A COMPARATIVE ANALYSIS OF STANDARDIZED ACHIEVEMENT TESTS WITH LEARNING DISABLED AND NON-LEARNING DISABLED ADOLESCENT BOYS

B. Claire McCullough
Computer Services Division
and Department of Sociology
University of South Carolina
Columbia, South Carolina

and

Barbara A. Zaremba
Learning Disabilities/Juvenile Delinquency Project
The National Center for State Courts
Williamsburg, Virginia

October, 1979

This report has been prepared with support from Grants 78-JN-AX-0038 and 78-JN-AX-0022 from the National Institute for Juvenile Justice and Delinquency Prevention, U. S. Department of Justice. Points of view or opinions stated herein are those of the authors and do not necessarily represent the official positions or policies of the U.S. Department of Justice.

B. Claire McCullough was a Staff Associate at the National Center for State Courts when this research was conducted.

A Comparative Analysis of Standardized Achievement Tests with Learning Disabled and Non-Learning Disabled Adolescent Boys

Abstract

The characteristics of the KeyMath Diagnostic Arithmetic Test and the Woodcock Reading Mastery Test, when used with a sample of learning disabled boys, are compared to those when used with a sample of non-learning disabled boys. The Woodcock is found to be quite similar, with high reliability and a unidimensional factor structure, for both the learning disabled and non-learning disabled samples. The KeyMath is also quite similar for the two samples. It has very high reliability, although the assumption of a unidimensional factor structure of math achievement is less defensible for the learning disabled sample.

The use of achievement tests is popular for the purpose of making screening decisions, placement decisions (Johnson and Myklebust, 1967; Dechant, 1968; Wallace and McLoughlin, 1975), and evaluation decisions (Tinney, 1975; Keilitz and Saks, 1979) for learning-disabled students. The characteristics of achievement tests have not been thoroughly examined among "normal" populations, let alone among learning disabled children.

This paper examines the characteristics of two achievement tests, the KeyMath Diagnostic Arithmetic Test (Connolly, Nachtman and Pritchett, 1976) and the Woodcock Reading Mastery Tests (Woodcock, 1973), comparing their characteristics among samples of learning disabled and non-learning disabled adolescents. The implications of these comparisons for the use of the tests in research on, and the screening and diagnosis of, learning disabilities are discussed.

Achievement Tests in LD Screening and Assessment

When achievement test scores are used to screen children for possible learning disabilities, those who are suspected of being learning disabled, typically are those whose test scores are below what is expected based on their tested I.Q. Conventionally, scores of verbal and performance I.Q. are compared to scores in reading and math achievement. The diagnosis of particular types of LD is often based to some extent on scores on particular parts of an achievement test.

Achievement Tests in LD Research

There are two ways in which achievement tests are most often used in LD research. One is to compare achievement test scores of learning disabled children to those of non-learning disabled children. The second is to compare two or more groups of learning disabled children who differ in some other way. Typically the difference has to do with their being in (or not being in) one or another kind of educational program for learning disabled children. The achievement test scores of the groups are compared to determine the relative effectiveness of the educational program(s).

Implication of Assumptions about Achievement Tests

Each of the above applications of achievement test scores with learning disabled children is based generally on several assumptions about the tests and about the abilities of learning disabled and non-learning disabled children. The tests are assumed to be equally reliable for learning disabled and non-learning disabled children. The implications of unreliable measurement are well established. If standardized achievement tests are not reliable for use with learning disabled children, any screening or assessment based on them is suspect. Any comparison of learning disabled and non-learning disabled children which does not consider the possible differential reliability of the tests is questionable.

The tests also are assumed to be measures with comparable validity, of characteristics which both learning disabled and non-learning disabled children are presumed to have. This involves two assumptions: first, that the learning disabled and non-learning disabled children can be characterized by general mathematics and verbal achievement; and, second, that these general characteristics can be measured with equal validity by the same standardized tests. It is assumed that both math and verbal achievement can be measured as unidimensional characteristics. This seems quite feasible for "normal" children but may be less so when considering learning disabled children.

The implications of this assumption being incorrect can be great. The assumption that math and verbal ability are unidimensional characteristics leads to the practice of using some overall score in each of the two areas. This practice may cover up differences between learning disabled and non-learning disabled children which would be very useful in learning disabilities screening and very important in learning disabilities research.

