Gender Bias in IQ-Discrepancy and Post-Discrepancy Definitions of Reading Disability

David L. Share and Phil A. Silva

Abstract

The present study investigated the hypothesis that the higher prevalence of reading disability (RD) often observed among boys is partly an artifact of gender bias in the prediction of reading from IQ. The relevant regression statistics derived from a sample of more than 900 children revealed a statistically significant intercept bias. Predicted reading scores for boys were systematically overestimated, thereby inflating IQ–reading discrepancies; the converse was found for girls. When defined separately for girls and boys, severe underachievement in reading was found to be equally prevalent in both genders and, furthermore, was associated with qualitatively and quantitatively similar patterns of deficits. Because the bias arose from general differences between boys and girls in reading score distributions (a lower mean and greater variance for boys) rather than from differences in IQ scores, gender bias poses a potential threat not only to traditional IQ-discrepancy definitions but also to post-discrepancy definitions that are based solely on reading score cutoffs. Future classification criteria for RD need to take heed of the possibility that when the distributions of reading scores for boys and girls are not identical, performance cutoffs designating low achievement that are based on data pooled from both genders are likely to result in the overidentification of boys with RD and the underidentification of girls with RD.

The essence of discrepancy definitions of reading disability is a reading performance that falls well below expectations based on age (or grade level) and IQ. Traditional IQ-based discrepancy formulae, however, may introduce a gender bias such that boys with a reading disability (RD) are more likely to be identified than girls with RD.

According to Jensen (1980), prediction is biased if it either overestimates or underestimates an individual's criterion performance depending on group membership. "Predictive bias means *systematic error* (in contrast to random errors of measurement) in the prediction of the criterion variable for persons of different subpopulations as a result of basing prediction on a common regression equation for *all* persons regardless of their subpopulation memberships, or basing prediction for persons of one subpopulation on the regression equation derived on a different subpopulation" (Jensen, 1980, p. 380). Jensen formally defined *bias* as a statistically significant difference between subpopulations in either slope, intercept, or standard error of estimate. In such cases, prediction based on a common regression equation will systematically overestimate or underestimate criterion scores.

Jensen (1980) provided an example of gender bias in the prediction of college grade point average (GPA) from Scholastic Aptitude Test (SAT) scores. Because of their lower GPA scores, the regression intercept is lower for boys. When boys and girls are combined in a common regression equation, the intercept is shifted upward toward the girls' intercept. This higher intercept systematically overestimates GPA scores for boys and underestimates GPA scores for girls.

In the identification of RD, it is customary to use a common discrepancy formula such as a regression equation for both boys and girls when predicting reading from age (or grade level) and IQ. An *intercept bias* similar to that just described may inflate the prevalence of severe underachievement among boys. Significant gender differences in intercept (with a lower intercept for boys) would cause the systematic overestimation of boys' reading scores, thereby inflating discrepancies between actual and predicted reading scores for underachieving boys. By the same token, underprediction of girls' reading scores would result in smaller discrepancies for underachieving girls, leading to overidentification of RD in boys and underidentification in girls.

If gender bias accounts for the preponderance of boys in discrepancy-defined groups, then defining the disorder separately for girls and boys should show similar prevalence rates *within* each gender. Furthermore, if both the prevalence rates and the pattern of deficits are found to be similar for girls and boys with RD, then current gender-pooled classification methods will need to be revised.

Using data from a longitudinal study of more than 900 New Zealand children followed since birth, we investigated the extent to which the higher prevalence of specific reading disability among boys was an artifact of gender bias. This study also compared the profile of cognitive, neuropsychological, and academic achievement deficits in girls and boys with RD.

Method

Sample

The sample consisted of children from the Dunedin Multidisciplinary Health and Development Research Unit. This sample has been studied since birth, with most of the children tested within 1 month of their birthday every second year from age 3. The study and sample have been described in detail by McGee and Silva (1982; see also Silva & Stanton, 1996). In summary, these children were part of a cohort born between April 1, 1972, and March 31, 1973, at Queen Mary Hospital in Dunedin. The children were first traced at age 3 (1975). A total of 1,139 lived in the Dunedin metropolitan area of the province of Otago and were, thus, eligible for inclusion in the study. Of the 1,139 children, 1,037 were assessed within 1 month of their third birthdays (1975–1976). The present analyses are primarily based on data collected at ages 5 (n = 991), 7 (n = 954), 9 (n = 955), 11 (n = 925), and 13 (n = 850).

When compared to all New Zealand children, the Dunedin sample is slightly biased, with more children being represented at the upper socioeconomic status (SES) levels and fewer at the lower SES levels according to Elley and Irving's (1972) index. Moreover, the sample is mainly of European origin, with only 2% of Maori and Polynesian background, compared with about 10% for the country as a whole (Department of Statistics, 1976).

Measures

IQ and Reading. Intelligence was assessed at age 11 with the *Wechsler Intelligence Scale for Children–Revised* (WISC-R; Wechsler, 1974). Only eight subtests were given in order to reduce testing time: Information, Similarities, Arithmetic, Vocabulary, Picture Completion, Block Design, Object Assembly, and Coding. Verbal, Performance, and Full Scale IQs were prorated using the method described in the test manual. Reading was assessed with the *Burt Word Reading Test* (Scottish Council for Research in Education, 1976).

Social Disadvantage. At age 11, a measure of the adversity of the child's home background was obtained, based on Rutter's adversity index (Rutter, 1978). The index, described by McGee, Williams, and Silva (1985), was based on low social class, large family size, low maternal intelligence score, solo parent status, poor maternal mental health, and a relatively low score on a measure of family social environment. All children with disadvantage scores of 2 or more (approximately 10% of the sample at age 11) were classified as *high disadvantage,* and the remainder were classified as *low disadvantage.*

Language and Speech Development. At age 9, the Verbal Comprehension (Auditory Reception) and Verbal Expression subscales of the *Illinois Test of Psycholinguistic Abilities* were administered (Kirk, McCarthy, & Kirk, 1968). Speech articulation was assessed at age 9 with the *Dunedin Articulation Check* (Justin, Lawn, & Silva, 1983). At age 11, listening comprehension was tested with the *Progressive Achievement Tests* (PAT) Listening Comprehension test (Elley & Reid, 1971).

