Order Number 1336903

Calculating the correlation coefficient between selected ability and achievement tests using validity generalization

Shiroma, Paul Shigeo, M.A.

The University of Arizona, 1989

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

CALCULATING THE CORRELATION COEFFICIENT BETWEEN SELECTED ABILITY AND ACHIEVEMENT TESTS USING VALIDITY GENERALIZATION

by

Paul Shigeo Shiroma

A Thesis Submitted to the Faculty of the

DIVISION OF EDUCATIONAL FOUNDATIONS AND ADMINISTRATION

In Partial Fulfillment of the Requirements
For the Degree of

MASTER OF ARTS
WITH A MAJOR IN EDUCATIONAL PSYCHOLOGY

In the Graduate College

THE UNIVERSITY OF ARIZONA

1 9 8 9

STATEMENT BY THE AUTHOR

This thesis has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under the rules of the Library.

Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgement the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Paul Shyue Shin

APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:

Darrell Sabers

Darrell Sabers, Ph.D.
Professor of Educational Psychology

4/11/89
Date

ACKNOWLEDGMENTS

If life's challenges were viewed as closed doors, I would like to thank my parents for pushing me to this one and making me knock, Dr. Darrell Sabers for opening the door and inviting me in, Ken and Imi Yamashita for making sure the framework didn't collapse, my wife for stepping through with me, and Dr. Lawrence Aleamoni and Dr. Harley Christiansen for assisting me through the educational passage.

Table of Contents

ABSTRACT

The correlation coefficient between an ability test and achievement test is necessary in order to estimate the effects of the regression of IQ on achievement when calculating a discrepancy between them for the purpose of classifying a child as learning disabled. Weighted mean correlations and their respective variances were computed across studies for one of 11 ability-achievement test pairs using a meta-analysis procedure called Validity Generalization. The results indicated that there is no "global correlation coefficient" that will accurately represent the relationship between all ability and achievement tests. Furthermore, although the estimated variation of correlations between specific test pairs were reduced to insignificant levels when studies were subgrouped according to clinical/educational classification, critical data necessary to adjust correlations and their respective variances for statistical artifacts were not reported in the literature. Thus, the results obtained from the subgroups could be due to capitalization on chance.

Calculating the Correlation Coefficient Between Selected Ability and Achievement Tests Using Validity Generalization

In classifying a child as learning disabled, an important criterion predicated by PL 94-142 is a "discrepancy between achievement and intellectual ability in one or more of seven areas relating to communication skills and mathematical abilities" (Federal Register, 1977, 42, p. 65082). The pragmatic effects of this mandate was demonstrated in various studies conducted since its inception. In their survey of 49 States, Frankenberger and Harper (1987) found an increase (16 States in 1981/1982 to 33 States in 1985/1986) in the number of States that utilized an ability-achievement discrepancy criterion in procedurally defining learning disabled students. Furthermore, Ysseldyke, Algozzine, Regan, and McGue (1981) found that 159 educators and school psychologists, experienced in determining learning disability classification, rated scores on achievement and intelligence tests, and the disparity between the two, as the greatest influence on both eligibility and classification of hypothetical students. Using naturalistic methodology, Huebner and Cummings (1986) obtained similar results. In their study, 30 psychologists were asked to provide data concerning the last two students that they recommended for special education services due to learning problems and the last two students that were referred but not recommended for special education services. The results of the analysis of these data indicated that the ability-achievement discrepancy was an essential criterion in determining the psychologists' recommendations.

Although the discrepancy between ability and achievement is established as a pragmatic, if not theoretical, indicator of a learning disability, there is no paramount procedure utilized to determine this discrepancy. Cone and Wilson (1981) identified four general mathematical procedures for determining an ability-achievement discrepancy. Of these procedures, the use of regression analysis was singularly recommended. Deviation from grade level, expectancy formulae, and standard score comparison were found to be technically inadequate although the use of standard score comparison was suggested as a supplemental procedure when "assessment instruments fail to meet accepappendix psychometric standards [Thurlow & Ysseldyke, 1977; Shepard & Smith, 1981] or the necessary correlation between the measures has not been adequately investigated". The major inadequacies found in the deviation from grade level and expectancy formulae were (1) the limited mathematical properties of ordinal data (i.e., grade level), (2) sensitivity to violations of the dubious assumption that rate of growth follows a consistent, stable, and linear path, and (3) inability to account for the effects of regression to the mean (Hanna, Dyck, & Holen, 1979; Shepard, 1980; Cone & Wilson, 1981; Reynolds, 1981; Wilson, Cone, Busch, & Allee, 1983; Berk, 1984). The standard-score procedure has been criticized because it does not account for the effects of regression to the mean (Shepard, 1980; Cone & Wilson, 1981; Berk, 1984; Wilson & Cone, 1984; Reynolds, 1984-1985) although the Department of Education, Special Education Programs Work Group on Measurement Issues in the Assessment of Learning Disabilities suggested that if its weakness were corrected, this procedure may be more appropriate than regression analysis (Reynolds, 1984-1985). Conversely, the Work Group recommended regression analysis as a statistically sound procedure due in part to its accountability for the effects of regression to the mean (Reynolds, 1984-1985). This characteristic of regression analysis was recognized as a critical component in the accurate measurement of a severe discrepancy (McLeod, 1979; Shepard, 1980; Cone & Wilson, 1981; Wilson, Cone, Busch, & Allee, 1983; Wilson & Cone, 1984; Reynolds, 1984-1985). In spite of this importance, researchers have neglected to discuss procedures for calculating a necessary component of the analysis; the correlation between tests of ability and achievement. Although the Work Group suggested that the calculation of correlation coefficients could be parsimoniously obtained if ability and achievement tests were co-normed (Reynolds, 1984-1985), Cone and Wilson (1981) implied that, among other practical problems, the financial cost of co-norming tests would be a formidable obstacle. They cited an alternative approach utilizing relevant data from past studies to estimate "regression parameters", but did not specify procedural steps.

Validity Generalization (VG) developed by Schmidt, Hunter, and Jackson (1986) is a statistical procedure whereby research findings (e.g., correlation coefficients) can be integrated across studies. Although past applications of this procedure were primarily in the field of employment testing, it has been successfully used in a variety of fields (Schmidt, 1988). The purpose of this study is to demonstrate the utility of VG procedures in calculating the correlation between ability-achievement and the implications for the measurement of ability-achievement discrepancy.

METHOD

Data Collection

A pool of research studies were obtained by using Bibliographic Research Services After Dark (BRS), a computer-based "search" service designed to provide references pertaining to various topics in various fields.

The articles were gathered from two "data bases" (Psyc Info and ERIC), chosen because of their content relevance to this study. Both data bases provided references for studies published from 1966 to a maximum of two months prior to the date of the search. Initially, 379 articles containing reference to specific pairs of individually administered ability and achievement tests were selected from these data bases. The selection process (search) utilized a "free text term" which meant that any article with the specific test pair mentioned in its title, abstract, or descriptor field was included in the pool of research studies. Only individually administered tests were included because group administered tests are inappropriate in determining ability-achievement discrepancies due to the lack of situational/environmental control (Reynolds, 1984-1985).

The following 11 test-pairs were investigated: (1)Peabody Individual Achievement Test (PIAT)/Peabody Picture Vocabulary Test-Revised (PPVT-R), (2)PIAT/Wechsler Intelligence Scale for Children-Revised (WISC-R), (3)WISC-R/Wide Range Achievement Test (WRAT), (4)WISC-R/Woodcock Johnson Psychoeducational Battery (WJ), (5)WRAT/WJ, (6)PPVT-R/WJ, (7)Kaufman Assessment Battery for Children (K-ABC)/WJ, (8)K-ABC/PIAT, (9)PPVT-R/WRAT, (10)PIAT/WJ, (11)WRAT/K-ABC.

