The Definition and Prevalence of Learning Disabilities

Paul B. Campbell

March 3, 1978

Presented at the

1978 ACLD Annual Conference
Kansas City, Missouri

Prepared under Grants 76NI-99-0133 and 76-99-0022 from the National Institute of Juvenile Justice and Delinquency Prevention, Law Enforcement Assistance Administration, U. S. Department of Justice.

The Definition and Prevalence of Learning Disabilities

Purpose

This study was designed and implemented to provide further informacion on the question of whether or not there is a link between adjudicated delinquency and learning disabilities (LD). Indicators of differential prevalence of LD in adjudicated and non-delinquent populations are expected if a link, in fact, exists.

The study builds upon the review of research on this issue conducted by Murray and his associates (1976). Data were also collected to examine the relation of self reported delinquency to LD. These data were subjected to separate analysis not reported herein.

Sample

The populations of interest for this study are those comprised of 12- to 15-year old boys who attend public school and have no record of adjudicated delinquency and boys in the same age cohort who do not have a record of adjudicated delinquency. The sample was selected in the non-delinquent, public school case to represent a highly heterogeneous group of boys in three cities -- Baltimore, Indianapolis, and Phoenix. In the adjudicated delinquent case, the sample represented all available boys who had been adjudicated and were in the juvenile justice system, and who were accessible during the time of the study.

Procedures

The process of data collection and decision-making about the presence or absence of LD occured in two stages. A detailed description of the process may be found in a recent report (Barrows, Campbell, Slaughter and Trainor, 1977) but a general description may help to illustrate the two basic principles. The first

principle is definitional. LD is characterized by pronounced intrapersonal differences in ability to perform a variety of verbal, quantitative, and maniplua-tive tasks because there is some non-obvious interference with the process of receiving information, of utilizing it in cognitive processes, or of communicating the results of cognition. The second principle is an operational one. At any point in the process where data are found to be insufficient, the decision will always require movement to an intensive data collection step.

These principles were applied through a review of records and a process of diagnostic assessment. In records review, a trained reviewer examined the record of each youth in the study for evidence of pronounced intrapersonal performance discrepancies in school subjects or test scores and for recorded clinical judgment of LD, mental retardation, or emotional distrubance. Additionally, any supporting data of anecdotal or clinical observation was noted. Inappropriate age, physical handicap, and the youth's major language were also noted. The rules specified that the youth be excluded from the sample if physical handicaps, native language differences, mental retardation or emotional disturbance were the likely causes of disparate performance. Exclusions for mental retardation required more rigorous rules than those usually applied, and emotional disturbance required a second review before exclusion from the sample. If no evidence of discrepancy of performance or of clinical or anecdotal suggestions of LD were present, the youth was assigned to the non-LD category and was subsequently contacted for an interview. The interview was designed to provide an adequate description of the sample and was not used to determine LD. All other youth were referred for diagnostic assessment, including those for whom there were inadequate records.

The diagnostic assessment process included a battery of tests and an observa-tion schedule. Including breaks and the interview, following the same schedule used

for youth who were interviewed only, the assessment time averaged about three and one half hours of individual work with each youth. When administration of the assessment battery was completed, the data were recorded on a Basic Data Form and prepared for subsequent analysis.

The Assessment Battery

The assessment battery included three major parts: tests for determining the presence or absence of learning disability, an observation schedule to assist with the determination, (see Table I) and a set of marker tests designed to be used in a subsequent analysis for construct validation. The tests used in the LD decision were as follows:

WISC-R
Woodcock Reading Mastery
Key Math
Bender Gestalt (Koppitz scoring)

The markers, although not used in the LD determination, included tests such as the Swinton-Wepman Visual Memory Test, the Rosner Auditory Analysis Test, the Thurstone Flags, and others.

Quality Control

Maintenance of quality in a complex undertaking such as this project required a variety of procedures suited to the many aspects of the project. The quality control checks which were utilized will be discussed sequentially in the remainder of this section.

