to examine causal factors related to academic underachievement, academic performance within specific subject areas, and the long-term impact of underachievement in specific subject areas. The relationship between placement and achievement and the effects of gender and ethnicity should also be scrutinized. Moreover, understudied academic areas, such as written language and content areas, should receive attention. Reporting standards for participants should also be improved, to allow for analysis of important moderator variables. From a practitioner's perspective, the findings from this study emphasize the need for teachers to continue to monitor and measure the academic performance of students with EBD and to use scientifically based academic interventions that address deficits across all academic subjects and settings. There is also a need to attend to math and spelling, which may be serious deficit areas. From an administrator's perspective, the findings underscore the need to train teachers and paraprofessionals to use scientifically based instructional programs to positively affect the academic performance of students with EBD. From a policy perspective, the findings from this study emphasize the need for policies that encourage the measurement of academic performance and outcomes at every stage of development to determine the effectiveness of interventions.

AUTHORS' NOTES

1. Preparation of this article was supported in part by Doctoral Leadership (H325D990035), Post-Doctoral Leadership (H325D0016), and Research and Demonstration Center for Behavior (H324X010010) grants from the U.S. Department of Education. Opinions expressed do not necessarily reflect the position of the U. S. Department of Education, and no endorsement should be inferred.
2. We extend our thanks to Renae Oestmann for her contributions to the preparation of this manuscript.

NOTE

The fit statistics Q_T, Q_B, and Q_W are analogous to the partitioning of the total variance of effects into the variance occurring between [Q_B] and within [Q_W] groups such that $Q_T = Q_B + Q_W$.

REFERENCES

Adams, M. J. (1996). *Beginning to read.* London: MIT Press.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

Anderson, J. A., Kutash, K., & Duchnowski, A. J. (2001). A comparison of the academic progress of students with EBD and students with LD. *Journal of Emotional and Behavioral Disorders, 9,* 106–115.

Brier, N. (1995). Predicting antisocial behavior in youngsters displaying poor academic achievement: A review of risk factors. *Journal of Developmental & Behavioral Pediatrics, 16,* 271–276.

Browne, T., Stotsky, B. A., & Eichorn, J. (1977). A selective comparison of psychological, developmental, social, and academic factors among emotionally disturbed children in three treatment settings. *Child Psychiatry and Human Development, 7,* 231–253.

Cohen, J. (1988). *Statistical power for the behavioral sciences* (2nd ed.). New York: Academic Press.

Connoly, A. J., Nachtman, W., & Pritchett, E. M. (1971). *Key math: Diagnostic arithmetic test.* Circle Pines, MN: American Guidance Service.

Cooper, H., & Hedges, L. V. (1996). *The handbook of research synthesis.* New York: Russell Sage.

Coutinho, M. J. (1986). Reading achievement of students identified as behaviorally disordered at the secondary level. *Behavioral Disorders, 11,* 200–207.

Cullinan, D. (2002). *Students with emotional and behavior disorders: An introduction for teachers and other helping professionals.* Upper Saddle River, NJ: Merrill/Prentice Hall.

Dixon-Floyd, I., & Johnson, S. W. (1998). Variables associated with assigning students to behavioral classrooms. *Journal of Educational Research, 91,* 123–127.

Duncan, B., Forness, S. R., & Hartsough, C. (1995). Students identified as seriously emotionally disturbed in school-based day treatment: Cognitive, psychiatric, and special education characteristics. *Behavioral Disorders, 20,* 238–252.

Durlak, J. A., & Wells, A. M. (1998). Evaluation of indicated preventive intervention (secondary prevention) mental health programs for children and adolescents. *American Journal of Community Psychology, 26,* 775–802.

Epstein, M. H., Kinder, D., & Bursuck, B. (1989). The academic status of adolescents with behavioral disorders. *Behavioral Disorders, 14,* 157–165.

Foley, R. M., & Epstein, M. H. (1992). Correlates of the academic achievement of adolescents with behavioral disorders. *Behavioral Disorders, 18,* 9–17.

