

Are IQ Scores Valid for Children Who Are Poor Readers?

Rex B. Kline, Susan A. Graham, and David Lachar

Remedial services for children with reading problems are often allocated according to discrepancies between reading and IQ scores. Results of some recent research suggest, however, that IQ scores of poor readers do not covary with their levels of functioning in other cognitive domains. This study evaluated whether the external validity of IQ scores (from the Wechsler Intelligence Scale for Children—Revised) was moderated by reading levels within 2 separate samples of referred children. We found that IQ scores had expected correlations with external measures of verbal, visual-spatial, short-term memory, and arithmetic ability, and that these relations were invariant across levels of reading skill.

Many school-age children have reading difficulties, but only some may be classified as reading disabled, which indicates a discrepancy between specific reading proficiency and general cognitive ability. Children so classified may be eligible for general learning disability services or for more specialized programs for the reading disabled. Poor readers who do not meet ability-achievement discrepancy definitions of reading disability may receive other remedial services but, in some instances, may receive none at all. For example, poor readers with IQ scores in the borderline range (e.g., 70-79 on Wechsler scales) may be ineligible for special education services because they are considered to be "slow learners." Poor readers with even lower IQ scores (e.g., 60-69), however, may be eligible for placement in classrooms for the educable mentally impaired, where they may at least receive more individual attention than "slow learners" who remain in regular classrooms.

Considering the implications for children's educational careers, it is crucial to demonstrate the validity of allocating special education resources on the basis of discrepancies between reading and general ability levels. There have been two types of criticisms about this method of allocation, however, the second of which provides the main focus of this study. First, as with the broader notion of learning disability, there are numerous conceptual and statistical problems with the operational definition of "significant" ability-reading discrepancies (e.g., see Kamphaus, Frick, & Lahey, 1991; Reynolds, 1984-1985). Second, there has been much debate about whether IQ tests provide valid estimates of general cognitive capacity for children with reading problems. For example, some researchers have argued that poor readers may obtain artificially low IQ scores because

IQ tests probably tap some of the same language processes involved in poor reading. Thus, poor readers could appear to be "slow learners" and thus fail to qualify for services (e.g., Siegel, 1989a, 1989b). A related criticism concerns the "Matthew effect," the possibility that reading problems lead to accumulated deficits in vocabulary, factual knowledge, and nonreading achievement, which in turn result in lower IQ scores (e.g., Stanovich, 1986, 1989, 1991).

There are empirical findings consistent with the above criticisms. For example, verbal IQ scores of poor readers seem to decline with age (e.g., Share & Silva, 1987), and early reading skills (at age 7) are significantly related to later measured general verbal intelligence (at age 13) even when initial differences in verbal ability are considered (e.g., Share, McGee, & Silva, 1989). Also, findings that correlations between IQ and reading scores increase with age (from about .40 in the early elementary grades to about .70 by early adulthood; Stanovich, 1986) are consistent with the hypothesis of reciprocal effects between reading and general cognitive ability as measured by IQ tests.

Even stronger evidence about the potential limited external validity of IQ tests with poor readers has been presented by Linda S. Siegel. For example, Siegel (1988) administered tests of memory, language, grammatical, and scholastic skills to regular and special education Canadian children who were classified as poor or normal readers. Readers of each type were also partitioned into subgroups on the basis of their IQ scores (<80, 80-90, 91-109, >110). Among the poor readers, performances of children with different IQ scores were similar across most measures. Siegel (1992) compared the performance of dyslexics (children whose reading achievement scores were > 15 standard score points below their IQ scores); poor readers (children with low reading and IQ scores); and normal controls across tasks similar to those used in her earlier study (Siegel, 1988). Dyslexics and normal controls had higher arithmetic scores than poor readers, but the dyslexics and poor readers had similar scores across most of the other tasks. Results of other recent studies also suggest limited differences between low- and normal-IQ poor readers across external cognitive domains (e.g., Fletcher, Francis, Rourke, Shaywitz, & Shaywitz, 1992; Pennington, Gilger, Olson, & DeFries, 1992; Shaywitz, Fletcher, Holahan, & Shaywitz, 1992). On the basis of these types of findings, Siegel

Rex B. Kline and Susan A. Graham, Department of Psychology, Concordia University, Montréal, Québec, Canada; David Lachar, Department of Psychiatry and Behavioral Sciences, University of Texas Medical School, Houston.

Portions of these results were presented at the 1990 Annual Meeting of the Canadian Psychological Association in Ottawa and the 1991 Annual Meeting of the American Psychological Association in San Francisco.

Correspondence concerning this article should be addressed to Rex B. Kline, Department of Psychology, Concordia University, Montréal, Québec, Canada H4B 1R6.

and others have argued that IQ scores should not be included in the operational definition of reading disability (e.g., Fletcher et al., 1989; Siegel, 1989a, 1989b).