Educational Attainment. At age 11, scores on the PAT Reading Comprehension test and Mathematics test (Elley & Reid, 1969; Reid & Hughes, 1974) were available for 616 children. These data were obtained from a broader study of the educational attainment of 2,600 Form I and II children from Dunedin's six intermediate schools (Silva, 1984). PAT scores were recorded as age-based percentiles using norms from the test manuals.

A measure of written expression was also available at age 12 for 592 children living in the Dunedin area. The children were given 30 minutes to write a story in response to a picture. A global measure of writing competence was based on eight subscales, each rated on a 0–5 scale. These subscales were as follows: Organization, Development of Ideas, Clarity of Communication, Reader Impact, Punctuation, Sentence Structure, Spelling, and Handwriting. This measure was described in detail by Adler (1986).

Spelling was assessed at age 11 with the *Dunedin Spelling Test* (Silva et al., 1981). The *Dunedin Spelling Test* comprises 25 words graded in difficulty with parallel-form reliability of .94.

Motor Development. At age 9, motor abilities were assessed with the *Basic Motor Ability Test* (Arnheim & Sinclair, 1974). This test is composed of nine subtests: Long Jump, Agility Run, Target Throwing, Push-Ups, Face-Down to Standing, Tapping, Static Balance, Bead Stringing, and Hamstring Stretch.

Neurological Assessment. At age 3, each child was examined by a pediatrician for neurological signs, including assessment of motility, passive movements, reflexes, facial musculature, strabismus, nystagmus, foot posture, and gait. The method of assessment was a modification of that described by Touwen and Prechtl (1970), and the results were described by McGee, Clarkson, Silva, and Williams (1982). The children were divided into two groups: those with fewer than two abnormalities and those with two or more abnormalities.

Neuropsychological Measures. At age 13, a battery of neuropsychological measures was administered. These measures included both language-based and nonverbal tasks. The language-based tasks included the following measures:

1. The *Rey Auditory Verbal Learning Test* (Rey, 1964; Taylor, 1959), which consisted of four presentation trials with immediate recall of a 15-word list, presentation of an interference list, and a sixth recall trial following a 15-minute delay;
2. a shortened version of the *Controlled Oral Word Association Test* (Benton & Hamsher, 1978), in which the child was asked to say as many words as possible beginning with the letter *a* (and then *s*) in 1 minute;
3. the *Trail Making Test*, Forms A and B (Lewinsohn, 1973; Reitan, 1958). For this test, the child drew lines to connect a sequence of numbered circles on a sheet (A) and then to connect numbered and lettered circles in alternating sequence (B).

The nonverbal test battery included

1. the *Grooved Pegboard Test* (Klove, 1963; Knights & Moule, 1968), consisting of a small board containing slotted holes angled in varying directions. The child was timed while inserting notched pegs into the board, with separate trials for each hand;
2. the WISC-R Mazes subtest (Wechsler, 1974), in which the child drew the way out of a series of increasingly difficult mazes;
3. The *Rey-Osterrieth Complex Figure Test* (Osterrieth, 1944; Waber & Holms, 1985), which required the child to copy a complex figure and then, after 3 minutes of interpolated activity (the *Grooved Pegboard Test*), to reproduce the figure from memory.

Each of these tests yielded scores for quantitative aspects (e.g., number of errors), for qualitative aspects (e.g., type of errors), and for timed aspects of task performance. All scores were coded such that higher scores reflected better performance, consistent with the WISC-R.

Identification of RD Groups

At age 11, underachieving groups were defined using Rutter and Yule's (1975) regression technique. A regression equation was computed to predict reading from IQ. To maintain continuity with previous studies of RD in this sample (Share, McGee, & Silva, 1989; Silva, McGee, & Williams, 1985), nonverbal (Performance) IQ rather than Verbal IQ was used as the benchmark of general ability. Many advocates of discrepancy definitions have argued that Verbal IQ is more appropriate, because it is the discrepancy between language and oral potential and reading that is most relevant. The choice of Performance IQ in the present study was motivated by our own (Share et al., 1989) and others' (e.g., Bishop & Butterworth, 1980) data showing significantly greater declines in Verbal IQ than in Performance IQ associated with poor reading progress. This finding indicates that Performance IQ is more independent of reading than Verbal IQ and, thus, provides a cleaner measure of the classical concept of general intellectual ability embodied in traditional notions of specific reading disabilities (see, e.g., Rutter & Yule, 1975). In any case, the present argument regarding potential gender bias is unaffected by the choice of IQ measure, as all three distributions (Verbal, Performance, and Full Scale; see Table 3) were very similar for both genders.

Regression equations were computed separately for girls (n = 443), for boys (n = 471), and for the combined gender groups (n = 914). Age was not needed in the regression equations because it was constant in our sample. In each of the three groups, children whose reading scores were more than 1.5 standard errors below prediction were designated as having specific reading disabilities. The regression parameters, together with the size and composition of the three groups of children with RD are presented in the Results section (see Table 1).