Forness, S., Bennett, L., & Tose, J. (1983). Academic deficits in emotionally disturbed children revisited. *Journal of the American Academy of Child Psychiatry, 22,* 140–144.

Forness, S. R., Youpa, D., Hanna, G. L., Cantwell, D. P., & Swanson, J. M. (1992). Classroom instructional characteristics in attention deficit hyperactivity disorder: Comparison of pure and mixed subgroups. *Behavioral Disorders, 17,* 115–125.

Glassberg, L. A. (1994). Students with behavioral disorders: Determinants of placement outcomes. *Behavioral Disorders, 19,* 181–191.

Glassberg, L. A., Hooper, S. R., Mattison, R. E. (1999). Prevalence of learning disabilities at enrollment in special education students with behavioral disorders. *Behavioral Disorders, 25,* 9–21.

Gottfredson, D. C., Gottfredson, G. D., & Skroban, S. (1996). A multimodel school-based prevention demonstration. *Journal of Adolescent Research, 11,* 97–115.

Graham, S., Harris, K. R., & Chorzempa, B. F. (2002). Contribution of spelling instruction to the spelling, writing, and reading of poor spellers. *Journal of Educational Psychology, 94,* 669–686.

Gresham, F. M., MacMillan, D. L., Bocian, K. M., Ward, S. L., & Forness, S. R. (1998). Comorbidity of hyperactivity-impulsivity-inattention and conduct problems: Risk factors in social, affective, and academic domains. *Journal of Abnormal Child Psychology, 26,* 393–406.

Grizenko, N., Papineau, D., & Sayegh, L. (1993). Effectiveness of a multimodal day treatment program for children with disruptive behavior disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 32,* 127–134.

Grizenko, N., & Sayegh, L. (1990). Evaluation of the effectiveness of a psychodynamically oriented day treatment program for children with behavior problems: A pilot study. *Canadian Journal of Psychiatry, 35,* 519–525.

Hammill, D. D., Ammer, J., Cronin, M. E., Mandlebaum, L. H., & Quinby, S. S. (1987). *Quick-score achievement test.* Austin, TX: PRO-ED.

Harcourt Educational. (2001). *Metropolitan achievement test.* San Antonio, TX: Author.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* New York: Academic Press.

Hinshaw, S. P. (1992). Externalizing behavior problems and academic underachievement in childhood and adolescence: Causal relationships and underlying mechanisms. *Psychological Bulletin, 111,* 127–155.

Hodges, K., & Plow, J. (1990). Intellectual ability and achievement in psychiatrically hospitalized children with conduct, anxiety, and affective disorders. *Journal of Consulting and Clinical Psychology, 58,* 589–595.

Hoover, H. D., Hieronymous, A. N., Frisbie, D. A., & Dunbar, S. B. (1996). *Iowa tests of basic skills.* Itasca, IL: Riverside.

Individuals with Disabilities Education Act Amendments of 1997, 20 U.S.C. 1401 (26).

Kaufman, A. S., & Kaufman, N. L. (1998). *Kaufman test of educational achievement.* Circle Pines, MN: American Guidance Service.

Kauffman, J. M. (2001). *Characteristics of emotional and behavioral disorders in children and youth.* (7th ed.). Upper Saddle River, NJ: Merrill/Prentice Hall.

Kauffman, J. M., Cullinan, D., & Epstein, M. H. (1987). Characteristics of students placed in special programs for the seriously emotionally disturbed. *Behavioral Disorders, 12,* 175–184.

Kavale, K. A., & Nye, C. (1985–1986). Parameters of learning disabilities in achievement, lingustic, neuropsychological, and social-behavioral domains. *The Journal of Special Education, 19,* 443–458.

Lee, S. W., Elliott, J., & Barbour, J. D. (1994). A comparison of cross-informant behavior ratings in school-based diagnosis. *Behavioral Disorders, 19,* 87–97.