Accuracy of Testing

To achieve this aspect of quality, all testing procedures used were treated as standardized, that is, there were uniform procedures for the administration of

the tests to each youth. Originally there were two quality control processes planned to maximize the uniformity of test administration. The first of these, applied after the initial training period, was a systematic observation of the testing process by the diagnostic supervisor in charge of each site. Originally this was planned to occur in every eighth test administration. Because of scheduling problems we were not able to maintain anticipated numbers of administrations per day and it became necessary to schedule the observations by the day rather than by the number of test administrations. We were, however, able to maintain a schedule of at least partial observation on a daily basis, thereby assuring that each diagnostic assessor would be observed somewhat more frequently than once every eight days. Depending upon whether public school or institutionalized students were being tested or whether parolees or probationers were the test candidates, the observations of test administration stayed within one or two administrations of the desired schedule. The change was in the direction of more administrations being observed for the latter group, because there were actually fewer administrations in the eight day time span.

The second form of quality control for uniformity of test administration involved the use of third party evaluators. Depending upon the site, three people participated in this activity. They were Mr. Richard Harsh, Ms. Nadine Lambert, or the Project Director, Paul Campbell. Harsh and Lambert observed the Phoenix site, Lambert and Campbell observed the Indianapolis site, and Lambert observed the Baltimore site.

The outside evaluators sat in on a test administration in its entirety, making notes on the accuracy with which the test procedures were followed. After this observation was completed, the site evaluators met with the diagnostic assessors being observed to review the findings of the observation. They then reviewed these findings with the diagnostic supervisor of each site. The supervisor

in turn shared the observation, particularly as it referred to any deviations from testing procedures, with all the diagnostic assessors at the site. This was done in a training session which was scheduled immediately following the quality control checks.

A third quality control was not included in the original proposal. The original plan called for pretraining before testing began, and refresher training midway through the testing. The two previously described observational opportunities comprised the remainder of the plan for maintenance of quality. However, it became evident within the first two weeks of testing that the diagnostic staff was assembling weekly for making the coded LD/nonLD decisions. This provided an opportunity for a weekly refresher in which the diagnostic supervisor and the diagnostic assessors were able to compare notes and deal with problems in testing procedure which may have occurred during the preceeding test administrations. This activity occurred at all of the sites. The emerging pattern of quality control can thus be seen to be a continuing function of the diagnostic supervisor, augmented by outside evaluators, and regularly supported by the observations and experiences of the diagnostic assessors themselves as the testing proceeded.

The second concern about the accuracy of testing related to the precision with which the results of testing could be recorded for subsequent data processing. The demands of the study required that immediate decisions be made about the LD/nonLD status of the adjudicated sample. The tests required to meet the estimate of prevalence, as well as the diagnostic function for the remediation sample, were necessarily complex. Each separate operation required to record results presented an opportunity for a clerical error. There were many such operations in the selected test battery. Initially it was expected that the training and oversight would provide an adequate level of clerical accuracy. However, a review of several test protocols about midway through the testing

demonstrated that there were frequent errors. An initial check by ACLD staff in the field and subsequently by a statistical aide and other ETS project staff members demonstrated that, although errors changing the LD/nonLD decision seemed to be quite rare, minor errors were very frequent. As expected, the greatest numbers of errors occurred with the most complex recording requirements, such as those of the Woodcock Reading Mastery Test. It was therefore decided to record only the raw scores. These were keypunched and computer algorithms were prepared to calculate the derived scores, thereby removing all of the complexity except the counting of responses and the recording of single numbers of no more than three digits. The principle followed in this process is that of reducing to an absolute minimum the number of opportunities to change the value of scores by erroneous transciption, reference to the wrong table or computational error. To facilitate this correction, the previously prepared data tapes were used to produce labels containing the student's name, identification number, and location code. The computer calculated converted scores which were included on the tape at the end of each student's record but which did not replace the previously hand computed scores. It is therefore possible to estimate the magnitude of clerical error while at the same time removing the consequences of this error upon frequency distributions, LD/nonLD decisions or other statistical processing to which the data may be subjected.