It is important to note that using a pure discrepancy between reading and IQ does not require an absolute low level of achievement. Thus, in theory at least, a child could have a very high IQ and an above-average reading score yet be considered to have RD according to the discrepancy method. In this particular sample, however, we found no discrepancy-defined child with above-average reading scores. Moreover, approximately 90% of the discrepancy-defined sample scored at least 1 *SD* below the mean. Using the combined (boys and girls) regression equation, only 8 children (all boys) out of a total of 74 children (10.8%) classified as having RD performed above this cutoff, and only 5 (4 boys and

TABLE 1
Descriptive Statistics and Regression Data for All Participants at Age 11 by Gender

	Boys[a]		Girls[b]	
Measure	***M***	***SD***	***M***	***SD***
Reading (BWRT)	69.9	20.91	75.6	18.44
IQ (WISC-R)	111.7	15.55	110.2	15.66
Regression				
slope	.53557	.05697	.51570	.05041
intercept	10.07715	6.42656	18.72201	5.61189
SE	19.20286		16.60012	

Note. BWRT = *Burt Word Reading Test* (Scottish Council for Research in Education, 1976); WISC-R = *Wechsler Intelligence Scale for Children–Revised* (Wechsler, 1974).
[a]n = 471. [b]n = 443.

1 girl) out of 66 (7.6%) scored above the cutoff using the within-gender regression equations.

Statistical Analysis

In the analysis of deficits associated with RD, planned contrasts were used to test the significance of differences between children with and without RD on individual measures and of group-by-gender interactions indicating whether those differences varied across gender. Because of the large number of significant tests being conducted, the Bonferroni inequality (Grove & Andreasen, 1982) was used to maintain the familywise error rate at .05. Dependent measures were grouped into four families: educational attainment; WISC-R subtests; speech–language, motor, and neurological measures; and neuropsychological measures. Within a family of measures, each planned contrast was tested by dividing alpha by the number of contrasts being tested in that family. For example, alpha for each of the ten measures of educational attainment was set at .005 (.05/10).

Results

Evidence for Gender Bias in the Definition of RD

Means and standard deviations of reading and IQ scores for 443 girls and 471 boys at age 11 are reported in Table 1, together with the regression statistics for predicting reading from IQ.

Consistent with previous reports, the mean of the distribution of reading scores was higher for girls, although the variance was lower than for boys. Means and variances of the IQ distributions for both genders were similar.

There were no statistically significant gender differences in either slopes, t (910) < 1.0, or standard errors of estimate, t (910) = 1.34, $p > .05$. However, a clear gender difference emerged in intercepts, t (910) = 5.47, $p < .05$, with a lower intercept for boys amounting to almost 9 points.

The effect of this intercept bias can be illustrated by comparing the predicted reading score derived from the common regression equation (all 914 children combined) to the separate within-gender intercepts. The common regression equation was Predicted reading = .515889 × IQ + 15.390574 (standard error of estimate = 18.25544). The predicted reading score for a boy with an IQ of 100 is 63.6 when based on the *boys'* regression equation, compared to 67.0 when based on the *combined* regression equation. For a girl of IQ 100, the intercept bias reduces the predicted reading score from 70.3 based on the *girls'* equation to 67.0 based on the combination equation. That is, reading scores for boys are systematically overpredicted by approximately 3 points, whereas girls' scores are underpredicted by a similar amount. This general overestimation of boys' reading scores inflates the magnitude of the discrepancies between the actual and predicted reading scores for boys, thereby increasing the number of cases falling below a given cutoff for underachievement. The converse occurs for girls. This is graphically depicted in Figures 1 and 2, in which 1.5 standard error cutoffs are plotted for boys and girls, respectively. Children falling below this cutoff were classified as having specific reading disabilities. In each figure, within-gender cutoffs (solid lines) and combined-gender cutoffs (broken lines) are plotted.

As is evident from Figure 1, the customary combined-gender cutoff classifies an additional 23 boys as having RD who would not otherwise be so identified. In Figure 2, the combined-gender cutoff *excludes* 16 girls who would be identified as having RD using the gender-specific cutoff.

The gender bias inherent in conventional discrepancy-based classification methods appears sufficient to account for the gender differences in RD prevalence rates observed in this sample. When computed separately for boys and girls, the prevalence of specific reading disability is very similar: 32 out of 471, or 6.8%, for boys; 34 out of 443, or 7.7%, for girls. Clearly, severe underachievement in reading is equally prevalent across genders once this prediction bias is controlled.

Characteristics of Boys and Girls with RD

It is evident that underachieving girls risk not being identified as having RD if classification is based on combined-gender regression equations. To what extent is the degree of underachievement in girls comparable to the degree of underachievement in boys? Reading and IQ scores for children with RD are compared to those of children without RD in Table 2.

Although both girls and boys with RD differed substantially from their same-gender peers with RD on reading scores, there were no significant differences on Performance IQ. The absence of significant group-by-gender interactions indicated that relative to their peers, both RD groups were underachieving to the same extent. Also, boys and girls differed from each other in reading scores.

Having established that severe reading underachievement is equally prevalent in girls and boys, the question arises as to whether underachieving girls show a pattern of reading deficits similar to that of underachieving boys. Means and standard deviations for WISC-R subtests and language, speech, motor, neurological, and disadvantage measures can be compared in Table 3.

As expected, children with RD (both girls and boys) had lower Verbal IQs and scored significantly lower than control children on all Verbal subtests. There were no significant differences either on overall Performance IQ or on individual Performance subtests. The decrement in Full Scale IQ can therefore be attributed to lower Verbal IQ. Significant differences between RD groups and controls were also found on speech articulation at age 9 and on listening comprehension (age 12), but not on ITPA scores at age 9. No significant differences were obtained on motor, neurological, or disadvantage