Siegel's findings have obvious implications for the evaluation of discrepancy definitions of reading disability, but they also raise more general questions about the clinical uses of IQ tests. That is, IQ tests are routinely administered to children who are referred due to poor achievement, and many such children have histories of reading difficulties. If IQ tests have limited external validity for poor readers, then practitioners would need to exercise appropriate caution in their interpretations of test results. Whether Siegel's results support the latter, broader possibility is not entirely clear, however, given certain limitations of her methods. For example, although some children in Siegel's (1988, 1992) studies were administered the full Wechsler Intelligence Scale for Children—Revised (WISC-R; Wechsler, 1974), IQ levels for most cases were established using either scores from two WISC-R subtests (Vocabulary, Block Design) or from the Peabody Picture Vocabulary Test (PPVT; Dunn, 1959), a multiple-choice measure of receptive vocabulary. Neither procedure yields clinically acceptable substitutes for full-battery IQ scores. Another problem concerns expected loss of statistical power for analyses wherein quantitative IQ scores are used to partition cases into groups for comparison on other measures (e.g., Swanson, 1989). There are other, more powerful statistical procedures that retain the full quantitative information of IQ and reading scores and allow researchers to test whether children's reading status moderates the external validity of IQ scores.

In this study we evaluated whether IQ scores have comparable external validity for children of varying reading skill. Within two samples of referred children, we administered the WISC-R, reading achievement tests, and numerous other standardized, individually administered measures of verbal, visual-spatial, short-term memory, quantitative, and fine motor coordination ability that served as the external criteria for IQ scores. Readers should note that numerous other types of variables could be used as external criteria for study of the external validity of IQ scores, including, for example, classroom performance in reading and nonreading achievement areas, teacher reports, neuropsychological tasks, special education history, and eventual educational outcome. Thus, the focus of this study concerns the concurrent validity of IQ scores relative to other measures commonly administered to referred children in school settings. We used moderated multiple regression (i.e., with interaction terms) to test whether covariations between IQ scores and the external criteria varied as a function of children's reading skills. Because regression analyses yield little information about discrete groups of children, we also classified children as being either poor or normal readers and as having either low or normal IQ scores, and we compared the scores of these groups across the external criteria. Using these analyses we sought to determine whether poor readers of different IQ levels obtained different patterns of scores across the external criteria.

Method

Subjects

Data for this study comprised cognitive ability and achievement test scores obtained from two samples. Children in the school sample ($N =$

134) were enrolled in public schools (7 in total within 4 school systems) in western and eastern Canada and were referred by their teachers for psychological testing. These cases represent all children in these schools who were referred due to poor achievement over a 2-year period; children referred due to primary emotional-behavioral problems were not included in this sample. All children attended English-language schools and resided in suburban, middle-class areas. Their average age was 9.5 years ($SD = 1.9$ years, range = 7½–12 years), 72% were boys, and almost all were White (96%). The average WISC-R Full Scale IQ score of the entire school sample was 95.2 ($SD = 15.4$).

A total of 102 children (77%) in the school sample were not receiving special education services when they were tested; they were currently attending regular classrooms. All but 3 of the remaining 32 children were receiving learning disability remedial services (about 2–4 hours a week in small groups of about 8–12 children) but otherwise were mainstreamed in regular classrooms. Three children attended a self-contained learning disability classroom in one school. Readers should note that, as in the United States, the criteria for the classification of children as learning disabled varies according to region in Canada. Also, there was no common definition of *learning disability* (e.g., a specific IQ-achievement discrepancy) that was common to all of the schools attended by children in our school sample. Thus, the learning disabled children in our sample are probably very heterogeneous apart from their special education status. Also, most children classified as learning disabled in Canada receive part-time services but otherwise attend regular classrooms; placement of learning disabled children in self-contained classrooms is relatively rare. Readers can consult Wiener and Siegel (1992) for description of learning disability assessment and placement practices in Canada.

Children in the clinic sample ($N = 248$) were referred for mental health services to a large, urban, child psychiatric facility located in an economically distressed, inner-city area in the midwestern United States. Most of these children resided in families that either had very low incomes or received public assistance. Although these children may have presented with other concurrent emotional or behavioral problems, they were all administered IQ and scholastic achievement measures due to concern about their cognitive status. These children represent all full-battery assessments of school-age children conducted at this clinic over a 5-year period. In contrast, about 1,000 other children received mental health services over the same period who were not seen for full-battery evaluations. About 40% of clinic sample cases were referred by their parents and another 30% were referred by schools. Other children were referred by other mental health agencies (20%) or by private-practice physicians, psychiatrists, or the court (10% in total). A total of 7% of these children were receiving special education services at the time of their evaluations, but their clinical files usually did not specify the types of services (e.g., learning disability, emotionally impaired) or whether they received part-time or full-time services. The remaining 93% were attending regular classrooms. The modal diagnosis based on the *Diagnostic and Statistical Manual, Third Edition (DSM-III; American Psychiatric Association, 1980)* diagnosis was attention deficit disorder with hyperactivity (20%); other diagnoses included adjustment reaction (17%), major depression (13%), overanxious disorder (15%), conduct disorder (8%), and mental retardation (8%). A total of 18% of the cases were not assigned a psychiatric diagnosis.¹ The mean age of clinic sample cases was 12.3 years ($SD = 2.9$, range = 7–16 years), 67% were boys, 57% were White, and 43% were African American. Their average WISC-R Full Scale IQ score was 89.6 ($SD = 17.6$).