The Decision Process

The second major quality control consideration deals with the accuracy of the decision process. There were two points of decision making, one for screening into or out of diagnostic assessment and one for the LD/nonLD decision after diagnostic assessment. The decision to screen into diagnostic assessment was

designed to "fail safe" by referring doubtful cases to diagnostic assessment rather than screening them out. The diagnostic supervisor was specifically instructed to review questionable cases, particularly if the reason for possible screenout was emotional disturbance. It was subsequently decided to include for diagnostic assessment the cases for which there were insufficient data to make a reliable assessment decision. Consequently, most cases with insufficient data were diagnostically assessed if the testing candidates themselves could be located. The LD/nonLD decision presented a different sort of problem. The original design called for two blind decisions by diagnostic assessors based on the perception of the diagnostic assessor who had administered the battery and then on the perception of a second diagnostic assessor who reviewed the information recorded on the basic data form. The second diagnostic assessor was not aware of the first assessor's decision because the decisions were coded. These were then reviewed by the supervisor who broke ties or, if he/she felt that the information was improperly evaluated, entered his/her own decision and justified reversing that of the two previous judgments. In order to check the consistency of this process, selected samples of cases were exchanged among site supervisors, again in coded form, so that the decisions were unknown when the cases were checked by the reviewing diagnostic supervisors. The results of the initial check showed enough inconsistency that a second check designed in the same manner was performed. The outcome of these two cross-site checks is presented in Table II. Of the 42 cases reviewed, 13 were challenged by one judge, and 9 by two judges. This result, although not unusual in comparison with the reported findings in the literature on clinical judgments, was not satisfactory for the purposes of this study. It was therefore decided to prepare

a decision algorithm for uniform application by computer. A review of the LD/nonLD decision process is in order here. On page 26 of the Research Procedures we find the following description (Barrows, et al, 1977).

> The first consideration in the LD/Not LD judgment will be a review of the profiles for discrepancies at least equivalent to the difference in group means of groups two years apart. These differences may occur within the score patterns of the WISC-R, including the Witkin factor scores, between the WISC scores and any or all of the achievement scores or between the achievement scores. When discrepancies of the indicated magnitude exist among all three sources of score data, the decision is clearly LD. When two sources show discrepancies, the same decision will apply if there is any supporting evidence from the Bender or the observations. When only one pair of scores show discrepancy, supportive evidence from the Bender and two or more pronounced characteristics from the behavioral observations will indicate LD. Cases which show no significant discrepancies as defined, but demonstrate two years below level achievement, and include observations of difficulty in following oral directions, motor difficulty, paper rotation, productive language problems, distractability, and at least one of the WISC observations will also be judged LD if the full scale WISC-R score is at least 33. Cases which do not meet any of these criteria will be judged non-LD.

It should be noted that these decision rules include both discrepancies among test scores (ability and achievement), evidence from other test sources such as the Bender and consideration in a clinical sense of ordinally characterized observations. The rules are not completely explicit because some latitude for judgment was intended. The great variability demonstrated by the cross-site protocol checks however, argued for more precise explication. Accordingly, the role of clinical judgment was limited to the initial categorizing of the observations applying to the WISC-R itself and to the general behavior of the youth in the testing situation. Specific requirements for the decision were defined to sequentially consider each element of data in algorithmic form.

This algorithm was organized around the following statements considered in order of presentation:

1. If the difference between verbal and performance scores on the WISC-R is 10 points or greater, count one point toward the LD/nonLD decision.

2. If a difference of 10 points or greater is present among the Witkin factors, count one point toward the decision. Count only one difference among the Witkin scores.

3. If a 10 point difference is present between the conventional scoring and the Witkin scoring, count one point toward the decision. Count only one difference between these score sets.

4. If two differences or discrepancies are present involving a single score, eliminate the most extreme score that would account for another discrepancy when compared to the two achievement scores. If after this restriction is satisfied a 10 point discrepancy is present between the Reading score and any ability score or between the Arithmetic score and any ability score, count one point for each toward the decision. A maximum of two discrepancies may be counted by these ability/achievement comparisons, one for each achievement area.

5. If a 10 point difference or discrepancy is present between the Arithmetic and the Reading scores, count one toward the decision.

6. If three discrepancies are accumulated by these comparisons, classify the case as LD.

7. If two discrepancies are present among the six comparison sets, and if any one of the following conditions is also present, classify the case as LD.

   a. A Bender score of three or more.
   b. Pronounced characteristics (a score of 1) on the WISC-R observations on two or more cases.
   c. Three or more occasionally observed characteristics (a score of 2) in the WISC-R observations.
   d. Three or more (a score of 1) pronounced characteristics in the behavioral observations.