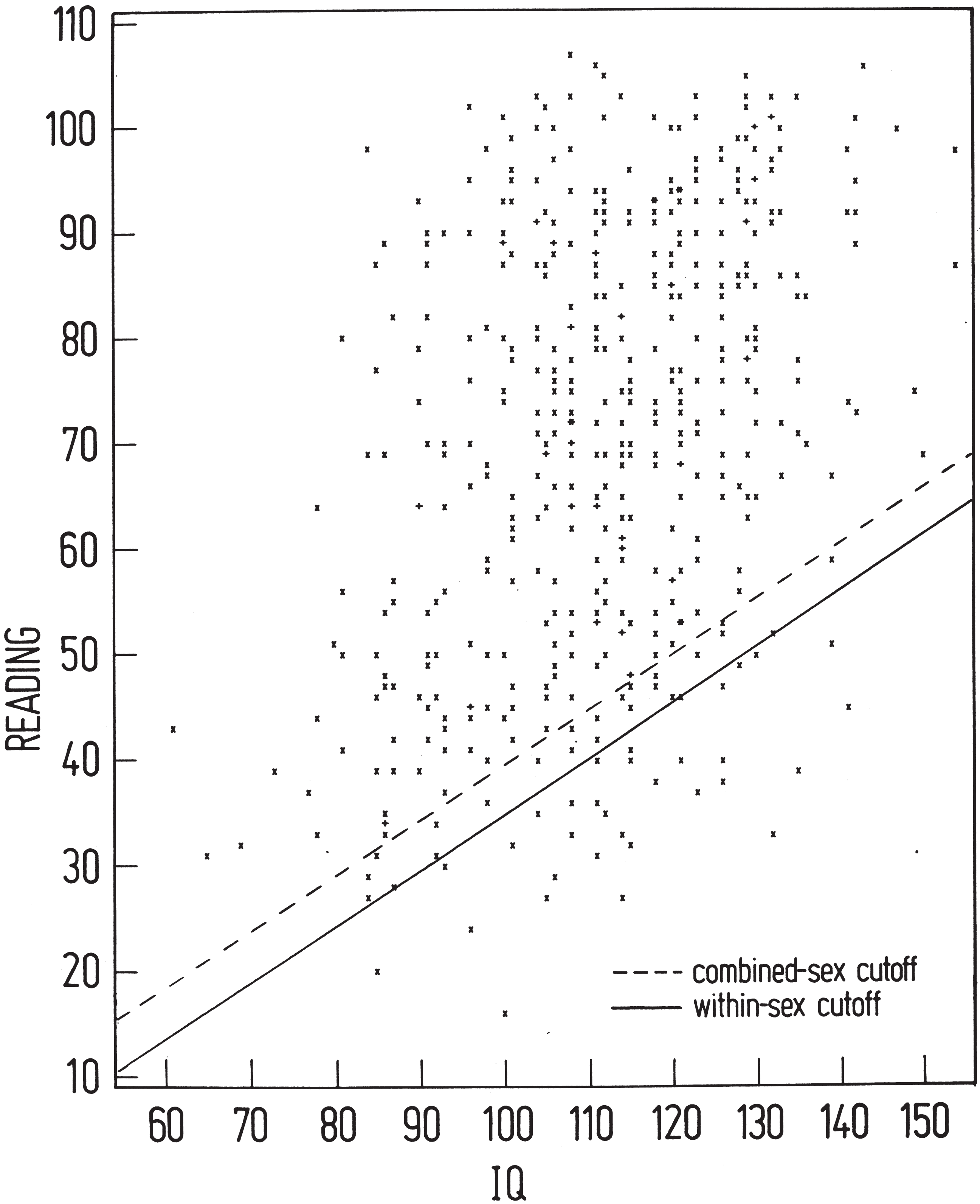

FIGURE 1. Scatterplot of boys' reading and IQ scores with 1.5 standard error cutoffs for specific reading disability. Solid line indicates cutoff based on within-gender regression equation; broken line indicates cutoff based on combined-gender regression equation.