Of our two samples, the demographic composition of the school sample would be most comparable to groups of children studied by Siegel.

¹ These children were evaluated before the publication of the revised version of *DSM-III*, the *DSM-III-R* (American Psychiatric Association, 1987).

For example, the gender and ethnic composition, language of instruction, and socioeconomic backgrounds of the Canadian children in our school sample are similar to those of Canadian children tested by Siegel (1988, 1992). There are some differences, however, between the school sample of the present study and samples studied by Siegel that should be noted by readers. All children in our school sample had been referred, and about one quarter were already receiving remedial services. Siegel's samples also included these types of children but also nonreferred, normal-achieving children. Also, children in Siegel's samples ranged in age from 7–16 years; the age range of children in the school sample of the present study is not as wide (7½–12½ years).² The demographic characteristics of our clinic sample are quite different from those of Siegel's samples, and results from both respective sources may not be directly comparable. The availability of the clinic sample in the present study, however, allowed us to evaluate the generalizability of results across demographically diverse groups.

Measures

All children were administered the WISC-R, which was the most widely used IQ scale for school-age children (Chattin & Bracken, 1989) until publication of the WISC-III (Wechsler, 1991). Reviews of the WISC-R are available in Bortner (1985), Detterman (1985), and Witt and Gresham (1985). School sample children were also administered the Kaufman Assessment Battery for Children (K-ABC; Kaufman & Kaufman, 1983a, 1983b) and the Fourth Edition Stanford-Binet Intelligence Scale (SB4; Delaney & Hopkins, 1987; Thorndike, Hagen, & Sattler, 1986a, 1986b). The K-ABC is normed for ages 2½–12½ years, and the SB4 is normed for ages 2–23 years. Both tests have features that may make them well-suited for use with children with reading problems. For example, subtests on two K-ABC scales (Sequential Processing, Simultaneous Processing) were designed to minimize language demands and measure cognitive processes believed crucial in the acquisition of reading skills (e.g., Das & Mensink, 1989). The K-ABC also has a separate Achievement scale that contains two reading tasks for school-age children—Reading-Decoding (a word recognition task) and Reading-Understanding (a reading comprehension measure). One of four scales on the SB4 is intended as a measure of fluid intelligence (Abstract-Visual Reasoning), and the language requirements of another scale (Short-Term Memory) are generally minimal. Other SB4 scales provide estimates of more clearly crystallized abilities, including Verbal Reasoning and Quantitative Reasoning. Although it is beyond the scope of this work to summarize the numerous validity studies of the K-ABC and SB4, several reviews and critiques are available (Bracken, 1985; Kamphaus & Reynolds, 1987; Keith, 1985; Laurent, Swerdlik, & Ryburn, 1992; Sattler, 1988; Spruill, 1987; R. M. Thorndike, 1990).

School sample children were also administered the Wide Range Achievement Test—Revised (WRAT-R; Jastak & Wilkinson, 1984) and the Bender Visual-Motor Gestalt Test (Bender, 1938). The WRAT-R is an individually administered measure that is normed for ages 5–75 years with three subtests, Reading (letter, word recognition), Spelling, and Arithmetic (see test review by Reinehr, 1987). The Bender-Gestalt is a widely used paper-and-pencil drawing task comprising abstract geometric figures. We used the Koppitz (1964, 1975) scoring system to yield a total error score for each child, and we converted error scores to age-based standard scores using Table C-62 of Sattler (1988, p. 885). Koppitz error scores have adequate test-retest and interrater reliability (e.g., .70–.90; Koppitz, 1975) and generally correlate approximately .50 to .40 with IQ and achievement scores and approximately .50 with error scores from other measures of visual-motor coordination (Sattler, 1988).