Consider, for example, sets of hypothetical scores for two children on an achievement test which has six

UNITED STATES DEPARTMENT OF JUSTICE
OFFICE OF JUVENILE JUSTICE AND DELINQUENCY PREVENTION
WASHINGTON, D.C. 20531

ASAP.
Thanks
Emily

October 29, 1979

Ms. Emily Johnson
National Criminal Justice
Reference Service
1015 20th Street, N.W. 4th Flr
Washington, D.C. 20036

Dear Emily:

Enclosed are the latest reports from the Learning Disabilities - Juvenile Delinquency R&D Project entitled "A Comparative Analysis of Standardized Achievement Tests with Learning Disabled and Non-learning Disabled Adolescent Boys" and "The Evaluation of the Learning Disabilities/Juvenile Delinquency Remediation Program: Evaluation Design and Interim Results."

Please enter these into the NCJRS data base and advise me when access numbers have been assigned. I am having 50 copies made of the Comparative Analysis and 150 of the Evaluation report for distribution.

Thank you.

Sincerely,

Pam

Pamela Swain
National Institute for Juvenile Justice
and Delinquency Prevention/OJJDP

cc: Paul Broder
National Center for State Courts

subtests. One child scores in the 75th percentile on all six subtests, the other scores in the 50th percentile on the first three subtests and in the 100th percentile on the other three. If the area of achievement is considered to be unidimensional and an overall score (say, the mean percentile) is used, the two children get the same score --the 75th percentile. If, on the other hand, it is assumed that this achievement area has two distinct subareas, represented by the first three and last three subtests, we come to different conclusions. The first child is at the 75th percentile in both subareas. The second child is at the 50th percentile in one area and the 100th percentile in the other. This characteristic (which could be typical of learning disabled children) would be lost under the unidimensional assumption.

This report examines data related to the assumptions about the uniform reliability and validity of achievement test scores for use with learning disabled and non-learning disabled children.

Data Available for Examination

The data on which our analyses are carried out come from a large-scale study of the relationship between learning disabilities and juvenile delinquency that is being conducted as a joint effort of the National Center for State Courts and the Association for Children with Learning Disabilities. A more general report of the study's findings to date can be found in McCullough, Zaremba and Rich (1979). The study is being conducted in three cities: Baltimore, Maryland; Indianapolis, Indiana; and Phoenix, Arizona. Data have been collected from 1699 youngsters, 1006 of whom were attending public schools and who had no official record of delinquency, and 693 of whom had been officially adjudicated delinquent and were either in institutions for delinquent youths, in the public schools after having been placed on probation or parole, or were school drop-outs. The youngsters participating in the study were twelve- to seventeen year-old boys and girls and their average age was 14.2 years. The sample includes youths from all socioeconomic levels, and a number of racial categories. Because the percentage of females in the sample was quite small, the analysis reported here is based on males only.

The Learning Disabilities Classification

Educational Testing Service was retained to develop and carry out the learning disabilities diagnostic procedures. As a first step in the classification of the participants into learning disabled and non-learning disabled groups, the school records of all the youngsters sampled were reviewed for evidence of possible learning disabilities. A child was suspected of being learning disabled if he had been so classified by the school or if there was evidence of irregular academic achievement in his school record. At this point in the diagnostic process, some youths were eliminated from the study for reasons which included (1) inappropriate age, (2) mental retardation, (3) severe emotional disturbance, (4) English as a second language, and (5) physical handicaps. Those who were not eliminated and for whom a learning disabilities classification could not be precluded were referred for further diagnostic evaluation. They were given a battery of tests by ETS diagnosticians which included the WISC-R, the Bender-Gestalt (Koppitz scoring), the KeyMath Diagnostic Arithmetic Test and the Woodcock Reading Mastery Test. Decisions as to the presence of learning disabilities were made by considering discrepancies between ability and achievement test scores, which were supplemented by the score on the Bender and observational notes taken during the testing (Campbell, 1978). Initially, these decisions were made by ETS diagnosticians, but a reliability check showed that the clinical decision-making process did not produce acceptably reliable results. To insure a more uniform application of the learning disabilities classification procedures, all elements to be considered were entered into a computerized algorithm which is described in detail by Campbell (1978). Using this process, 23% of the total sample was determined to be learning disabled.

The Achievement Tests

The KeyMath is a diagnostic test consisting of 14 subtests that are grouped into three areas: Content, which includes the three subtests of Numeration, Fractions, and Geometry and Symbols; Operations, which includes the six subtests of Addition, Subtraction, Multiplication, Division, Mental Computation and Numerical Reasoning; and Applications, which includes the five subtests of Word Problems, Missing Elements, Money, Measurement and Time. The test was developed on a sample of 1222 students in kindergarten through seventh grade and is individually administered.