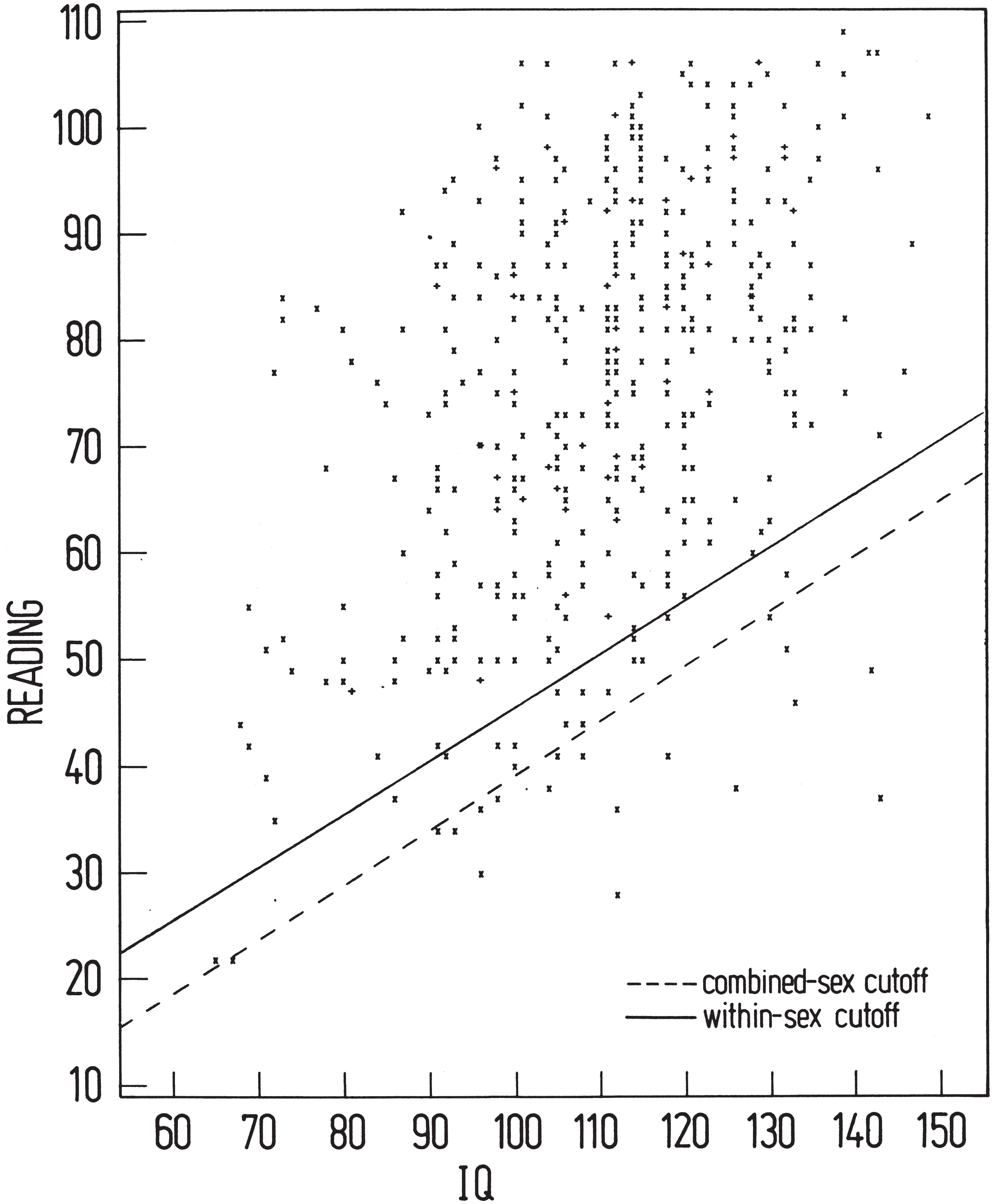

FIGURE 2. Scatterplot of girls' reading and IQ scores with 1.5 standard error cutoffs for specific reading disability. Solid line indicates cutoff based on within-gender regression equation; broken line indicates cutoff based on combined-gender regression equation.

TABLE 2
Means and Standard Deviations of Reading and IQ Scores by Gender and Disability Status

Group	WISC-R Performance IQ		BWRT Word Reading	
	M	*SD*	*M*	*SD*
Boys				
RD[a]	115.3	13.28	35.9*	8.29
NRD[b]	111.5	15.68	72.4	19.34
Girls				
RD[c]	108.2	18.06	41.5*	8.61
NRD[d]	110.4	15.46	78.4	16.05

Note. RD = reading disabilities; NRD = no reading disabilities; BWRT = *Burt Word Reading Test* (Scottish Council for Research in Education, 1976); WISC-R = *Wechsler Intelligence Scale for Children–Revised* (Wechsler, 1974).
[a]*n* = 32. [b]*n* = 439. [c]*n* = 34. [d]*n* = 409.
*$p < .05$, comparing RD and NRD groups.

measures. These results fully replicate findings (cumulative to age 9) reported earlier by Silva et al. (1985) for boys with specific reading disabilities. The significant differences between children with and without RD on listening comprehension at age 12 may be attributable either to differences in the content of the language measures used at ages 9 and 12 or to a decline in general language skills resulting from impoverished reading opportunities. This issue is addressed more fully elsewhere (see Share & Silva, 1987). Of major interest is the lack of any significant gender-by-group interactions, indicating that girls with specific reading disabilities show the same pattern of deficits as boys with specific reading disabilities.

Table 4 reports means and standard deviations for the neuropsychological measures administered at age 13. Significant differences between children with and without RD were found on only two measures—oral word association and delayed verbal recall. These data are therefore consistent with the findings at age 11, indicating a specific verbal deficit in children with RD. Again, there were no gender-by-group interactions, indicating that both boys and girls with RD show the same pattern of deficits.

The results for tests of educational attainment appear in Table 5. The data in Table 5 indicate that children defined as underachieving in reading on a test of word recognition also exhibited underachievement in a wide range of academic areas, including spelling, written expression, and mathematics. Once again, girls and boys with RD were no different in this regard.

Discussion

It has long been observed that groups of children with RD tend to have a large proportion of boys. A variety of explanations for this gender difference have been discussed in the research literature, including gender-related genetic factors (e.g., Geschwind & Galaburda, 1985; Tallal, Ross, & Curtiss, 1989), referral biases (e.g., Shaywitz, Shaywitz, Fletcher, & Escobar, 1990; Wadsworth, DeFries, Stevenson, Gilger, & Pennington, 1992), and statistical artifacts (Share, McGee, McKenzie, Williams, & Silva, 1987; Van der Wissel & Zegers, 1985).

It is widely acknowledged that in school-identified and teacher-identified groups, as well as in clinical samples, referral and ascertainment biases are responsible for most, if not all, of the gender imbalance. Clinic- and school-identified samples typically have two to three boys to each girl (Naiden, 1976; Vogel, 1990; Wadsworth et al., 1994). Boys with RD, like boys generally, are perceived to be more overactive and to have greater behavioral difficulties than girls with RD—qualities regarded as disruptive in the classroom and hence requiring intervention (Shaywitz et al., 1990). In research-defined (i.e., test-defined) samples, the ratio of boys to girls is usually much lower than in referred samples, although still slightly above a one-to-one ratio (closer to three boys to two girls; Flynn & Rahbar, 1994; Naiden, 1976; Shaywitz et al., 1990). Thus, although referral biases certainly appear to be the main reason for the typical overidentification of boys, they may not always be the only source.

The present study identified another potential source of bias—statistical bias. We found that when the general distribution of reading scores for boys was lower on average than the distribution for girls, the IQ-discrepancy regression formula produced a preponderance of boys with RD due to the fact that reading scores were systematically overestimated for boys, thereby inflating the discrepancies. The converse occurred for girls. Although some research studies have reported significantly lower mean reading scores for boys compared to girls (Jorm, Share, Maclean, & Matthews, 1986; Rutter, Tizard, & Whitmore, 1970; Silva et al., 1985), others have found small or negligible differences favoring girls (Flynn & Rahbar, 1994; Shaywitz et al., 1990). Thus, gender bias need not be an inevitable product of IQ-discrepancy formulae. However, in the case of extreme cutoffs, even small differences can result in the overidentification of boys as having RD (Van der Wissel & Zegers, 1985), particularly when coupled with their greater reading score variance. In our study, the standard deviation of boys' reading scores was several points larger than that for girls (see also Anastasi, 1976). Picture the consequences of an extreme cutoff in the lower tail of two overlapping distributions (one for each gender) with the same mean but with a greater spread (i.e., platykurtic) for boys. The farther out in the tail the