From the initial pool of articles, approximately 73 studies contained relevant and sufficient information required for data analysis using VG procedures. Some of these studies used more than one sample of subjects and therefore, a correlation coefficient for one or more of the previously mentioned test pairs was available for a total of 86 samples. The type and frequency of descriptive information that was reported on these samples are presented in the proceeding paragraphs. These sample characteristics are the basis for the selection of moderator variables which will be used to delineate homogeneous subgroups of samples. Subgrouping samples is an important aspect of VG procedures and will be discussed in the next section.

Information on socioeconomic status (SES) was available for 16 samples or 19% of the total. All of the subjects in these samples were classified middle, lower-middle, or low SES with close to 63% of the subjects in the 16 samples classified lower-middle to low SES. Geographic location for 11 samples of children were available (13%) with approximately 76% of the subjects residing in urban areas (in contrast to rural areas). The percentage of males and females for 42 (50%) samples were ascertained and nearly 62% of the subjects were male. The racial composition of 33 samples of children (39%) were as follows: 15%-Navajo, 55%-Caucasian, 11%-Black, 18%-Mexican-American. The mean age of the children ranged from 7 to 13 years old for 63 samples (74%). The majority of the mean ages (66%) clustered between 8.6 and 10.6 years with 11% below and 23% above this range. In many of the samples, mean-age was accompanied by the age range of the children. Over 75% (47) of the samples with mean-age data also contained age ranges for subjects and overall, 62% (53) of all the samples contained subjects' age range. The average span of the range across the 53 samples was 5.8 years.

In addition to demographic information, data were collected on clinical/educational classification and test scores. Of the 85 samples, 35% or 30 samples were identified with clinical/educational classifications. An additional 36% (31) of the total number of samples were referred for assessment and 3% (3) had been given medical diagnosis. The percentages of clinical/educational classification for sample subjects were as follows: 46%-Learning Disabled (LD), 18%-mixed diagnosis, 13%-Mentally Retarded (MR), 7%-Emotionally Handicapped (EH), 8%-Gifted, 3%-Behavior disordered (BD), 5%-Low Achievers (LA). For those children referred for assessment, 43% were referred for specific learning problems while the purpose of the remaining 57% of the referrals were either unspecified or multiple in nature. The 3% of the total number of study samples that utilized medical diagnostic descriptor were composed of children with Spina Bifida (48%) and sensori-neural impairment (52%).

Scores for mean ability and achievement measures were available for 50 (59%) and 46 (54%) samples, respectively. Although the range of scores varied considerably across samples (from 50 to 120 points on ability measures and 60 to greater than 130 points on achievement measures), more than 84% of the ability and achievement test scores ranged between 70 to 109 points and 73% fell within a 29 point range (80 to 109 for ability measures and 70 to 99 for achievement measures). Furthermore, test scores above 109 and below 60 accounted for less than 7% of all sample tests score means.

Procedure

VG procedures allow the researcher to determine the magnitude and variance of the relationship between two variables across studies. Furthermore, systematic variance across studies can be accounted for if sufficient descriptive data for the studies are available. The basic procedures are described below. A more in-depth description and rationale for the procedures are presented by Schmidt, Hunter, and Jackson (1986).

The magnitude of the relationship between two variables is determined by weighing the correlation coefficient reported for each study by its respective sample size, summing these values across studies, and dividing by the sum total of subjects across studies. Thus, the magnitude of the relationship can be expressed as a weighted mean correlation coefficient. Furthermore, in most cases this weighted mean is the best estimate of the population correlation coefficient. The variance of the weighted mean correlation coefficient is determined by weighing the squared difference between the estimated population correlation coefficient and the correlation coefficient reported for each study by its respective sample size, summing these values across studies, and dividing by the total sum of the sample sizes across studies. In order to estimate the variance of the population correlation coefficient, the variance of the correlation coefficient across studies is adjusted for sampling error. Sampling error is calculated by essentially weighing the square of the variance unaccounted for by the estimated population correlation coefficient by the total number of studies and dividing by the sum total of the samples across studies. VG procedures also contain provisions for adjusting the variance for test reliability and sample range restriction/enhancement provided that measures of test reliability and standard deviations are provided. Using the Chi Square Distribution and the test of significance outlined by Schmidt, Hunter, and Jackson (1986) the probability of obtaining a specific estimated population variance can be determined. If this probability is greater than expected by chance and assuming that there are sufficient descriptive data, then the variables that account for the variance can be identified. This is a critical component of VG procedures because when there is a large adjusted variance the underlying assumption is that the population, represented by the aggregated study samples, is heterogeneous and can be subgrouped into more homogenous populations.

The identification of homogeneous populations has two distinct advantages: (1) greater accuracy of measurements and (2) group differences become identifiable. Logically, one would seek to establish subgroups based on the characteristic or characteristics that effect the relationship under study. That is, if there is systematic variance in the relationship between two variables across studies, one should be able to account for this variance by delineating subgroups of samples based on the characteristic or characteristics that impact the relationship. For example, test scores on achievement and ability measures may be effected by ethnicity and therefore, it is plausible that subgroups based on ethnicity would account for some of the variance in the correlation coefficient between an achievement and an ability measure across studies. Once subgroups are established, it is possible to calculate the weighted mean and variance of the correlation coefficient across the studies in each subgroup. A meaningful subgroup is one in which the variance of the weighted mean for that subgroup is zero or near zero (i.e., homogeneous in nature). If the variance of the weighted mean of a subgroup is greater than that expected by chance, then this subgroup can again be delineated by the reiterative process described above. Through this process a more meaningful measure of the relationship for specific subgroups is estimated. The descriptive data available for study samples and the logic for selecting certain characteristics (i.e., moderator variables) as a basis for differentiation is critical to reducing variance. A summary description of the characteristics for the samples that were analyzed in this study has been previously given and the following section provides the rational for the moderator variable that was selected as a basis of group differentiation.

Rational for group differentiation

In VG procedures, the purpose for group differentiation of study samples is to reduce the variance of, in this case, the correlation coefficient across studies and thus provide evidence that elucidates group differences on the measure under scrutiny. Group differentiation is accomplished through the use of variables that described study sample characteristics or "moderator variables".

The rational for determining moderator variables upon which groups were separated was influenced by two factors: (1) the quality and quantity of the available descriptive variables and (2) the purpose of the research study. Therefore, although other variables may influence LD placement, clinical/educational classification was the only appropriate basis upon which to determine the moderator variable in light of the quality and quantity of descriptive variables, as well as the purpose of this study. The following subgroup descriptions, based on the moderator variable, are listed below.

(1) Referred Learning Disabled (RLD)-Samples of children that were classified as LD or referred for specific learning problems.

(2) Referred for psychological evaluation (R)-Samples of children that were referred for an unspecified or multiple reasons.

(3) Miscellaneous Diagnosis (Mx Diag)-Samples of children that were classified EH, BD, LA, MR, medically diagnosed with a sensori-neural impairment or Spina Bifida. Additionally, samples with a mixture of classified and unclassified children or children who were clinically or educationally labeled but could not be differentiated by this label were included in this group.

(4) Unidentified (UI)-Samples of children that were not classified/diagnosed.

If a moderator variable contained less than two study samples, it was not included as a category in the analysis and the study sample was placed in the next most similar moderator variable category. For instance, if the UI subgroup contained fewer than two study samples, the study sample would be combined with the MX Diag subgroup. Similarly, if the R subgroup contained only one study sample, it would be placed in the Mx Diag subgroup. No group differentiations were made if only one RLD study sample was available.

RESULTS AND DISCUSSION

Validity Generalization Results

Coefficients for the PPVT-R/WJ, K-ABC/WJ, PPVT-R/WRAT test-pairs were found in no more than a single study and none of the studies reporting PIAT/WJ subtest coefficients contained more than one coefficient between specific subtest pairs. From the remaining nine test-pairs the weighted average correlation, the variance of the correlation, and the correction for the effects of sampling error were calculated for 75 subtest and total test score combinations based on the 86 study samples of children. Insufficient data on test reliability for study samples would necessitate the assumption of a known reliability and zero variation of reliability across studies if adjustments for the effects of range restriction/enhancement were to be made with VG procedures. Therefore, although there was a sufficient number of studies to adjust for range restriction/enhancement, the unavailability of data on the reliability of test scores for study samples precluded the utilization of VG procedures to correct for attenuation due to test unreliability and adjust for the effects of range restriction/enhancement of study samples on the correlation coefficient. The results of VG procedures for specific ability-achievement measures are presented in Appendix A.