In addition to the WISC-R, clinic sample children were administered the Peabody Individual Achievement Test (PIAT; Dunn & Markwaldt, 1970). The PIAT is an individually administered multiple-choice test

that is normed for grades Kindergarten–12. The five PIAT subtests include Reading Recognition, Reading Comprehension, Spelling, Arithmetic, and General Information (see test reviews by Lyman, 1971, and Williams & Vincent, 1985). Clinic sample children were also administered the Peabody Picture Vocabulary Test (PPVT; Dunn, 1959), an individually administered, multiple-choice format, pictorial test of receptive vocabulary breadth that is normed for ages 2A–18 years (see test review by Lyman, 1965).³

Assessment of reading skill. We used children's scores on reading recognition tasks to operationally define their level of reading proficiency. Within the school sample, we calculated average standard scores for all children on the K-ABC Reading-Decoding and WRAT-R Reading subtests. The correlation between these subtest scores was very high ($r = .85$), which suggests that a mean score represents performance on a common skill domain. Within the clinic sample, children's scores on the PIAT Reading Recognition subtest served as our reading criterion measure.⁴ Although there is no single type of measure with which to operationally define poor versus good reading skill (e.g., see Swanson, 1989), reading recognition tasks may offer some advantages over alternative types of tests. For example, Siegel (1988) has argued that reading recognition tasks provide "purer" measures of basic reading processes (i.e., phonological processing) than do measures such as reading comprehension tasks, which may have high "ecological validity" but could confound children's reading ability with their vocabulary breadth or familiarity with text content (Siegel, 1988). There is evidence, however, that results of reading recognition and comprehension tasks identify many of the same children as poor readers (e.g., Rispen, van Yperen, & van Duijn, 1991), which suggests that the distinction between these two types of measures may not be so crucial. A particular advantage of using reading recognition tasks in the present study, however, is that Siegel has used the Reading subtest of the WRAT-R in her studies, which enhances comparability of results from our school sample with those of Siegel's (1988, 1992).

Procedure

All children in the school sample were tested in their schools over two half-day (2–3 hours) sessions conducted within the same week. The IQ and achievement measures were administered in counterbalanced order by either Rex Kline or two school psychologists. Rex Kline conducted about one half of the evaluations and also checked all test protocols for scoring accuracy. To evaluate possible examiner effects, we compared mean IQ scores from the WISC-R, K-ABC, SB4, and mean WRAT-R subtest scores across all examiners. All of these comparisons were nonsignificant. Also, we conducted a reliability study of the scoring of the IQ tests. We randomly selected 15 test protocols from the school sample that included test administrations by all three examiners. We removed the original scores, and all protocols were independently rescored by Rex Kline and the two school psychologists. We then derived

² One of the authors (Susan Graham) contacted Linda Siegel to discuss possible samples differences (L. S. Siegel, personal communication, January 25, 1993). No differences other than those already mentioned in the text seemed obvious to Linda Siegel.

³ Revised versions of the PIAT (Markwardt, 1989) and PPVT (Dunn & Dunn, 1981) are available. Previous version/revised version correlations are typically .70–.90.

⁴ Although we administered different reading recognition measures to school and clinic sample children, other investigators have reported high intercorrelations among these subtests. Kaufman and Kaufman (1983b) reported a mean correlation of .74 for the K-ABC Reading-Decoding and PIAT Reading Recognition subtests, and Jastak and Wilkinson (1984) reported an average correlation of .87 for the WRAT-R Reading and PIAT Reading Recognition subtests.

intraclass correlation coefficients for all summary scores across the three examiners. The intraclass correlations ranged from .91 (SB4 Abstract-Visual Reasoning scale) to 1.00 (SB4 Quantitative Reasoning scale); the average intraclass correlation across all test scales was .98.

Children in the clinic sample were tested by staff psychologists and (supervised) psychology interns. Typically, all tests were administered during a single clinic appointment. We estimate that about 40 different examiners conducted the evaluations of clinic sample children over the 5-year period in which these cases were tested. Consequently, it is impractical to test for examiner effects by comparing mean scores across such a large number of examiners. Because each examiner tested a relatively small number of children, however, effects specific to particular examiners may have had little impact on the whole data set. Also, we did not have access to the original record forms of tests administered to these children and thus could not evaluate the reliability of test scoring in the clinic sample.

Results

Moderator Variable Analyses

We used the technique of moderated multiple regression with residualized-centering to evaluate whether children's reading levels affected the external validity of IQ scores. We entered children's IQ and reading recognition scores and the products of these scores (i.e., the IQ \times reading interaction) as predictors of the external measures into regression equations. Significance of the interaction terms would indicate that IQ score validity was moderated by reading skill. Residualized-centering involves partialing out the main effects of IQ and reading from the interaction term to reduce intercorrelations (i.e., multicollinearity) among the predictor variables (Lance, 1988).