TABLE 3
Means and Standard Deviations by Gender and Disability Status on Measures of Intelligence, Speech, Language, Motor Skills, Neurological Function, and Social Disadvantage

	Boys				Girls			
	RD[a]		NRD[b]		RD[c]		NRD[d]	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
WISC-R								
Information	8.6*	2.30	10.7	2.80	8.2*	2.68	9.7	2.48
Similarities	10.4*	3.59	12.2	3.66	9.0*	3.56	11.9	3.45
Arithmetic	8.5*	2.19	10.5	3.13	7.9*	2.62	10.4	2.80
Vocabulary	9.8*	2.15	10.9	2.89	8.5*	2.46	10.8	2.57
Picture completion	12.8	1.87	11.9	2.57	11.3	2.44	11.1	2.60
Block design	12.2	2.77	12.2	3.21	10.5	3.77	11.6	3.13
Object assembly	14.6	2.96	12.7	3.11	12.5	3.25	12.1	3.21
Coding	9.0	2.68	9.7	2.91	10.3	3.42	11.1	2.82
Verbal IQ	95.6*	12.43	106.8	15.65	90.0*	15.20	104.1	14.58
Performance IQ	115.3	13.28	111.5	15.68	108.2	18.06	110.4	15.46
Full scale IQ	105.3	12.15	109.9	15.55	98.3	16.58	107.8	15.15
PAT listening comprehension	39.0*	27.63	56.4	27.75	28.9*	18.75	48.9	26.67
DAC speech articulation	13.8*	3.49	16.0	3.06	14.9*	4.02	16.7	2.68
ITPA								
Comprehension	33.4	6.73	35.3	7.69	33.0	9.02	35.1	7.35
Expression	36.9	11.73	35.7	8.95	34.5	9.51	35.8	9.08
BMAT Motor skills	.09	.49	.07	.47	–.27	.56	–.02	.48
Measure	%		%		%		%	
Neurological assessment	3.9		0.0		3.1		3.3	
Social disadvantage	7.1		9.5		9.7		8.1	

Note. RD = reading disabilities; NRD = no reading disabilities; WISC-R = *Wechsler Intelligence Scale for Children–Revised* (Wechsler, 1974); PAT = *Progressive Achievement Tests* (Elley & Reid, 1971); DAC = *Dunedin Articulation Check* (Justin, Lawn, & Silva, 1983); ITPA = *Illinois Test of Psycholinguistic Abilities* (Kirk, McCarthy, & Kirk, 1968); BMAT = *Basic Motor Ability Test* (Arnheim & Sinclair, 1974). Neurological assessment (Touwen & Prechtl, 1970) score reflects percentage of participants with at least two neurological abnormalities. Social disadvantage score reflects percentage of participants scoring 2 or higher on Rutter's adversity index (McGee, Williams, & Silva, 1985).
[a]n = 32. [b]n = 439. [c]n = 34. [d]n = 409.
*$p < .05$, comparing RD and NRD groups.

cutoff is set, the greater will be the gender imbalance.

It is important to note that all these potential artifacts can occur even if IQ is discarded. That is, even if RD were defined solely on the basis of reading scores, gender bias remains a genuine threat whenever the boys' mean is lower or the boys' variance greater. In the present study, IQ distributions (Verbal, Performance, and Full Scale) were very similar for boys and girls (both means and variances), demonstrating that the gender bias derives from the reading distributions, not from the IQ distributions. Applying the same logic of within-gender and combined-gender cutoffs using an absolute low-achievement criterion (1 *SD* below the mean) revealed that a total of 92 boys and 85 girls were classified as having RD on a within-gender basis. Combining the results for all children in the sample, 23 girls with RD are no longer classified as such, whereas 26 boys are now added to the RD group, creating a marked gender imbalance of almost two to one (118 RD boys compared to 62 RD girls). These data emphasize that the gender bias issue is just as applicable to post-discrepancy taxonomies (see, for example, Working Party of the Division of Educational and Child Psychology of the British Psychological Society, 1999) as it is to current IQ-discrepancy definitions.

These data beg the question of whether reading disability should be defined separately for boys and girls. This is a complex social policy issue and, as such, is well beyond the scope of the present study. Nonetheless, several preliminary comments can be offered.

The fact that reading disability is conventionally defined in reference to a child's age or current grade level recognizes both the reality of fundamental biological differences related to age and the social context of school-based learning and instruction. Gender, like

TABLE 4
Means and Standard Deviations by Gender and Disability Status on Neuropsychological Measures

	Boys				Girls			
	RD[a]		NRD[b]		RD[c]		NRD[d]	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
COWAT (words per minute)	10.6	2.77	12.09	3.42	10.8	3.77	12.5	3.38
RAVLT								
Immediate recall	11.2	2.52	11.9	2.23	11.7	3.29	12.5	2.03
Delayed recall	7.7	2.94	9.6	2.67	9.4	3.43	10.5	2.71
TMT (seconds to complete)								
Form A	17.5	5.16	18.1	5.25	29.9	9.58	16.9	5.90
Form B	43.4	16.92	38.0	21.06	46.0	20.66	34.9	17.90
ROCFT Errors								
Copying	32.8	2.69	32.5	3.68	30.0	7.60	32.7	3.77
Delayed recall	19.4	7.30	30.6	6.37	17.4	8.45	18.7	6.68
ROCFT Time (%)								
Copying	.63	.26	.56	.26	.67	.28	.54	.29
Delayed recall	188.0	52.29	200.4	70.68	195.6	52.11	192.3	57.13
WISC-R Mazes	25.9	3.06	25.5	3.10	24.1	4.02	25.2	3.36
GPBT (seconds)								
Preferred hand	96.1	10.08	95.3	10.98	95.7	10.30	92.4	11.35
Total both hands	133.2	14.35	135.6	17.87	151.3	26.91	140.0	21.34

Note. RD = reading disabilities; NRD = no reading disabilities; COWAT = *Controlled Oral Word Association Test* (Benton & Hamsher, 1978); RAVLT = *Rey Auditory Verbal Learning Test* (Rey, 1964); TMT = *Trail Making Test* (Lewinsohn, 1973); ROCFT = *Rey-Osterrieth Complex Figure Test* (Waber & Holms, 1985); WISC-R = *Wechsler Intelligence Scale for Children–Revised* (Wechsler, 1974); GPBT = *Grooved Pegboard Test* (Knights & Moule, 1968).
[a]$n = 32$. [b]$n = 439$. [c]$n = 34$. [d]$n = 409$.