The Woodcock Reading Mastery Test is an individually administered test consisting of five subtests: Letter Identification, Word Identification, Word Attack, Word Comprehension, and Passage Comprehension. The test was developed on two samples. The first consisted of approximately 1000 kindergarten through seventh grade students, the second included 4000 kindergarten through twelfth grade students. A total of 50 school districts across 20 states contributed to the norming samples.

Both tests are standardized tests of achievement with evidence of their reliability in the samples on which they were developed. Neither test has been examined thoroughly for use with other types of samples. The following analysis addresses their appropriateness for use with learning disabled adolescent boys.

Procedures of the Present Analysis

Since not all of the youngsters in the sample were referred for testing, only a subset is available for this analysis. Restricting the analysis to boys for whom complete achievement tests were available results in 384 learning disabled and 603 non-learning disabled boys for comparison.

The reliability of the achievement tests was compared by computing coefficient alpha (standardized) separately for the learning disabled and non-learning disabled boys. For this analysis, the subtest scores are considered the items and the reliability is calculated for the whole test.

To test the validity of using these achievement test scores as measures of general achievement in math and reading, a factor analysis of the subtests in each test was done separately for learning disabled youngsters and the results compared. If the unidimensionality assumption holds, there should be only one reliable factor among learning disabled, as well as non-learning disabled, youths.

Results

The reliability of the tests is quite comparable for the two groups. For the learning disabled group, the reliabilities were .88 and .94 respectively for the

Woodcock and KeyMath. The reliabilities were slightly higher for the non-learning disabled group; .92 and .97 respectively. These results are comparable to the published reliabilities for the samples on which the tests were developed. Woodcock (1973) reports a reliability of .83 for a sample in grade 7.9, and Connolly, Nachtman and Pritchett (1976) report a reliability for the KeyMath of .96 for a seventh grade sample. These results indicate that, as far as reliability is concerned, these two achievement tests are quite comparable for learning disabled and non-learning disabled adolescent boys.

The results concerning validity are less straightforward. The factor analysis of the fourteen KeyMath subtests resulted in one factor for the non-learning disabled sample, but two for the learning disabled boys. The loadings (based on varimax rotation) of the fourteen subtests for this group are presented in Table 1.

(INSERT TABLE 1)

These results indicate that there are two relatively distinct dimensions of math achievement for learning disabled boys. The subtests which most heavily define the first are Numeration, Fractions, Geometry and Symbols, Word Problems, Missing Elements, Money, Measurement and Time. The second dimension of math achievement for learning disabled boys is defined most heavily by the subtest testing Addition, Subtraction, Multiplication and Division. The Mental Computation and Numerical Reasoning subtests load almost equally on the two dimensions. The second dimension defines fairly clearly the operations area of the test as defined by the test developers. The first dimension is defined by the content and applications areas of the test. What these results indicate is that learning disabled boys can be distinguished from non-learning disabled boys by their patterns of scores on the KeyMath subtests. While non-learning disabled boys tend to score equally well or equally poorly on all subtests of the KeyMath, the scores of learning disabled boys tend to cluster into two areas, operations versus content and applications.

While these results seem to call into question the use of an overall KeyMath score for learning disabled boys, other information from the factor analysis sheds further light on the issue. The determination of whether or not a factor is a reliable one, which was used here, is the conventional one based on the eigenvalue of the factor. If a factor has an eigenvalue greater than or equal to

1.0, it is considered reliable. The second factor for the learning-disabled sample had an eigenvalue of 1.15, only slightly greater than 1.0. The first factor explained the majority of the common variance in the test (57%), while the second factor explained only an additional 8.2%. These two pieces of information, coupled with the rather high reliability of the whole set for the learning disabled sample (.88), indicate that although there are two dimensions of math achievement for this sample, the test can be used quite reliably as a whole.

With this idea in mind, the factor analysis was repeated for the learning disabled boys, this time "forcing" one factor. The loadings from this analysis are compared to the loadings for the non-learning disabled sample in Table 2.

(INSERT TABLE 2)

A comparison of the factor loadings in Table 3 confirms the finding that the assumption of a unidimensional math achievement is more reasonable among non-learning disabled than among learning disabled boys. This is indicated by the fact that the factor loadings are all uniformly high (the lowest is .68) for the nonlearning disabled sample, and in every case higher than the loadings for the learning disabled sample. This is also confirmed by the fact that more of the variance on the test is explained by this factor for the non-learning disabled sample (70.1 vs. 57.0).

An examination of the loadings in Table 3 for the learning disabled sample, however, indicates that they are all uniformly high as well. The pattern of loadings is also quite similar in the two groups, with the higher loadings tending to be the same in the two groups. The results seem to suggest that while the assumption that math achievement is unidimensional is not as defensible among learning disabled boys as among non-learning disabled boys, it is a reasonable assumption in both groups.