Using Schmidt, Hunter, and Jackson's (1986) test of significance for variance across studies, I found that the estimated standard deviation for 18 of the 75 coefficients were significantly greater than expected by chance ($p < .01$). Data from three coefficients were not calculated because it was not possible to differentiate subgroups. The following adjustments were made to the subgroups of the remaining 15 coefficients due to the lack of sufficient study samples. A study sample belonging in the UI subgroup was placed in the R subgroup under the WISC-R FSIQ/PIAT General Information coefficient. The UI subgroup consisted entirely of Navajo children for the coefficients between the WISC-R VIQ, PIQ, FSIQ ability measures and PIAT Total Test achievement measure, and a study sample belonging in the R subgroup was placed in the MX Diag subgroup. Finally, for the coefficients between the WISC-R PIQ and FSIQ measures and PIAT Reading Recognition measure, study samples of a R and UI subgroup were placed in the MX Diag subgroup. The results yielded by VG procedures on the subgroups are presented in Appendix B.

Discussion

The ability to yield a variability measure of a specific outcome (e.g., correlation coefficient) across studies is the essence of utilizing VG procedures. The importance of accounting for variability across studies is demonstrated using the WISC-R Full Scale IQ/WRAT Math subtest (FSIQ/MATH) correlation coefficient.

Initially the correlation between the FSIQ/MATH across 36 studies and a total sample size of 2679 was .5026 with an estimated standard deviation of .1396 (see Appendix A) which was significantly greater than expected by chance ($p < .01$). The results of categorizing the 36 studies into subgroups (i.e., UI, R, RLD, and MX Diag) were presented in Appendix B. The estimated variances of the studies across each of three respective subgroups (R, RLD, and MX Diag), as indicated by their estimated standard deviations, were smaller than the estimated variance across all 36 studies. The effects of categorizing studies according to sample characteristics were shown using the studies that comprised the UI subgroup (see Appendix C). Visual inspection of the correlation coefficients in Appendix C indicates that the greater range of the coefficients accounted for the greater variability of the studies in the UI subgroup as compared to the variability of the studies in the other subgroups. Based on the descriptive variables presented for the studies in the UI subgroup, the moderator variables that differentiated between the lower and higher coefficients were (1) a combined gifted and minority (i.e., Hispanic and Navajo) subgroup and (2) a combined Anglo and unidentified subgroup. The results of VG procedures on the subgroups differentiated by these two moderator variables were presented in Appendix D and show that the estimated standard deviations of both groups were insignificant ($p < .01$). The use of moderator variables to successively differentiate subgroups of the 36 studies used to determine the relationship between the FSIQ/MATH subtest resulted in a corresponding decrease in variability within each subgroup. These results suggests that the variation across studies of the correlation coefficient for the FSIQ/MATH were due to specific sample characteristics (e.g., referral for psychological evaluation). Thus, it is prudent to recognize the effects of these variables when utilizing the FSIQ/MATH correlation coefficient. There are two limitations that not only apply to the results mentioned above but to all the results presented in Appendices A and B. The first limitation involves the adjustments for the effects of sample range restriction/enhancement and attenuation due to test unreliability on the variance of the coefficient. The second limitation is based on the concept of capitalization on chance.

<u>An hypothetical example</u>

Assume that the reliability measures of the FSIQ/MATH had means of .80 and variances of .05 across studies. Then using the available test score standard deviations from the studies comprising the FSIQ/MATH coefficient to adjust for sample range restriction/enhancement, and the assumed reliability measures and their variances to adjust for attenuation, the resulting estimated variance is zero. Thus, accounting for the effects of test unreliability and sample range restriction/enhancement on the variance of the correlation coefficient across studies may reduce a statistically significant estimated correlation variance to no variance across studies. Indeed, if the assumptions that were made for the FSIQ/MATH coefficient (i.e., reliability was .80 with a variance of .05) were imposed on all the coefficients in Appendix A, the estimated variances of only 15 out of the 75 coefficients would be greater than zero, and of these 15 coefficients, only 5 would vary significantly greater than would be expected by chance ($p <$

.01).

Threat of capitalization on chance

Schmidt, Hunter, and Jackson (1986) stated that if artifacts (i.e., sampling error, correction for attenuation due to test unreliability, and study sample range restriction/enhancement) were not corrected for, then the relationship of study sample outcome with study sample characteristics may be due to these artifacts. Furthermore, the higher the ratio between study sample characteristics and the number of study samples, the greater the potential that the relationship between variation across study samples and study sample characteristics were due to chance. With these points in mind, the results presented in Appendices A and B, and particularly results based on subgroups with small numbers of study samples such as the WISC-R FSIQ/PIAT Total Test subgroup results (3 subgroups across 9 studies), must be interpreted and used with caution. The variabilities across test pairs and within test pairs have important implications when calculating an aptitude-achievement discrepancy for the use of classifying a learning disabled population. As displayed in Appendix A, the relationship between aptitude and achievement measures across test pairs varied considerably; .2658 (K-ABC Sequential Processing/PIAT Spelling) to .8000 (WJ Broad Cognitive/WRAT Reading. The relationship within a specific aptitude-achievement test pair across defined subgroups of children also varied considerably. For instance, the correlation coefficients for the four WISC-R FSIQ/WRAT Math subtest subgroups ranged from .2972 (MX Diag) to .5713 (R). In addition to the aforementioned limitations there is another concern when interpreting subgroup coefficients. The integrity of the moderator variables directly affects the quality of the coefficients derived from these variables.

Threats due to subgroup delineation

There is evidence to suggest that utilizing school-based educational diagnosis of LD to differentiate samples of subjects is a precarious practice. Morrison, MacMillan, and Kavale (1985) identified two general problems when school-based LD samples were used in research. There were variations in the situations which caused a child to be referred for a learning disability and there were variations in the identification process. A child may be referred for LD placement for reasons other than academic difficulties, and the assessment process, definition of LD, and actual placement may vary considerably across and within settings. In accord with Morrison et al., Ysseldyke and Algozzine (1979) suggested that bias during the pre-assessment, assessment, and post-assessment stages negatively affected the integrity of the school-based LD sample. Sigman (1985) and Torgenson (1986) argued that school-based LD samples were influenced by socio-political forces. The former introduced the concept that the school system using the classification of LD to explain the system's inadequacies and the latter suggested that even if uniform assessment procedures were enacted, social and political forces (e.g., funding for programs) would bias school-based samples. Large scale surveys of professionals involved in the identification and placement of LD children in 14 school districts (Perlmutter & Parus, 1983) and in Child Service Demonstration Centers (Mann, Davis, Boyer, Metz, & Wolford, 1983; Thurlow & Ysseldyke, 1979) uniformly supported the contention that LD placement procedures greatly varied and resulted in inconsistent placement decisions. In their examination of the files of a probability sample of 800 children served as LD in Colorado, Shepard, Smith, and Vojir (1983) found that less than 50% of the sample had characteristics associated with the federal definition of LD. Friedrich, Fuller, and Davis (1984) examined the discriminative power of 94 empirically derived formulas for assessing learning disabilities for 1,600 referred children and concluded that the definition of LD is so vague that it "does not exist as a unique psychological entity". On a myopic level, Ysseldyke, Algozzine, Richey, and Graden (1982) reviewed 20 randomly sampled videotapes of placement team meetings and using a coding scheme, found that although 14 of the 20 referred children were declared eligible for LD services most of the data presented in these meetings were irrelevant to the Federal Criteria for LD placement.