Regression results significant at the .01 level are reported in Table 1, including squared multiple correlations and standardized regression weights for IQ, reading, and the interaction.⁵ In both samples, the interaction of IQ and reading was never a significant predictor of the external measures. Thus, the relations of IQ scores to the external criteria were unaffected by children's reading skills. As expected, IQ scores were not only significant predictors of most external measures, they were also usually the most important predictors. In contrast, reading skill was the strongest predictor of spelling ability and text comprehension, both of which are specific language-arts skills. In the school sample, reading skills were also the best predictor of scores on the WRAT-R Arithmetic subtest, which is based on a paper-and-pencil format. All of the other arithmetic subtests have visual-spatial formats, however, and scores on these tasks were best predicted by IQ scores. Also, reading skill was often a significant secondary predictor, especially for external measures with obvious verbal-language content (e.g., K-ABC Riddles) and for tasks that required recall of orally or sequentially presented stimuli (e.g., SB4 Short-Term Memory scale).

Group Comparisons

Although we believe the above analyses provide statistically rigorous tests of IQ \times Reading interactions, results of regression analyses yield little information about discrete groups of children. That is, psychologists often evaluate children who are already members of particular groups (e.g., already in special education) or are asked to assist in the classification of referred

Table 1
Moderator Variable Analyses of the Validity of IQ Scores

External variables	Standardized beta weights			R ²
	IQ	Reading	IQ \times reading	
School sample				
Cognitive scales				
K-ABC				
Sequential processing	.51	.25	—	.47
Simultaneous processing	.76	—	—	.55
SB4				
Verbal Reasoning	.65	.24	—	.67
Abstract-Visual Reasoning	.60	—	—	.36
Quantitative Reasoning	.50	—	—	.39
Short-Term Memory	.66	.23	—	.68
Achievement tasks				
WRAT-R				
Spelling	—	.82	—	.72
Arithmetic	.33	.48	—	.53
K-ABC				
Reading-Understanding	.17	.78	—	.81
Arithmetic	.60	.24	—	.59
Riddles	.64	.25	—	.67
Clinic sample				
PIAT				
Arithmetic	.60	.29	—	.68
Reading Comprehension	.25	.70	—	.78
Spelling	—	.85	—	.70
General Information	.65	.24	—	.69
PPVT	.62	.22	—	.62

Note. All results are significant at $p < .01$. A dash indicates a nonsignificant finding. K-ABC = Kaufman Assessment Battery for Children; SB4 = Fourth Edition Stanford-Binet Intelligence Scale; WRAT-R = Wide Range Achievement Test; PIAT = Peabody Individual Achievement Test; PPVT = Peabody Picture Vocabulary Test.

children. For these analyses we classified children as having either low or normal IQ scores and either poor or normal reading skill. We used the same definition of poor reading used by Siegel in her studies: Poor readers had reading recognition scores equal to below the 25th percentile (i.e., a standard score of 90). Within the clinic sample we classified children as poor readers on the basis of their PIAT Reading Recognition scores and within the school sample on the basis of their mean WRAT-R Reading and K-ABC Reading-Decoding scores. We also applied Siegel's definition of low IQ scores (Full Scale IQ < 80). Siegel (1988) also classified children with higher IQ scores into three groups (80-90, 91-109, and > 110), but our samples in the present study were not large enough to permit similar partition of normal-range scores. These definitions of poor reading and low IQ scores are, of course, not absolute, but they do enhance the comparability of our group comparisons with those from other studies.

⁵ We omitted from our analyses of school sample data results for the Faces and Places subtest of the K-ABC Achievement scale. Because Canadian children have difficulty with the U.S.-specific content of this task, we felt that these results would not be informative.

A total of 16 children (12%) in the school sample and 69 children (28%) in the clinic sample had low IQ scores. All low-IQ children in the school sample were classified as poor readers, as were most cases with low IQ scores in the clinic sample (86% = 59/69). Thus, no low-IQ school-sample children were normal readers, and few cases in the clinic sample (10) were so classified.⁶ Among children with normal-range IQ scores, 64% (75/118) of school sample cases and 36% of clinic sample cases were poor readers. Reported in Table 2 are group mean scores and results of significance tests and effect size indexes (Cohen's *d*, which indicates mean differences in standard deviation units; Cohen, 1977) across the external measures. Within both samples, mean comparisons across all groups would confound IQ and reading levels, and thus were not conducted. Also, comparisons between normal-IQ groups who differ in reading level would indicate only that reading problems do not ordinarily occur in isolation, which is already known. Accordingly, the only comparisons of real interest were between poor readers of different IQ levels (both samples) and between normal readers of different IQ levels (clinic sample).

Almost all comparisons of low- versus normal-IQ levels among poor readers across the external measures were significant for both samples. Thus, poor readers with low IQ scores had poorer skills across a wide range of verbal, quantitative, visual-spatial, memory, and fine motor coordination abilities than did poor readers with higher IQ scores. Also, the magnitudes of these group differences were generally large. (By convention, *d* values around .20 are interpreted as "small" effects, *d* values about .50 indicate "medium" differences, and *d* values greater than .80 are considered "large" effects; Cohen, 1977.) The only exceptions to this general pattern of group differences occurred for specific language-arts skills, including spelling (nonsignificant in both samples) and reading comprehension (nonsignificant only in the school sample). Scores on these measures were generally equally low for poor readers regardless of their IQ levels. Comparisons between normal readers with low versus normal IQ scores in the clinic sample indicated a similar pattern. That is, the spelling and reading comprehension abilities of these children were comparable (i.e., both normal-range) regardless of their IQ levels, but normal readers with higher IQ scores had better arithmetic skills and funds of general information.