TABLE 5
Means and Standard Deviations by Gender and Disability Status on Academic Achievement Measures

	Boys				Girls			
	RD[a]		NRD[b]		RD[c]		NRD[d]	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Reading comprehension	20.4*	12.35	54.0	28.36	20.0*	19.67	52.8	26.55
Spelling	5.9*	4.19	15.2	5.71	8.6*	4.72	17.7	4.14
Writing competence	38.0*	18.50	67.3	21.49	49.5*	16.39	80.1	19.46
Mathematics	26.3*	21.28	48.1	27.53	21.9*	17.39	48.0	24.78

Note. RD = reading disabilities; NRD = no reading disabilities. Reading comprehension (Elley & Reid, 1969) and mathematics (Reid & Hughes, 1974) reflect age-based percentile scores from the *Progressive Achievement Tests;* spelling scores from the *Dunedin Spelling Test* (Silva et al., 1981); writing competence scores from Adler (1986).
[a]$n = 32$. [b]$n = 439$. [c]$n = 34$. [d]$n = 409$.
*$p < .05$, comparing RD and NRD groups.

age, might be considered a basic biological variable, engendering differential expectations (particularly in early language development) with respect to academic outcomes and their prediction in specific social and cultural contexts. However, whereas in-school learning is almost invariably segregated by age, segregation by gender—in Western cultures at least—is far less common. When schooling is gender-segregated (e.g., for religious or ideological reasons), it would seem appropriate to define reading disability by reference to the specific context of gender-segregated learning. Such a contextual approach draws support from recommendations of the Committee on the Prevention of Reading Difficulties in Young Children (Snow, Burns, & Griffin, 1998) regarding the needs of children (typically from affluent families) who are reading poorly relative to their classmates but who attend schools where the distribution of reading scores is well above the national average. On the question of whether such children should qualify for special support in spite of the fact that their reading scores do not fall below a cutoff point (e.g., 25th percentile) based on national norms, the committee's affirmative answer was based on research evidence showing that such children read in ways *similar* to failing readers defined by conventional criteria and that these children are also at risk for the same negative educational and occupational *outcomes* as other poor readers. In the present gender-relative context, the first criterion certainly appears, because the differences between the present sample of girls and boys with RD were found to be quantitative rather than qualitative. The issue of long-term outcomes for boys with RD compared to girls with RD remains to be examined.

ABOUT THE AUTHORS

David L. Share, *dshare@construct.haifa.ac.il.* ***Phil A. Silva,*** *PhD, is chairman of the D.A.R.E. Foundation of New Zealand, an adjunct professor in the faculty of health studies at Auckland University of Technology, and a principal investigator in the Pacific Islands Families Study. Address: David L. Share, Faculty of Education, University of Haifa, Mt. Carmel, 31905, Haifa, Israel.*

REFERENCES

Adler, C. (1986). *The development of descriptive criteria and an assessment of children's writing language.* Unpublished MA thesis, University of Otago, New Zealand.

Arnheim, D. D., & Sinclair, W. A. (1974). *The clumsy child.* St. Louis: Mosby.

Anastasi, A. (1976). *Psychological testing* (4th ed.). New York: Macmillan.

Benton, A. L., & Hamsher, K. de S. (1978). *Multilingual aphasia examination manual.* Iowa City: University of Iowa.

Bishop, D. V. M., & Butterworth, G. E. (1980). Verbal performance discrepancies: Relationship to birth risk and specific reading retardation. *Cortex, 16,* 375–389.

Department of Statistics. (1976). *New Zealand yearbook.* Wellington: New Zealand Government Printer.

Elley, W. B., & Irving, J. C. (1972). A socioeconomic index for New Zealand based on levels of education and income from the 1966 census. *New Zealand Journal of Education Studies, 7,* 155–167.

Elley, W. B., & Reid, N. A. (1969). *Progressive Achievement Tests: Teachers' manual. Reading comprehension–Reading vocabulary.* Wellington: New Zealand Council for Educational Research.

Elley, W. B., & Reid, N. A. (1971). *Progressive Achievement Tests: Teachers' manual, Listening comprehension.* Wellington: New Zealand Council for Educational Research.

Flynn, J. M., & Rahbar, M. H. (1994). Prevalence of reading failure in boys compared with girls. *Psychology in the Schools, 31,* 66–71.

Geschwind, N., & Galaburda, A. M. (1985). Cerebral lateralisation. Biological mechanisms, associations and pathology. II: A hypothesis and a programme for research. *Archives of Neurology, 42,* 521–552.

Grove, W. M., & Andreasen, N. C. (1982). Simultaneous tests of many hypotheses in exploratory research. *Journal of Nervous and Mental Disease, 170,* 3–8.

Jensen, A. R. (1980). *Bias in mental testing.* London: Methuen.

Jorm, A. F., Share, D. L, Maclean, R., & Matthews, R. (1986). Cognitive factors at school entry predictive of specific reading retardation and general reading backwardness: A research note. *Journal of Child Psychology and Psychiatry, 27,* 45–54.

Justin, C., Lawn, L., & Silva, P. A. (1983). *The Dunedin Articulation Check (DAC): A clinical instrument from the Dunedin Multidisciplinary Health and Development Research Unit.* Dunedin, New Zealand: Otago Speech Therapy Association.