In direct contrast to this evidence, Wilson and his colleagues examined data from 1,839 referrals from 7 special education agencies in Iowa and concluded that placement decisions based on aptitude-achievement discrepancies and exclusionary clauses could adequately differentiate an LD population (Cone, Wilson, Bradley, & Reese, 1985; Wilson, Cone, Bradley, & Reese, 1986; Wilson, 1985). Furlong and Yanagida (1985), utilizing MANOVA and discriminant analysis procedures on 117 LD and 106 non-LD children in two school districts, also found that although profiles of LD children were different across the two districts, the severity of academic underachievement consistently differentiated LD from non-LD children. Finally, through the discriminant function analysis of archival test data based on 80 files obtained from 30 school psychologists Huebner and Cummings (1986) found that the aptitude-achievement discrepancy score differentiated between children placed in an LD program and children placed in regular education.

Implications

The VG results on subgroup subtest-pairs tended to support the position of Wilson and his colleagues. The estimated standard deviations of 10 of the RLD subgroups were smaller than their respective subtest pair indicating that the relationship between aptitude and achievement measures did not appreciably vary for this subgroup across studies. Furthermore, only one RLD subgroup subtest pair estimated variance was significantly greater than what would be expected by chance ($p<.01$). These results also supported the finding of Ysseldyke, Algozzine, Regan, and McGue (1981) who in analyzing computer generated fictional data on the effects of various factors (aptitude/achievement measures, discrepancy scores) influencing the placement decisions of 159 educators and school psychologists found that the reason for referral significantly determined placement. Research conducted by O'Reilly, Northcraft, and Sabers (1989) support this effect. Thus, a child referred for a learning disability would probably be placed in an LD program. Assuming this is the case, children referred for learning disabilities and children actually labelled LD would display similar characteristics. This is indicated by the generally lower estimated standard deviation of the RLD subgroups compared to their respective subtest-pair correlation estimated standard deviation. The critical assumption that cannot be made from these data is that this subgroup represents a truly LD sample in the sense that the subgroup samples demonstrated a discrepancy between aptitude and achievement. Some research provides evidence that groups of children labeled as LD may indeed be homogeneous, but the discriminating characteristic of the group is that of low achievement and not a severe discrepancy between aptitude and achievement (Epps, Ysseldyke, & Algozzine, 1985; Algozzine & Ysseldyke, 1983). Further validation of the correlation coefficient for the RLD subgroup is necessary. Of particular value to this validation procedure is the accurate and consistent identification of moderator variables. Indeed, the lack of descriptive data in LD research has been documented by Olson and Mealor (1981). They reviewed 113 studies on LD samples and found that 37% of the researchers described their sample population only in terms of label or placement. Furthermore, age range and intelligence were highly variable across studies. Both Lovitt and Jenkins (1979) and Wong (1986) offered solutions to this problem. The former suggested that situational and instructional variables, demographic features, and motivational levels be specified in sample descriptions and the latter suggested selecting operationally defined "target characteristics", cluster analysis, or marker variables that included descriptive (age, sample size, etc.), substantive (aptitude, achievement measures, etc.), topical (activity level, attention span, etc.), and background markers (years of study, exclusionary criteria, etc.). In light of VG procedures, it is highly advantageous to develop uniform identification procedures for samples so that well defined groups of samples can be differentiated through the use of moderator variables.

CONCLUSION

The argument of whether an aptitude-achievement discrepancy is a valid measure of LD is not an important issue from the viewpoint of operationalizing the Federal Definition of LD. It is arguable that the utility of this measure is indeterminable until such time that a mathematically sound equation is developed and systematically integrated into LD research. It is generally accepted that this equation must include an adjustment for regression to the mean, and this adjustment is dependant upon the relationship between specific aptitude-achievement measures. This study demonstrated a statistical technique that allows one to calculate such a measure and provides initial estimates of this relationship for selected measures. The incorporation of additional data into the calculation of coefficients will lead to a precise understanding of the effects of various moderator variables on the relationship between ability and achievement measures. Three recommendations can be made based on the results of this study. (1) The data indicate that there is no "global" aptitude-achievement correlation coefficient.

The specific ability and achievement measures, and the characteristics of the population will affect the coefficient. (2) The interpretation of the data is limited because correlation coefficients and standard deviations could not be adjusted for the effects of two statistical artifacts (i.e., test unreliability and sample range restriction/enhancement). (3) Further research is needed to develop a statistical model that can be used to determine the probability of the estimated standard deviation of a correlation coefficient occurring with respect to a specific number of studies and total sample size.

Appendix A

Validity Generalization Test-Pair Results

Achievement test	Number of studies	Number of students	Correlation Weighted mean	Correlation Estimated SD
		WISC-R Verbal IQ		
PIAT				
RR	9	1099	.6064	.0793*
RC	10	1221	.6214	.0557*
Spelling	9	1099	.5146	.0762*
GI	9	1099	.7353	.0714
Math	9	1099	.6448	.0671
Total Test	9	954	.7271	.0707**
WOODCOCK-JOHNSON				
Reading	4	326	.5070	.0436
Written Language	4	326	.4713	.0387
Knowledge	3	197	.6085	.1249*
Skill	2	149	.4268	0
Math	4	326	.4937	0

Achievement test	Number of studies	Number of students	Correlation Weighted mean	Correlation Estimated SD
		WISC-R Verbal IQ		
WRAT				
Reading	36	2688	.4928	.1411**
Spelling	35	2537	.4413	.1349**
Math	36	2575	.5240	.1225**
		WISC-R Performance IQ		
PIAT				
RR	9	1099	.4810	.1063**
RC	10	1221	.4339	.0700*
Spelling	9	1099	.3369	.0300*
GI	9	1099	.4658	.0361
Math	9	1099	.4684	.0436
Total Test	9	954	.4910	.1136**
WRAT				
Reading	33	2702	.2839	.1473**
Spelling	31	2485	.2783	.1503**
Math	32	2523	.4284	.1241**

Achievement test	Number of studies	Number of students	Correlation Weighted mean	Correlation Estimated SD
		WISC-R Performance IQ		
WOODCOCK-JOHNSON				
Reading	4	326	.2896	.0849
Written Language	4	326	.4186	.2744**
Knowledge	3	197	.5367	0
Skill	2	149	.6223	.1954**
Math	4	326	.3367	.0469
		WISC-R Full Scale IQ		
WRAT				
Reading	40	2809	.4313	.1446**
Spelling	33	2523	.4203	.1797**
Math	36	2679	.5026	.1396**
WOODCOCK-JOHNSON				
Reading	4	326	.4792	.0906
Written Language	4	326	.4235	.1453*
Knowledge	2	158	.3899	.2232**
Skill	2	149	.3903	0
Math	4	326	.4882	0

Achievement test	Number of studies	Number of students	Correlation Weighted mean	Correlation Estimated SD
		WISC-R Full Scale IQ		
PIAT				
RR	9	1099	.6264	.0964**
RC	10	1221	.5968	.0583
Spelling	9	1099	.4838	.0346
GI	12	1343	.6752	.1082**
Math	9	1099	.6579	.0520
Total Test	9	954	.7025	.0806**
		K-ABC Sequential Processing		
WRAT				
Reading	6	228	.4944	0
Spelling	6	228	.3291	0
Math	6	228	.3257	0
PIAT				
RR	5	188	.3734	.1277
RC	4	150	.5205	.0469
Spelling	4	137	.2658	.1428
GI	4	137	.3234	.1439

Achievement test	Number of studies	Number of students	Correlation Weighted mean	Correlation Estimated SD
K-ABC Sequential Processing				
PIAT				
Math	4	137	.3834	.1277
Total Test	6	223	.3922	.1072
K-ABC Mental Processing				
PIAT				
RR	4	137	.4323	.1063
RC	4	150	.5874	.1175
Spelling	4	137	.4232	0
GI	4	137	.4423	.0374
Math	4	137	.5045	0
Total Test	5	172	.4481	.0922
WRAT				
Reading	6	228	.4514	.1288
Spelling	6	228	.3631	.1382
Math	6	228	.5106	.0458