Discussion

In this study we found little evidence that IQ scores had less concurrent validity for children who are poor readers than for children with better reading skills. Results of both regression analyses and group comparisons conducted within two separate samples indicated that lower IQ scores were associated with poorer abilities in several domains regardless of children's reading skills. The only instance in which IQ scores had nil external validity in the present study concerned specific language-arts achievement areas, such as spelling and text comprehension. Reading ability, as measured by performance on word recognition tasks, was the only significant predictor of these achievement skills. That is, children with poor or normal reading skills tended to have (respectively) poor or normal language-arts skills regardless of their IQ scores.

Some findings of the present study correspond with those reported in other investigations of the relations between general ability and specific reading skill. For example, we found that reading ability had incremental validity over IQ scores in the prediction of a wide range of cognitive skills, including general verbal and visual-spatial reasoning, quantitative facility, and short-term memory. These findings are consistent with suggestions by other researchers regarding possible reciprocal effects among specific reading ability and other, more general cognitive skills (e.g., Share et al., 1989; Stanovich, 1986). Also, poor readers in the present study as well as those in Siegel's (1988, 1992) studies performed poorly on spelling and text comprehension tasks regardless of their IQ levels. Siegel (1992) has interpreted such findings to mean that poor readers with higher IQ scores cannot apply their general cognitive resources to compensate for specific deficits in phonological processing, which they share with low-IQ poor readers. Siegel's hypothesis is consistent with a hierarchical view of the organization of cognitive abilities in which specific competencies (e.g., spelling) may be influenced by broader, higher level abilities (e.g., *g*) but are not wholly determined by them.

Unlike other results reported by Siegel (1988, 1992), however, we found numerous and sizable differences between low- and normal-IQ poor readers in other cognitive domains beyond phonological processing. Although in the present study we tried to make our procedures as similar as possible to those used by Siegel (e.g., same types of reading measures and definitions of poor reading and low IQ scores) and studied one sample that may be comparable with Siegel's, other factors could account for differences in results across these studies. For example, some of our statistical procedures (i.e., the regression analyses) are more powerful than the group comparisons conducted by Siegel (1988, 1992) and thus may indicate greater validity of IQ scores. Results of group comparisons in the present study, however, also indicated larger IQ-group differences than found by Siegel (1988, 1992). We administered a full IQ battery to every child in our samples instead of screening tests, which would probably enhance the external validity of IQ scores. Also, the samples of the present study were composed of referred children, whereas Siegel's (1988, 1992) samples contained both referred and non-referred cases. Although the ranges of ability in our samples may be smaller than those in Siegel's samples—which might restrict correlations between IQ scores and external measures—we nevertheless found stronger evidence for the validity of IQ scores. The sizes of the samples in this study were not large enough to form as many IQ groups as studied by Siegel (1988). We found many significant differences between poor and normal readers divided into only two IQ groups, however, despite the possibility that comparisons among greater numbers of IQ levels may have been more powerful. Finally, Siegel (1988, 1992) used different types of tests as external criteria for IQ scores, which raises the possibility of measure-specific results about the validity of IQ scores.

Although it is difficult to identify which of the above factors may account for differences between the results of this study

⁶ Siegel (1988) also found few low-IQ normal readers in her sample: only 14 of 947 cases (1%) had IQ scores < 80 and reading recognition scores > 90.

and those of Siegel's (1988, 1992), these issues offer suggestions for further study in this area. Researchers should study the validity of IQ tests using other types of samples, especially children with different types of special education classifications or academic dysfunctions. For example, the group of students classified as learning disabled in our school sample was small and very heterogeneous, both of which precluded us from studying the validity of IQ scores for more specific, homogeneous subgroups of children. Also, researchers should also use a wider range of external criteria for IQ scores, such as classroom performance, other types of individually administered measures, or educational outcome. The latter type of criterion would be especially relevant because it involves the predictive validity of IQ scores for poor readers. That is, although special educational resources are often allocated according to IQ-achievement discrepancies, there is actually little empirical evidence concerning the prognosis or differential response to remediation of poor readers with low versus normal IQ scores (e.g., Stanovich, 1991). For example, Shaywitz et al. (1992) found that IQ-discrepant poor readers showed improved reading skill relative to age-norms from Grade 2 through 5; the reading skill of low-IQ poor readers did not improve over this time. Although these results seem to lend support to the construct validity of reading disability, there are very few similar longitudinal studies about the predictive validity of IQ-reading discrepancies. Finally, it would also be important to study whether reading level moderates IQ or factor scores from the WISC-III (Wechsler, 1991), which may become as widely used as its predecessor, the WISC-R.