Levy, F., & Hobbes, G. (1989). Reading, spelling, and vigilance in attention deficit and conduct disorder. *Journal of Abnormal Child Psychology, 17,* 291–298.

Linn, R. L., & Gronlund, N. E. (2000). *Measurement and assessment in teaching* (8th ed.). Upper Saddle River, NJ: Merrill/Prentice Hall.

Lopez, M. F., Forness, S. R., MacMillan, D. L., Bocian, K. M., & Gresham, F. M. (1996). Children with attention deficit hyperactivity disorder and emotional or behavioral disorders in primary grades: Inappropriate placement in the learning disorder category. *Education and Treatment of Children, 19,* 286–299.

Maguin, E., & Loeber, R. (1996). How well do ratings of academic performance by mothers and their sons correspond to grades, achievement test scores, and teachers' ratings. *Journal of Behavioral Education, 6,* 405–425.

Markwardt, F. C. (1998). *Peabody individual achievement test–Revised.* Circle Pines, MN: American Guidance Service.

Mattison, R. E., Humphrey, F. J., Kales, S. N., Hanford, M. A., Finkenbinder, R. L., & Hernit, R. O. (1986). Psychiatric background and diagnoses of children evaluated for special class placement. *Journal of the American Academy of Child Psychiatry, 25,* 514–520.

Mattison, R. E., Spitznagel, E. L., & Felix, B. C. (1998). Enrollment predictors of the special education outcome for students with SED. *Behavioral Disorders, 23,* 243–256.

McEvoy, A., & Welker, R. (2000). Antisocial behavior, academic behavior, and school climate: A critical review. *Journal of Emotional and Behavioral Disorders, 8,* 130–140.

McGraw-Hill. (1990). *Comprehensive test of basic skills.* Monterey, CA: CTB Macmillan/McGraw-Hill.

Meadows, N. B., Neel, R. S., Scott, C. M., & Parker, G. (1994). Academic performance, social competence, and mainstream and nonmainstreamed students with serious behavioral disorders. *Behavioral Disorders, 19,* 170–180.

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

Mooney, P., Epstein, M. H., Reid, R., & Nelson, J. R. (2003). Status and trends of academic intervention research for students with emotional disturbance. *Remedial and Special Education, 24,* 273–287.

Morrison, G. M., & D'Incau, B. (1997). The web of zero-tolerance: Characteristics of students who are recommended for expulsion from school. *Education & Treatment of Children, 20,* 316–335.

National Research Council. (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

Naylor, M. W., Staskowski, M., Kenney, M. C., & King, C. A. (1994). Language disorders and learning disabilities in school-refusing adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry, 33,* 1331–1337.

O'Brien, J. D., Halperin, J. M., Newcorn, J. H., Sharma, V., Wolf, L., & Morganstein, A. (1992). Psychometric differentiation of conduct disorder and attention deficit disorder with hyperactivity. *Developmental and Behavioral Pediatrics, 13,* 274–277.

O'Malley, K. J., Francis, D. J., Foorman, B. R., Fletcher, J. M., & Swank, P. R. (2002). Growth in precursor and reading-related skills: Do low-achieving and IQ-discrepant readers develop differently? *Learning Disabilities Research and Practice, 17,* 19–34.

Rennick, P. M., Klinge, V., Hart, Z., & Lennox, K. (1978). Evaluation of intellectual, linguistic, and achievement variables in normal, emotionally disturbed, and learning disabled children. *Adolescence, 52,* 755–766.

Richmond, B. O., & Blagg, D. E. (1985). Adaptive behavior, social adjustment, and academic achievement of regular and special education children. *The Exceptional Child, 32,* 93–98.

Rosenberg, M. S., Adams, D. C., & Gurevitch, J. (2000). *MetaWin* (Version 2.1) [Computer software]. New York: Sinauer Associates.