Achievement test	Number of studies	Number of students	Correlation Weighted mean	Correlation Estimated SD
		K-ABC Simultaneous Processing		
WRAT				
Reading	6	228	.3840	.2189*
Spelling	6	228	.3122	.2081*
Math	6	228	.3257	0
PIAT				
RR	5	188	.3488	0
RC	3	99	.4760	0
Spelling	4	137	.4073	0
Math	4	137	.4522	.1500
		PPVT-R		
PIAT				
RR	5	146	.4878	0
RC	5	146	.4725	.0490
Spelling	5	146	.3759	.0632
GI	5	146	.5101	.0959
Math	5	146	.4878	.0866
Total Test	5	146	.4311	.1179

Achievement test	Number of studies	Number of students	Correlation Weighted mean	Correlation Estimated SD
Woodcock-Johnson Brief Cognitive				
WRAT				
Reading	2	60	.7700	.0671
Woodcock-Johnson Broad Cognitive				
WRAT				
Reading	2	60	.8000	.0616

Note. RR = Reading Recognition, RC = Reading Comprehension GI = General Information

*p < .05. **p < .01.

Appendix B

Subgroup Results

Subgroup	Number of studies	Number of students	Correlation Weighted mean	Correlation Estimated SD
WISC-R Full Scale IQ/PIAT General Information				
Referred	3	122	.3343	.1145
Referred (LD)	4	553	.5478	.2219**
Mixed Diagnosis	4	690	.7103	0
WISC-R Full Scale IQ/PIAT Reading Recognition				
Referred (LD)	4	431	.5572	.0372
Mixed Diagnosis	5	668	.6711	.0956**
WISC-R Full Scale IQ/PIAT Total Test				
Unidentified	2	72	.4935	0
Referred (LD)	3	251	.6516	0
Mixed Diagnosis	4	631	.7466	.0516
WISC-R Full Scale IQ/WRAT Reading				
Unidentified	10	544	.4605	.1854**
Referred	14	1337	.4689	.1218**
Referred (LD)	11	729	.3955	.0914
Mixed Diagnosis	5	199	.2300	.1028

Subgroup	Number of studies	Number of students	Correlation Weighted mean	Correlation Estimated SD
WISC-R Full Scale IQ/WRAT Math				
Unidentified	10	601	.4291	.1845**
Referred	13	1313	.5713	.0855
Referred (LD)	9	587	.4865	.0952
Mixed Diagnosis	4	178	.2972	0
WISC-R Full Scale IQ/WRAT Spelling				
Unidentified	10	601	.3616	.1812**
Referred	10	1157	.4342	.1364**
Referred (LD)	9	587	.3628	.1266
Mixed Diagnosis	4	178	.1971	.1880
WISC-R Performance IQ/WRAT Math				
Unidentified	7	451	.4076	.1265
Referred	11	1369	.4713	.0966
Referred (LD)	8	424	.3406	.1134
Mixed Diagnosis	5	279	.3848	.1571
WISC-R Performance IQ/PIAT Total Test				
Unidentified	2	72	.2865	0
Referred (LD)	3	251	.4648	.0745
Mixed Diagnosis	4	631	.5247	.1123

Subgroup	Number of studies	Number of students	Correlation Weighted mean	Correlation Estimated SD
WISC-R Performance IQ/WRAT Spelling				
Unidentified	7	451	.2873	.1258
Referred	11	1331	.2996	.1032
Referred (LD)	8	424	.2347	.1705
Mixed Diagnosis	5	279	.2283	.1609
WISC-R Performance IQ/WRAT Reading				
Unidentified	8	484	.3068	.1115
Referred	12	1393	.3042	.1439**
Referred (LD)	9	546	.2270	.1532
Mixed Diagnosis	5	279	.2544	.1585
WISC-R Performance IQ/PIAT Reading Recognition				
Referred (LD)	4	431	.4014	0
Mixed Diagnosis	5	668	.5324	.1177**
WISC-R Verbal IQ/WRAT Reading				
Unidentified	11	503	.4672	.1845**
Referred	11	1360	.5318	.1282**
Referred (LD)	9	546	.4249	.0934
Mixed Diagnosis	5	279	.4819	.1142

Subgroup	Number of studies	Number of students	Correlation Weighted mean	Correlation Estimated SD
WISC-R Verbal IQ/WRAT Math				
Unidentified (UI)	11	461	.4888	.1658
Referred	11	1369	.5631	.0765
Referred (LD)	8	424	.4742	.1321
Mixed Diagnosis	5	279	.5451	.1405
WISC-R Verbal IQ/WRAT Spelling				
Unidentified	11	503	.3900	.1661
Referred	11	1331	.4799	.1123**
Referred (LD)	8	424	.3894	.1460
Mixed Diagnosis	5	279	.4264	.0566
WISC-R Verbal IQ/PIAT Total Test				
Unidentified	2	72	.5269	0
Referred (LD)	3	251	.6957	.0502
Mixed Diagnosis	4	631	.7624	0

**$p < .01$

Appendix C

WISC-R Full Scale IQ/WRAT Math Unidentified (UI) Subgroup

Reference Citation	Descriptor Variable	Sample Size	Reported Correlation
Karnes (1984)	Gifted	90	.1900
Karnes et al. (1986)	Gifted	41	.3100
Mishra (1981)	Navajo	68	.1600
Mishra (1982)	Navajo	40	.1600
Mishra (1983)	Navajo	64	.4200
Hartlage & Steele (1977)	Unidentified	36	.7600
Hartlage & Boone (1977)	Unidentified	42	.4700
Drudge et al. (1984)	Unidentified	30	.4500
Wright & Piersel (1987)	Unidentified	160	.6300
Reilly et al. (1985)	Anglo	30	.7500

Appendix D

Identification of subgroups in the WISC-R Full Scale IQ/WRAT Math Unidentified (UI) Subgroup

Achievement test	Number of studies	Number of students	Correlation Weighted mean	Correlation Estimated SD
Unidentified	10	601	.4291	.1840**
UI-Subgroups				
Anglo/ Unidentified	5	298	.6171	.0616
Gifted/ Minority	5	303	.2441	0

$^{**}p < .01$

Appendix E

Algozzine, B., & Ysseldyke, J. E. (1981). An analysis of difference score reliabilities on three measures with a sample of low-achieving youngsters. Psychology in the Schools, 18(2), 133-138.

Allington, R. L. (1981). Sensitivity to orthographic structure in educable mentally retarded children. Contemporary Educational Psychology, 6(2), 135-139.

Appelbaum, A. S., & Tuma, J. M. (1982). The relationship of the WISC-R to academic achievement in a clinical population. Journal of Clinical Psychology, 38(2), 401-405.

Bracken, B. A. & Murray, A. M. (1984). Stability and predictive validity of the PPVT-R over an eleven month interval. Educational and Psychological Research, 4(1), 41-44.

Bracken, B. A., Prasse, D. P., & Breen, M. J. (1984). Concurrent validity of the Woodcock-Johnson Psycho-Educational Battery with regular and learning-disabled students. Journal of School Psychology, 22(2), 185-192.

Breen, M. J. (1983). WRAT-WISC R correlates for learning disabled, emotionally disturbed, educable mentally retarded, and regular education students. Educational and Psychological Research, 3(2), 103-110.

Brent, D. E., & Routh, D. K. (1978). Response cost and impulsive word recognition errors in reading-disabled children. Journal of Abnormal Child Psychology, 6(2), 211-219.

Burns, E. (1975). An evaluation of the Peabody Individual Achievement Test with primary age retarded children. Psychology in the Schools, 12(1), 11-14.

Burns, E., Peterson, D., & Bauer, L. (1974). The concurrent validity of the Peabody Individual Achievement Test. Training School Bulletin, 70(4), 221-223.