Kirk, S. A., McCarthy, J. J., & Kirk, W. D. (1968). *The Illinois test of psycholinguistic abilities* (rev. ed.) Urbana: University of Illinois Press.

Klove, H. (1963). Clinical neuropsychology. In F. M. Forster (Ed.), *The medical clinics of North America* (pp. 1647–1658). New York: Saunders.

Knights, R. M., & Moule, A. D. (1968). Normative data on the motor steadiness battery for children. *Perceptual and Motor Skills, 26,* 643–650.

Lewinsohn, P. M. (1973). *Psychological assessment of patients with brain injury.* Unpublished manuscript, University of Oregon, Eugene.

McGee, R. O., Clarkson, J. E., Silva, P. A., & Williams, S. M. (1982). Neurological dysfunction in a large sample of three year old children: A report from the Dunedin Multidisciplinary Child Development Study. *New Zealand Medical Journal, 95,* 693–696.

McGee, R., & Silva, P. A. (1982). *A thousand New Zealand children: Their health and development from birth to seven* [Special report no. 8]. Auckland: Medical Research Council of New Zealand.

McGee, R., Williams, S. M., & Silva, P. A. (1985). The factor structure and correlates of ratings of inattention, hyperactivity, and antisocial behaviour in a large sample of nine year old children from the general population. *Journal of Consulting and Clinical Psychology, 53,* 480–490.

Naiden, N. (1976). Ratio of boys to girls among disabled readers. *The Reading Teacher, 29,* 439–442.

Osterrieth, P. A. (1944). Le test de copie d'une figure complexe [Complex figure copying test]. *Archives de Psychologie, 30,* 206–356.

Reid, N. A., & Hughes, D. C. (1974). *Progressive Achievement Tests: Teachers' manual. Mathematics.* Wellington: New Zealand Council for Educational Research.

Reitan, R. M. (1958). Validity of the Trail Making Test as an indication of organic brain damage. *Perceptual and Motor Skills, 8,* 271–276.

Rey, A. (1964). *L'examen clinique en psychologie* [Clinical assessment in psychology]. Paris: Presses Universitaires de France.

Rutter, M. (1978). Family, area and school influences in the genesis of conduct disorders. In L. A. Hersov, M. Berger, & D. Schaffer (Eds.), *Aggression and antisocial behaviour in childhood and adolescence* (pp. 95–113). Oxford, UK: Pergamon Press.

Rutter, M., Tizard, J., & Whitmore, K. (1970). *Education, Health and Behaviour*. London: Longman.

Rutter, M., & Yule, W. (1975). The concept of specific reading retardation. *Journal of Child Psychology and Psychiatry, 16,* 181–197.

Scottish Council for Research in Education. (1976). *The Burt word reading test, 1974 revision.* London: Hodder & Stoughton.

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,* 97–100.

Share, D. L., McGee, R. O., McKenzie, D., Williams, S. M., & Silva, P. A. (1987). Further evidence relating to the distinction between specific reading retardation and general reading backwardness. *British Journal of Developmental Psychology, 5,* 35–44.

Share, D. L., & Silva, P. A. (1987) Language deficits and specific reading retardation. *British Journal of Disorders of Communication, 22,* 219–226.

Shaywitz, S. E., Shaywitz, B. A., Fletcher, J. M., & Escobar, M. D. (1990). Prevalence of reading disability in boys and girls. *Journal of the American Medical Association, 264,* 998–1002.

Silva, P. A. (1984). *The educational attainment of 2,600 Dunedin children: A report on the Dunedin Intermediate Schools' Collaborative Testing Project 1984: The NZCER tests.* Dunedin, New Zealand: Dunedin Multidisciplinary Health and Development Research Unit.

Silva, P. A., McGee, R. O., & Williams, S. M (1985). Some characteristics of nine year old boys with general reading backwardness and specific reading retardation. *Journal of Child Psychology and Psychiatry, 26,* 407–421.

Silva, P. A., Smith, C., & Pearce, D. (1984). *The Dunedin spelling tests.* Dunedin, New Zealand: University of Otago.

Silva, P. A., & Stanton, W. R. (1996). *From child to adult: The Dunedin Multidisciplinary Health and Development Study.* Auckland, New Zealand: Oxford University Press.

Snow, C., Burns, M., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

Tallal, P., Ross, R., & Curtiss, S. (1989). Unexpected gender ratios in families of language/learning-impaired children. *Neuropsychologia, 27,* 987–998.

Taylor, E. M. (1959). *The appraisal of children with cerebral deficits.* Cambridge, MA: Howard University Press.

Touwen, R. C., & Prechtl, H. R. (1970). *The neurological examination of the child with minor nervous dysfunction.* London: Heineman.

Van der Wissel, A., & Zegers, F. C. (1985). Reading retardation revisited. *British Journal of Developmental Psychology, 44,* 1–12.

Vogel, S. A. (1990). Gender differences in intelligence, language, visual–motor abilities, and academic achievement in males and females with learning disabilities: A review of the literature. *Journal of Learning Disabilities, 23,* 44–52.

Wadsworth, S. J., DeFries, J. C., Stevenson, J., Gilger, J. W., & Pennington, B. F. (1992). Gender ratios among reading-disabled children and their siblings as a function of parental impairment. *Journal of Child Psychology and Psychiatry, 33,* 1229–1239.

Waber, D. P., & Holmes, J. M. (1985). Assessing children's copy productions of the Rey-Osterrieth complex figure. *Journal of Clinical and Experimental Psychology, 7,* 264–280.

Wechsler, D. (1974). *Manual of the Wechsler Intelligence Scale for Children–Revised.* New York: Psychological Corp.

Working Party of the Division of Educational and Child Psychology of the British Psychological Society. (1999). *Dyslexia, literacy and psychological assessment.* Leicester, UK: British Psychological Society.