Rosenblatt, J. A., & Rosenblatt, A. (1999). Youth functional status and academic achievement in collaborative metal health and education programs: Two California care systems. *Journal of Emotional and Behavioral Disorders, 7,* 21–30.

Rosenthal, R. (1984). *Meta-analytic procedures for social research.* Beverly Hills, CA: Sage.

Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin, 188,* 183–192.

Rosenthal, R. (1996). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231–243). New York: Russell Sage.

Ruhl, K. L., & Berlinghoff, D. H. (1992). Research on improving behaviorally disordered students' academic performance: A review of the literature. *Behavioral Disorders, 17,* 178–190.

Schroeder, L. B. (1965). A study of the relationship between five descriptive categories of emotional disturbance and reading and arithmetic achievement. *Exceptional Children, 32,* 111–112.

Scruggs, T. E., & Mastropieri, M. A. (1986). Academic characteristics of behaviorally disordered and learning disabled students. *Behavioral Disorders, 11,* 184–190.

Sinclair, E., Forness, S. R., & Alexson, J. (1985). Psychiatric diagnosis: A study of its relationship to school needs. *The Journal of Special Education, 19,* 333–344.

Smith, S. W., & Coutinho, M. J. (1997). Achieving the goals of the National Agenda: Progress and prospects. *Journal of Emotional and Behavioral Disorders, 5,* 2–5.

Stanford Achievement Test. (1995). San Antonio, TX: Psychological Corp.

Stone, F. B., & Rowley, V. N. (1964). Educational disability in emotionally disturbed children. *Exceptional Children, 30,* 423–426.

Tramontana, M. G., Hooper, S. R., Curley, A. D., & Nardolillo, E. M. (1990). Determinants of academic achievement in children with psychiatric disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 29,* 265–268.

Treiman, R. (1993). Beginning to spell. New York: Oxford University Press.

Trout, A. L., Nordness, P. D., Pierce, C. D., & Epstein, M. H. (2003). Research on the academic status of children with emotional and behavioral disorders: A review of the literature from 1961–2000. *Journal of Emotional and Behavioral Disorders, 11,* 198–210.

U.S. Census Bureau. (n.d.). *United States Census 2000: Your gateway to Census 2000.* Retrieved March 11, 2003, from http://www.census.gov/main/cen2000.html

U.S. Department of Education. (2001). *Twenty-third annual report to Congress on the implementation of the Individuals with Disabilities Education Act.* Washington, DC: Author.

U.S. Department of Health and Human Services. (1999). *Mental health: A report of the Surgeon General.* Washington, DC: Author.

Venn, J. J. (2004). *Assessing students with special needs* (3rd ed.). Columbus, OH: Merrill/Prentice Hall.

Wagner, M. M. (1995). Outcomes for youths with serious emotional disturbance in secondary school and early adulthood. *The Future of Children, 5,* 90–111.

Weisz, J. R., Weiss, B., Han, S. S., Granger, D. A., & Morton, T. (1995). Effects of psychotherapy with children and adolescents revisited: A meta-analysis of treatment outcome studies. *Psychological Bulletin, 117,* 450–468.

Wiederholt, J. L., & Bryant, B. R. (2001). *Gray oral reading tests* (4th ed.). Austin, TX: PRO-ED.

Wilkinson, G. S. (1993). *Wide range achievement test* (3rd ed.). Wilmington, DE: Jastak Associates.

Woodcock, R. W. (1998). *Woodcock reading mastery tests–Revised.* Circle Pines, MN: American Guidance Service.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III tests of achievement* (Rev. ed.). Itasca, IL: Riverside.

Wilson, L., Cone, T., Bradley, C., & Reese, J. (1986). The characteristics of learning disabled and other handicapped students referred for evaluation in the state of Iowa. *Journal of Learning Disabilities, 19,* 553–557.

Zimet, S. G., & Farley, G. K. (1993). Academic achievement of children with emotional disorders treated in a day hospital program: An outcome study. *Child Psychiatry and Human Development, 23,* 183–202.