Cooley, E. J., & Ayres, R. (1985). Convergent and discriminant validity of the Mental Processing Scales of the Kaufman Assessment Battery for Children. Psychology in the Schools, 22(4), 373-377.

Covin, T. M., & Lubimiv, A. J. (1976). Concurrent Validity of the WRAT. Perceptual and Motor Skills, 43(2), 573-574.

Davis, J. M., Elfenbein, J., Schum, R., & Bentler, R. A. (1986). Effects of mild and moderate hearing impairments on language, educational, and psychosocial behavior of children. Journal of Speech and Hearing Disorders, 51(1), 53-62.

Dean, R. J. (1977). Analysis of the PIAT with Anglo and Mexican-American children. Journal of School Psychology, 15(4), 329-333.

Dean, R. S. (1980). The use of the Peabody Picture Vocabulary Test with emotionally disturbed adolescents. Journal of School Psychology, 18(2), 172-175.

Desmond, R., Singer, J. L., Singer, D. G., Calam, R., & Colimore, K. (1985). Family mediation patterns and television viewing: Young children's use and grasp of the medium. Human Communication Research, 11(4), 461-480.

Drudge, O. W., Reilly, T. P., Rosen, J. C., Fischer, M., & Loew, D. (1984). A comparison of the WISC-R, McCarthy Scales, Woodcock-Johnson and academic achievement: Concurrent and predictive validity. (ERIC Document Reproduction Service No. ED 244 996).

Eisert, D. C., & Shelburne, K. (1983). Children with Spina Bifida: Why do they fail in school? Paper presented at the annual convention of the American Psychological Association Washington, DC.

Gresham, F. M., Reschly, D. J., & Carey, M. P. (1987). Teachers as "tests": Classification accuracy and concurrent validation in the identification of learning disabled children. School Psychology Review, 16(4), 543-553.

Grossman, F. M., & Clark, J. H. (1982). The WISC-R as a predictor of WRAT achievement for educable mentally handicapped students. Educational and Psychological

Measurement, 42(2), 701-704.

Grossman, F. M., & Clark, J. H. (1984). Concurrent validity of WISC-R factor scores for educable mentally handicapped children. Perceptual and Motor Skills, 58(1), 227-230.

Grossman, F. M., & Johnson, K. M. (1982). WISC-R factor scores as predictors of WRAT performance: A multivariate analysis. Psychology in the Schools, 19(4), 465-468.

Gutterman, J. E., Ward, M., & Genshaft, J. (1985). Correlations of low vision children on the Perkins-Binet Tests of Intelligence for the blind, the WISC-R and the WRAT. Journal of Visual Impairment and Blindness, 79(2), 55-58.

Haddad, F. A. (1986). Concurrent validity of the Test of Nonverbal Intelligence with learning disabled children. Psychology in the Schools, 23(4), 361-364.

Hale, R. L. (1978). The WISC-R as a predictor of WRAT performance. Psychology in the Schools, 15(2), 172-175.

Hartlage, L. C., & Boone, K. E. (1977). Achievement test correlates of Wechsler Intelligence Scale for Children and Wechsler Intelligence Scale for Children-Revised. Perceptual and Motor Skills, 45(3), 1283-1286.

Hartlage, L. C., & Steele, C. T. (1977). WISC and WISC-R correlates of academic achievement. Psychology in the Schools, 14(1), 15-18.

Hutton, J. B., & Davenport, M. A. (1985). The WISC-R as a predictor of Woodcock-Johnson achievement cluster scores for learning-disabled students. Journal of Clinical Psychology, 41(3), 410-414.

Karnes, F. A., Edwards, R. P., & McCallum, R. S. (1986). Normative achievement assessment of gifted children: Comparing the K-ABC, WRAT, and CAT. Psychology in the Schools, 23(4), 346-352.

Karnes, F. A., Whorton, J. E., & Currie, B. B. (1984). Correlations between WISC-R IQs and Wide Range Achievement Test grade equivalents for intellectually gifted students. Psychological Reports, 54(1), 69-70.

Kaye, D. B., & Baron, M. B. (1984). The validity of the Detroit Test of Learning Aptitude. Journal of Psychoeducational Assessment, 2(2), 117-124.

Mishra, S. P. (1981). Relationship of WISC-R factor scores to academic achievement and classroom behaviors of Native American Navajos. Measurement and Evaluation in Guidance, 14(1), 26-30.

Mishra, S. P. (1983). Validity of WISC-R IQs and factor scores in predicting achievement for Mexican-American children. Psychology in the Schools, 20(4), 442-444.

Mishra, S. P., & Lord, J. (1982). Reliability and predictive validity of the WISC-R with Native-American Navajos. Journal of School Psychology, 20(2), 150-154.

Murray, A., & Bracken, B. A. (1984). Eleven month predictive validity of the Kaufman Assessment Battery for Children. Journal of Psychoeducational Assessment, 2(3), 225-232.

Naglieri, J. A. (1980). McCarthy and WISC-R correlations with WRAT achievements scores. Perceptual and Motor Skills, 51(2), 392-394.

Naglieri, J. A. (1981). Concurrent validity of the Revised Peabody Picture Vocabulary Test. Psychology in the Schools, 18(3), 286-289.

Naglieri, J. A. (1984). Concurrent and predictive validity of the Kaufman Assessment Battery for Children with a Navajo sample. Journal of School Psychology, 22(4), 373-380.

Naglieri, J. A. (1984). Predictive validity of the PPVT-R for Navajo children. Psychological Reports, 55(1), 297-298.

Naglieri, J. A. (1985). Normal children's performance on the McCarthy Scales, Kaufman Assessment Battery, and Peabody Individual Achievement Test. Journal of Psychoeducational Assessment, 3(2), 123-129.

Naglieri, J. A. (1985). Assessment of mentally retarded children with the Kaufman Assessment Battery for

Children. American Journal of Mental Deficiency, 89(4), 367-371.

Naglieri, J. A., & Haddad, F. A. (1984). Learning disabled children's performance on the Kaufman Assessment Battery for Children: A concurrent validity study. Journal of Psychoeducational Assessment, 2(1), 49-56.

Naglieri, J. A., & Pfeiffer, S. J. (1981). Correlations among scores on WISC-R and Woodcock-Johnson Achievement Tests for learning disabled children. Psychological Reports, 49(3), 913- 914.

Naglieri, J. A., & Pfeiffer, S. I. (1983). Stability, concurrent and predictive validity of the PPVT-R. Journal of Clinical Psychology, 39(6), 965-967.

Prasse, D. P., Siewert, J. C., & Breen, M. J. (1983). An analysis of performance on reading subtests from the 1978 Wide Range Achievement Test and Woodcock Reading Mastery Test with the WISC-R for learning disabled and regular education students. Journal of Learning Disabilities, 16(8), 458-461.

Raskin, L. M., Bloom, A. S., Klee, S. H., & Reese, A. (1978). The assessment of developmentally disabled children with the WISC-R, Binet and other tests. Journal of Clinical Psychology, 34(1), 111-114.

Reilly, T. P., Drudge, O. W., Rosen, J. C., Loew, D. E., & Fischer, M. (1985). Concurrent and predictive validity of the WISC-R, McCarthy Scales, W-J, and academic achievement. Psychology in the Schools, 22(4), 380-382.

Reynolds, C. R., & Gutkin, T. B. (1980, April). Predictive validity of the WISC-R for white and Mexican-American children. Paper presented at the Annual Meeting of the American Educational Research Association. Boston, MA.

Reynolds, C. R., & Hartlage, L. C. (1978). Comparison of WISC and WISC-R racial regression lines. Paper presented at the Annual Meeting of the Southeastern Psychological Association. Atlanta, Georgia

Reynolds, C. R., Wright, D., & Dappen, L. (1981). A comparison of the criterion-related validity (academic achievement) of the WPPSI and the WISC-R. Psychology in the Schools, 18(1), 20-23.

Sattler, J. M., & Ryan, J. J. (1981). Relationship between WISC-R and WRAT in children referred for learning difficulties. Psychology in the Schools, 18(3), 290-292.