(INSERT TABLE 3)

The results of the factor analysis of the Woodcock subtests are presented in Table 3. In this case, there is only one factor for both groups. The factor loadings are uniformly high except for the first subtest,[1] and the ranking of the loadings is the same whether considering learning disabled or non-learning disabled boys.

Implications

The implications of this analysis are that both the KeyMath and the Woodcock can be used with confidence for the screening and diagnosis of and for research on learning disabled children. Dividing the KeyMath subtests into two areas may also be useful in distinguishing learning disabled from non-learning disabled children. If, as Coles (1978) stresses, testing is the core of learning disabilities diagnosis, more attention should be paid to the reliability and validity of the tests used in learning disabilities screening and diagnosis. If competent research in the learning disabilities area is to be done, the same is true.

REFERENCES

Buros, Oscar K. (editor). The Eighth Mental Measurements Year Book. New York, The Gryphon Press, 1978.

Campbell, Paul B. "The definition and prevalence of learning disabilities." Paper presented at the 1978 ACLD Annual Conference, Kansas City, Missouri.

Coles, Gerald S. "The learning disabilities test battery: empirical and social issues." Harvard Educational Review, 1978, 48, 313-340.

Connolly, Austin J., Nachtman, William, & Pritchett, E. Milo. KeyMath Diagnostic Arithmetic Test Manual. Circle Pines, Minnesota: American Guidance Service, 1976.

Cronbach, L. J. "Coefficient alpha and the internal structure of tests." Psychometrika, 1951, 16, 297-334.

Dechant, Emerald. Diagnosis and Remediation of Reading Disability. West Nyack, N.Y.: Parker Publishing Co., 1968.

Johnson, Doris J. & Myklebust, Helmer R. Learning Disabilities: Educational Principles and Practices. New York: Grune and Stratton, 1967.

Keilitz, Ingo & Saks, Michael J. The Evaluation of the Learning Disabilities/Juvenile Delinquency Remediation Program: Evaluation Design and Interim Results. Unpublished report. Williamsburg, VA: National Center for State Courts, 1979.

McCullough, B. Claire, Zaremba, Barbara A. & Rich, William D. "The role of the juvenile justice system in the link between learning disabilities and delinquency." State Court Journal, 1979, 3, 24-26, 44-47.

Tinney, F. A. "A comparison of the KeyMath Diagnostic Arithmetic Test and the California Arithmetic Test used with learning disabled students." Journal of Learning Disabilities, 1975, 8, 57-59.

Wallace, Gerald and McLoughlin, James A. Learning Disabilities: Concepts and Characteristics. Columbus, Ohio: Charles E. Merrill Publ., 1975.

Woodcock, Richard W. Woodcock Reading Mastery Tests Manual. Circle Pines, Minnesota: American Guidance Service, 1973.

FOOTNOTE

1. The Woodcock test manual reports a very low reliability (.16) for this subtest using a sample at grade level 7.9 Buros (1978, 1306) reports that "...by the end of fourth grade most pupils are expected to achieve a perfect score." The low factor loading with our samples is consistent with the above.

Table 1

Factor Loadings for KeyMath Subtests

With the Learning Disabled Sample

Subtest	Factor 1	Factor 2
Numeration	.66	.48
Fractions	.66	.41
Geometry & Symbols	.57	.39
Addition	.30	.68
Subtraction	.32	.71
Multiplication	.30	.76
Division	.36	.77
Mental Computation	.53	.55
Numerical Reasoning	.54	.50
Word Problems	.74	.35
Missing Elements	.68	.19
Money	.62	.36
Measurement	.74	.32
Time	.69	.32

Table 2

Factor Loadings for KeyMath Subtests With Learning Disabled and Non-Learning Disabled Samples (One Factor Solution)

Subtest	Learning Disabled	Non-Learning Disabled
Numeration	.82	.89
Fractions	.77	.84
Geometry & Symbols	.69	.78
Addition	.69	.79
Subtraction	.69	.81
Multiplication	.71	.84
Division	.77	.83
Mental Computation	.77	.86
Numerical Reasoning	.74	.84
Word Problems	.78	.85
Missing Elements	.63	.68
Money	.70	.78
Measurement	.76	.87
Time	.72	.86

Table 3

Factor Loadings for Woodcock Subtests With

Learning Disabled and Non-Learning Disabled Samples

Subtest	Learning Disabled	Non-learning Disabled
Letter Identification	.45	.54
Word Identification	.97	.97
Word Attack	.79	.87
Word Comprehension	.76	.87
Passage Comprehnsion	.92	.95

END