We should note in closing that although the results of this study indicate that the external correlates of IQ scores do not vary by children's reading skills, we agree with concerns expressed by Siegel and other researchers that the practice of allocating special education resources on the basis of IQ-achievement discrepancies lacks a scientific basis. That is, the "weak link" in operational definitions of reading disability may not be IQ tests per se but, rather, unsubstantiated assumptions about the implications of IQ-achievement discrepancies for children's educational progress. Until appropriate, well-designed, and longitudinal studies of the validity of the IQ-achievement discrepancy model are conducted, we concur with the view that denial of remedial services to some children with low IQ scores may be unwarranted.

References

- American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.
- American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.
- Bender, L. (1938). *A visual motor gestalt test and its clinical use*. American Orthopsychiatric Association (Research Monograph, No. 3).
- Bortner, M. (1985). Review of Wechsler Intelligence Scale for Children—Revised. In J. V. Mitchell (Ed.), *The ninth mental measurements yearbook* (Vol. 2, pp. 1713–1714). Lincoln: Buros Institute of Mental Measurements of the University of Nebraska—Lincoln.
- Bracken, B. A. (1985). A critical review of the Kaufman Assessment Battery for Children. *School Psychology Review*, 14, 21–36.
- Chattin, S. H., & Bracken, B. A. (1989). School psychologists' evaluation of the K-ABC, McCarthy Scales, Stanford-Binet IV, and WISC-R. *Journal of Psychoeducational Assessment*, 7, 112–130.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.
- Das, J. P., & Mensink, D. (1989). K-ABC simultaneous-sequential scales and prediction of achievement in reading and mathematics. *Journal of Psychoeducational Assessment*, 7, 103–111.
- Delaney, E. A., & Hopkins, T. F. (1987). *Examiner's handbook*. Chicago: Riverside.
- Detterman, D. K. (1985). Review of Wechsler Intelligence Scale for Children—Revised. In J. V. Mitchell (Ed.), *The ninth mental measurements yearbook* (Vol. 2, pp. 1715–1716). Lincoln: Buros Institute of Mental Measurements of the University of Nebraska—Lincoln.
- Dunn, L. M. (1959). *Peabody Picture Vocabulary Test manual*. Circle Pines, MN: American Guidance Service.
- Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test—Revised*. Circle Pines, MN: American Guidance Service.
- Dunn, L. M., & Markwaldt, F. C. (1970). *Manual for the Peabody Individual Achievement Test*. Circle Pines, MN: American Guidance Service.
- Fletcher, J. M., Espy, K. A., Francis, D. J., Davidson, K. C., Rourke, B. P., & Shaywitz, S. E. (1989). Comparisons of cutoff and regression-based definitions of reading disabilities. *Journal of Learning Disabilities*, 22, 334–355.
- Fletcher, J. M., Francis, D. J., Rourke, B. P., Shaywitz, S. E., & Shaywitz, B. A. (1992). The validity of discrepancy-based definitions of reading disabilities. *Journal of Learning Disabilities*, 25, 618–629.
- Jastak, J., & Wilkinson, G. S. (1984). *Wide Range Achievement Test—Revised*. Wilmington, DE: Jastak.
- Kamphaus, R. W., Frick, P. J., & Lahey, B. J. (1991). Methodological issues and learning disability diagnosis in clinical practice. *Journal of Learning Disabilities*, 24, 613–618.
- Kamphaus, R. W., & Reynolds, C. R. (1987). *Clinical and research applications of the K-ABC*. Circle Pines, MN: American Guidance Service.
- Kaufman, A. S., & Kaufman, N. L. (1983a). *K-ABC administration and scoring manual*. Circle Pines, MN: American Guidance Service.
- Kaufman, A. S., & Kaufman, N. L. (1983b). *K-ABC interpretive manual*. Circle Pines, MN: American Guidance Service.
- Keith, T. Z. (1985). Questioning the K-ABC: What does it measure? *School Psychology Review*, 14, 9–20.
- Koppitz, E. M. (1964). *The Bender Gestalt Test for young children*. New York: Grune & Stratton.
- Koppitz, E. M. (1975). *The Bender Gestalt Test for young children: Research and application, 1963–1973* (Vol. 2). New York: Grune & Stratton.
- Lance, C. E. (1988). Residual centering, exploratory and confirmatory moderator analysis, and decomposition of effects in path models containing interaction effects. *Applied Psychological Measurement*, 12, 163–175.
- Laurent, J., Swerdlik, M., & Ryburn, M. (1992). Review of validity research on the Stanford-Binet Intelligence Scale: Fourth Edition. *Psychological Assessment*, 4, 102–112.
- Lyman, H. B. (1965). Peabody Picture Vocabulary Test. In O. K. Buros, (Ed.), *The sixth mental measurements yearbook* (pp. 530–532). Highland Park, NJ: Gryphon Press.
- Lyman, H. B. (1971). Review of the Peabody Individual Achievement Test. *Journal of Educational Measurement*, 8, 137–138.
- Markwardt, F. C. (1989). *Peabody Individual Achievement Test—Revised*. Circle Pines, MN: American Guidance Service.
- Pennington, B. F., Gilger, J. W., Olson, R. K., & DeFries, J. C. (1992). The external validity of age- versus IQ-discrepancy definitions of reading disability: Lessons from a twin study. *Journal of Reading Disabilities*, 25, 562–573.
- Reinehr, R. C. (1987). Wide Range Achievement Test. In D. J. Keyser