Schwarting, F. G., & Schwarting, K. R. (1977). The relationship of the WISC-R and WRAT: A study based upon a selected population. Psychology in the Schools, 14(4), 431-433.

Simpson, R. G. (1982). Correlation between the General Information Subtest of the Peabody Individual Achievement Test and the Full Scale Intelligence

Quotient of the WISC-R. _Educational and Psychological Measurement_, _42_(2), 695-699.

Singer, M. G., Vance, B., & Brenner, O. C. (1983). Predicting WRAT scores from the WISC-R for a selected sample of LD children and youth. _Journal of Clinical Psychology_, _39_(5), 758-761.

Smith, T. C., & Smith, B. L. (1986). The relationship between the WISC-R and WRAT for a sample of rural referred children. _Psychology in the Schools_, _23_(3), 252-254.

Sutter, E. G., & Bishop, P. C. (1986). Further investigation of the correlations among the WISC-R, PIAT, and DAM. _Psychology in the Schools_, _23_(4), 365-367.

Thompson, P. L., & Bassard, M. R. (1984). Validity of the Woodcock-Johnson Tests of Cognitive Ability: A comparison with the WISC-R in LD and normal elementary students. _Journal of School Psychology_, _22_(2), 201-208.

Tramill, J. L., Tramill, J. K., Thornthwaite, R., & Anderson, F. (1981). Investigations into the relationships of the WRAT, the PIAT, the SORT, and the WISC-R in low-functioning referrals. _Psychology in the Schools_, _18_(2), 149-153.

Valencia, R. R. (1985). Predicting academic achievement in Mexican-American children using the Kaufman Assessment Battery for Children. _Educational and Psychological_

Research, 5(1), 11-17.

Vance, B., Fuller, G. B., & Lester, M. L. (1986). A comparison of the Minnesota Perceptual Diagnostic Test Revised and the Bender Gestalt. Journal of Learning Disabilities, 19(4), 211-214.

Walden, J. (1979). A comparison of the PIAT and WRAT: A closer look. Psychology in the Schools, 16(3), 342-346.

Webster, R. E. (1985). The criterion-related validity of psychoeducational tests for actual reading ability of learning disabled students. Psychology in the Schools, 22(2), 152-159.

White, T. H. (1979). Correlations among the WISC-R, PIAT, and DAM. Psychology in the Schools, 16(4), 497-501.

Wikoff, W. L. (1978). Correlations and factor analysis of the Peabody Individual Achievement Test and the WISC-R. Journal of Consulting and Clinical Psychology, 46(2), 322-325.

Wikoff, W. L. (1979). The WISC-R as a predictor of achievement. Psychology in the Schools, 16(3), 364-366.

Wilson, J. D., & Spangler, P. F. (1974). The Peabody Individual Achievement Test as a clinical tool. Journal of Learning Disabilities, 7(6), 384-387.

Wright, D., & Dappen, L. (1982). Factor analysis of the WISC-R and WRAT with a referral population. Journal of School Psychology, 20(4), 306-312.

Wright, D., & DeMers, S. T. (1982). Comparison of the relationship between two measures of visual-motor coordination and academic achievement. Psychology in the Schools, 19(4), 473-477.

Wright, D., & Piersel, W. C. (1987). Usefulness of a group-administered ability test for decision making. Journal of School Psychology, 25(1), 63-71.

Ysseldyke, J. E., Algozzine, B., & Shinn, M. (1981). Validity of the Woodcock-Johnson Psycho-Educational Battery for learning disabled youngsters. Learning Disability Quarterly, 4(3), 244-249.

Zagar, R., & Mead, J. D. (1983). Analysis of a short term battery for children. Journal of Clinical Psychology, 39(4), 590-597.

Zins, J. E., & Barnett, D. W. (1983). The Kaufman Assessment Battery for Children and school achievement: A validity study. Journal of Psychoeducational Assessment, 1(3), 235-241.

References

Adelman, H. S. (1979). Diagnostic classification of LD: A practical necessity and procedural problem. Learning Disability Quarterly, 2(2), 56-62.

Adelman, H. S. (1979). Diagnostic classification of LD: Research and ethical perspectives as related to practice. Learning Disability Quarterly, 2(3), 5-16.

Adelman, H. S., & Taylor, L. (1985). Summary of the survey of fundamental concerns confronting the LD field. Journal of Learning Disabilities, 19(7), 391-393.

Adelman, H. S., & Taylor, L. (1986). The problems of definition and differentiation and the need for a classification schema. Journal of Learning Disabilities, 19(9), 514-520.

Algozzine, B., Forgnone, C., Mercer, C., & Trifiletti, J. (1979). Toward defining discrepancies for specific disabilities: An analysis and alternatives. Learning Disability Quarterly, 2(4), 25-31.

Algozzine, B., & Ysseldyke, J. E. (1981). Special education services for normal children: Better safe than sorry? Exceptional Children, 48(3), 238-243.

Algozzine, B., & Ysseldyke, J. (1983). Learning disabilities as a subset of school failure: The over-sophistication of a concept. Exceptional Children, 50(3), 242-246.

Algozzine, B., & Ysseldyke, J. E. (1986). The future of the LD field: Screening and diagnosis. Journal of Learning Disabilities, 19(7), 394-398.

Algozzine, B., Ysseldyke, J. E., & Shinn, M. (1982). Identifying children with learning disabilities: When is a discrepancy severe? Journal of School Psychology, 20(4), 299-305.

Berk, R. A. (1981). What's wrong with using grade-equivalent scores to identify children? Academic Therapy, 17(2), 133-140.

Berk, R. A. (1984). An evaluation of procedures for computing an ability-achievement discrepancy score. Journal of Learning Disabilities, 17(5), 262-266.

Braden, J. P. (1987). A comparison of regression and standard score discrepancy methods for learning disabilities identification: Effects on racial representation. Journal of School Psychology, 25(1), 23-29.

Brooks, R. (1979). Psychoeducational assessment: A broader perspective. Professional Psychology, 10(5), 708-722.

Castiglione, S. S., Jr. (1981). Assessment of the learning disabled with regard to cognitive functioning and cognitive complexity. Journal of Learning Disabilities, 14(2), 74-80.

Cone, T. E., & Wilson, L. R. (1981). Quantifying a severe discrepancy: A critical analysis. Learning Disability Quarterly, 4(4), 359-371.

Cone, T. E., Wilson, L. R., Bradley, C. M., & Reese, J. H. (1985). Characteristics of LD students in Iowa: An empirical investigation. Learning Disability Quarterly, 8(3), 211-220.

Danielson, L. C., & Bauer, J. N. (1978). A formula-based classification of learning disabled children: An examination of the issues. Journal of Learning Disabilities, 11(3), 163-176.

Davis, W. A., & Shepard, L. A. (1983). Specialists' use of tests and clinical judgement in the diagnosis of learning disabilities. Learning Disability Quarterly, 6(2), 128-138.

Epps, S., Ysseldyke, J. E., & Algozzine, B. (1985). An analysis of the conceptual framework underlying definitions of learning disabilities. Journal of School Psychology, 23(2), 133-144.

Forness, S. R., Sinclair, E., & Guthrie, D. (1983). Learning disability discrepancy formulas: Their use in actual practice. Learning Disability Quarterly, 6(2), 107-114.

Frankenberger, W., & Harper, J. (1987). States' criteria and procedures for identifying learning disabled children: A comparison of 1981/82 and 1985/86 guidelines. Journal of Learning Disabilities, 20(2), 118-121.

Friedrich, D., Fuller, G. B., & Davis, D. (1984). Learning disability: Fact and fiction. Journal of Learning Disabilities, 17(4), 205-210.

Furlong, M. J., & Yanagida, E.H. (1985). Psychometric factors affecting multidisciplinary team identification of learning disabled children. Learning Disability Quarterly, 8(1), 37-44.

Hanna, G. S., Dyck, N. J., & Holen, M. C. (1979). Objective analysis of achievement-aptitude discrepancies in LD classification. Learning Disabilities, 2(4), 32-38.

Harber, J. R. (1981). Learning disability Research: How far have we progressed? Learning Disability Quarterly, 4(4), 372-381.

Huebner, E. S., & Cummings, J. A. (1986). A naturalistic study of school psychologists' use of diagnostic test data in making decisions. Journal of Psychoeducational Assessment, 4, 183-190.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). Meta-analysis: Cumulating research findings across studies. Beverly Hills: Sage Publications.

Kavale, K., & Nye, C. (1981). Identification criteria for learning disabilities: A survey of the research literature. Learning Disability Quarterly, 4(4), 383-388.

Keogh, B. K. (1986). Future of the LD field: Research and practice. Journal of Learning Disabilities, 19(8), 455-460.

Lovitt, T. C., & Jenkins, J. R. (1979). Learning disabilities research: Defining populations. Learning Disability Quarterly, 2(3), 46-50.

Mann, L., Davis, C. A., Boyer, C. W., Jr., Metz, C. M., & Wolford, B. (1983). LD or not LD, that was the question: A retrospective analysis of Child Service Demonstration Centers' compliance with the Federal definition of learning disabilities. Journal of Learning Disabilities, 16(1), 14-17.

McLeod, J. (1979). Educational underachievement: Toward a defensible psychometric definition. Journal of Learning Disabilities, 12(5), 42-50.

McLoughlin, J. A., & Netick, A. (1983). Defining learning disabilities: A new and cooperative direction. Journal of Learning Disabilities, 16(1), 21-23.

McNutt, G. (1986). The status of learning disabilities in the States: Consensus or controversy? Journal of Learning Disabilities, 19(1), 12-16.

Mercer, C. D., Hughes, C., & Mercer, A. R. (1985).

Learning disabilities used by state education departments. Learning Disability Quarterly, 8(1), 45-55.

Morrison, G. M., MacMillan, D. L., & Kavale, K. (1985). System identification of learning disabled children: Implications for research sampling. Learning Disability Quarterly, 8(1), 2-10.

Murai, H. M. (1982). Eligibility criteria for learning disabilities programs: Institutionalized discrimination. Journal of Learning Disabilities, 15(5), 267.

O'Donnell, L. E. (1980). Intra-individual discrepancy in diagnosing specific learning disabilities. Learning Disability Quarterly, 3(1), 10-18.

Olson, J. C., & Mealor, D. J. (1981). Learning disabilities identifications: Do researchers have the answers? Learning Disability Quarterly, 4(4), 389-392.

O'Reilly, C. S., Northcraft, G. B., & Sabers, D. L. (1989). Confirmation bias in special education placement decisions. School Psychology Review, 18(1), 126-135.

Perlmutter, B. F., & Parus, M. V. (1983). Identifying children with learning disabilities: A comparison of diagnostic procedures across school districts. Learning Disability Quarterly, 6(3), 321-328.

Reger, R. (1979). Learning disabilities: Futile attempts at a simplistic definition. Journal of Learning Disabilities, 12(8), 529-532.

Reynolds, C. R. (1981). The fallacy of "two years below grade level for age" as a diagnostic criterion for reading disorders. Journal of School Psychology, 19(4), 350-358.

Reynolds, C. R. (1984-1985). Critical measurement issues in learning disabilities. Journal of Special Education, 18(4), 451-476.

Reynolds, C. R. (1985). Measuring the aptitude-achievement discrepancy in learning disability diagnosis. Remedial and Special Education, 6(5), 37-48.

Reynolds, C. R., & Hartlage, L. (1979). Comparison of WISC and WISC-R regression lines for academic prediction with black and with white referred children. Journal of Consulting and Clinical Psychology, 47(3), 589-591.

Rotatori, A. F., & Mauser, A. J. (1980). IEP assessment for LD children. Academic Therapy, 16(2), 141-153.

Schmidt, L. (1988). Validity generalization and the future of criterion-related validity. In H. Wainer, & H. I. Braun (Eds.), Test Validity (pp. 173-190). Hillsdale, NJ: Lawrence Erlbaum Associates.

Schulte, A., & Borich, G. D. (1984). Considerations in the use of difference scores to identify learning-disabled children. Journal of School Psychology, 22, 381-390.

Singer, M. C., Vance, B., & Brenner, O.C. (1983). Predicting WRAT scores from the WISC-R for a selected sample of LD children and youth. Journal of Clinical Psychology, 39(5), 758-761.

Shepard, L. A. (1980). An evaluation of the regression discrepancy method for identifying children with learning disabilities. Journal of Special Education, 14(1), 79-90.

Shepard, L. A. (1983). The role of measurement in educational policy: Lessons from the identification of learning disabilities. Educational Measurement: Issues and Practice, 2(3), 4-8.

Shepard, L. A., & Smith, M. L. (1983). An evaluation of the identification of learning disabled students in Colorado. Learning Disability Quarterly, 6(2), 115-127.

Shepard, L. A., Smith, M. L., & Vojir, C. P. (1983). Characteristics of pupils identified as learning disabled. American Educational Research Journal, 20(3), 309-331.

Sigman, S. B. (1985). A radical perspective on the development of American special education with a focus on the concept of "learning disabilities". Dissertation Abstracts International, 47, 503A.

Smith, J. D., & Polloway, E. A. (1979). Learning disabilities: Individual needs or categorical concerns? Journal of Learning Disabilities, 12(8), 525-528.

Thibodeau, J. J. (1985). Beyond the test scores. Academic Therapy, 20(3), 307-315.

Thurlow, M. L., & Ysseldyke, J. E. (1979). Current assessment and decision-making practices in model LD programs. Learning Disability Quarterly, 2(1), 15-24.

Torgesen, J. K. (1986). Learning disabilities theory: Its current state and future prospects. Journal of Learning Disabilities, 19(7), 399-407.

Tucker, J., Stevens, L. J., & Ysseldyke, J. E. (1983). Learning disabilities: The experts speak out. Journal of Learning Disabilities, 16(1), 6-14.

United States Office of Education (1977). Rules and regulations implementing Public Law 94-142 (The education for All Handicapped Children Act of 1975). Federal Register, 42.

Valdez, R. S., & Valdez, C. (1983). Detecting predictive bias: The WISC-R vs. achievement scores of Mexican-American and non-minority students. Learning Disability Quarterly, 6(4), 440-447.

Wilson, L. R. (1985). Large-scale learning disability identification: The reprieve of a concept. Exceptional Children, 52(1), 44-51.

Wilson, L. R., & Cone, T. (1984). The regression equation method of determining academic discrepancy. Journal of School Psychology, 22(1), 95-110.

Wilson, L. R., Cone, T. E., 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(9), 553-557.

Wilson, L. R., Cone, T. E., Busch, R., & Allee, T. (1983). A critique of the expectancy formula approach:Beating a dead horse? Psychology in the Schools, 20(2), 241-249.

Wong, B. Y. L. (1986). Problems and issues in the definition of learning disabilities. In J. K. Torgesen, & B. Y. L. Wong (Eds.), Psychological and educational perspectives on learning disabilities. (pp. 3-22). Orlando: Academic Press, Inc.

Ysseldyke, J. E., & Algozzine, B. (1979). Perspectives on assessment of learning disabled students. Learning Disability Quarterly, 2(4), 3-12.

Ysseldyke, J. E., & Algozzine, B. (1983). LD or not LD: That's not the question. Journal of Learning Disabilities, 16(1), 29-31.

Ysseldyke, J. E., Algozzine, B., Regan, R., & McGue, M. (1981). The influence of test scores and naturally occurring pupil characteristics on psychoeducational decision making with children. Journal of School Psychology, 19(2), 167-177.

Ysseldyke, J. E., Algozzine, B., Richey, L., & Graden, J. (1982). Declaring students eligible for learning disability services: Why bother with the data? Learning Disability Quarterly, 5(1), 37-44.