- & R. C. Sweetland (Eds.), *Test critiques compendium* (pp. 583–586). Kansas City, MO: Test Corporation of America.
- Reynolds, C. R. (1984–1985). Critical measurement issues in learning disabilities. *Journal of Special Education, 18*, 451–476.
- Rispens, J., van Yperen, T. A., & van Duijn, G. A. (1991). The irrelevance of IQ to the definition of learning disabilities: Some empirical evidence. *Journal of Learning Disabilities, 24*, 434–438.
- Sattler, J. M. (1988). *Assessment of children* (3rd ed.). San Diego: Author.
- Share, D. L., McGee, R., & Silva, P. A. (1989). IQ and reading progress: A test of the capacity notion of IQ. *Journal of the American Academy of Child and Adolescent Psychiatry, 28*, 91–100.
- Share, D. L., & Silva, P. A. (1987). Language deficits and specific reading retardation: Cause or effect? *British Journal of Disorders of Communication, 22*, 219–226.
- Shaywitz, B. A., Fletcher, J. M., Holahan, J. M., & Shaywitz, S. E. (1992). Discrepancy compared to low achievement definitions of reading disability: Results from the Connecticut Longitudinal Study. *Journal of Learning Disabilities, 25*, 639–648.
- Siegel, L. S. (1988). Evidence that IQ scores are irrelevant to the definition and analysis of reading disability. *Canadian Journal of Psychology, 42*, 201–215.
- Siegel, L. S. (1989a). IQ is irrelevant to the definition of learning disabilities. *Journal of Learning Disabilities, 22*, 469–486.
- Siegel, L. S. (1989b). Why we do not need intelligence test scores in the definition and analysis of learning disabilities. *Journal of Learning Disabilities, 22*, 514–518.
- Siegel, L. S. (1992). An evaluation of the discrepancy definition of dyslexia. *Journal of Learning Disabilities, 25*, 618–629.
- Spruill, J. (1987). Stanford-Binet Intelligence Scale Fourth Edition. In D. J. Keyser & R. C. Sweetland (Eds.), *Test critiques* (Vol. 6, pp. 544–559). Kansas City, MO: Test Corporation of America.
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360–407.
- Stanovich, K. E. (1989). Has the learning disability field lost its intelligence? *Journal of Learning Disabilities, 22*, 487–492.
- Stanovich, K. E. (1991). Discrepancy definitions of reading disability: Has intelligence led us astray? *Reading Research Quarterly, 26*, 7–29.
- Swanson, H. L. (1989). Phonological processes and other routes. *Journal of Learning Disabilities, 22*, 493–497.
- Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986a). *Stanford-Binet Intelligence Scale: Guide for administering and scoring the fourth edition*. Chicago: Riverside.
- Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986b). *Stanford-Binet Intelligence Scale Fourth Edition: Technical manual*. Chicago: Riverside.
- Thorndike, R. M. (1990). Would the real factors of the Stanford-Binet Fourth Edition please come forward? *Journal of Psychoeducational Assessment, 8*, 412–435.
- Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children—Revised*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (1991). *WISC-III manual*. San Antonio, TX: Psychological Corporation.
- Wiener, J., & Siegel, L. (1992). A Canadian perspective on learning disabilities. *Journal of Learning Disabilities, 25*, 340–350.
- Williams, R. E., & Vincent, K. R. (1985). Peabody Individual Achievement Test. In D. J. Keyser & R. C. Sweetland (Eds.), *Test critiques* (Vol. 3, pp. 480–487). Kansas City, MO: Test Corporation of America.
- Witt, J. C., & Gresham, F. M. (1985). Review of Wechsler Intelligence Scale for Children—Revised. In J. V. Mitchell (Ed.), *The ninth mental measurements yearbook* (Vol. 2, pp. 1716–1719). Lincoln: Bureau of Mental Measurements of the University of Nebraska-Lincoln.

Received August 31, 1992

Revision received March 5, 1993

Accepted March 15, 1993 ■