8 If only one profile discrepancy is observed, a score of three or more on the Bender, and two or more pronounced behavioral characteristics will classify as LD.

9. If no discrepancies are present but achievement t-scores of 40 or less and occasional or pronounced characteristics are present in behavioral observations of difficulty in following oral direction, of gross or fine motor difficulty, of difficulty in oral expression, of distractability, and in at least one observation in the WISC performance observations, and the WISC Full Scale score is at least 33, classify as LD.

10. All other cases are classified as non-LD.

This decision algorithm rests on the hypothesis that pronounced intrapersonal differences in ability to perform a variety of verbal, quantitative, and manipulative tasks is associated with LD. Because the stability of a single score comparison may be questioned, the existence of differences must be verified either by occurrence among several score comparisons or by the increasing prevalence of signs from the other measures and/or observations.

The Final Decision Algorithm

Concurrently with the examination of the effects of applying a uniform Decision rule, it became possible to examine the characteristics of the data themselves. Of particular interest were the intercorrelations between the various parts of the assessment battery. Because differences between the performance of a youth on one task is to be compared with his performance on another, the reliability of these differences is at issue. Table III shows the correlations of the Woodcock, Key Math, and WISC-R. The reliabilities reported are taken from the publisher's manuals. The intercorrelations, however, are calculated for the data from this study. The WISC-R was reported in two forms, as conventionally reported in Verbal (V) and Performance (P) subscores and as the Witkin factor scores - Analytic Functioning (AF), Verbal Comprehension (VC), and Attention Concentration (AC). The AF score is composed of the Block Design, Picture Completion, and Object Assembly subtests. The VC score is composed of the Vocabulary, Information, Comprehension, and Similarities subtests. AC combines the scores from Digit Span, Arithmetic, and Coding (Witkin, et al, 1974).

Although the original algorithm accepted a discrepancy between achievement and either of these sets of ability scores, the magnitude of the correlations between V and VC and P and AF makes any pair of discrepancies using these scores most likely to be the result of using the same information twice. Only one of the two possible scoring procedures for the WISC-R is therefore appropriate. A

second consideration of the data in Table III deals with the amount of unique information about intrapersonal differences available from the two scoring systems. For conventional scoring, approximately 64 percent of the variance is unique if both V and P are used. In the Witkin scoring system, however, the unique variance attributable to the part scores ranges from 68 to 78 percent. After taking into account the reliability of the measures, it still appears that the Witkin scoring provides the most usable information. The remainder of Table III presents the measurement error bands which must be exceeded before a discrepancy is judged to be significant.

The Final Decision Rule

Taking into account the characteristics of the data, the final decision rule is expressed in the following sequence of decision points:

1. A difference or discrepancy of 10 points (11 if AC is a contributing score) within the three Witkin factors will count as one toward the LD/nonLD decision. Only one discrepancy may be counted from this source.

2. A difference of 15 points between the reading and math tests will count as one discrepancy toward an LD decision.

3. A discrepancy of 10 points between the reading score and any Witkin score will count as one toward an LD decision. Ony one reading/Witkin discrepancy may be counted.

4. A discrepancy of 15 points between the math score and any Witkin score will count as one toward the decision. Only one math/Witkin discrepancy may be counted.

5. If three discrepancies are accumulated according to these rules, the youth is classified as LD.

6. If two discrepancies are present among the six comparison sets, and if any one of the following conditions is also present, the case is classified as LD.

   a. A Bender score of three or more.
   b. Pronounced characteristics (a score of 1) in the WISC-R observations on two or more cases.
   c. Three or more occasionally observed characteristics (a score of 2) in the WISC-R observations.
   d. Three or more (a score of 1) pronounced characteristics in the behavioral observations.

7. If only one profile discrepancy is present, a score of three or more on the Bender, and the presence of two or more pronounced behavioral characteristics will classify as LD.

8. If no discrepancies are present but achievement t-scores of 40 or less and occasional or pronounced characteristics in behavioral observations of difficulty in following oral direction, of gross or fine motor difficulty, of difficulty in oral expression, of distractability, and at least one score of 1 or 2 in the WISC performance observations are present, and the WISC Full Scale score is at least 33, classify as LD.

9. All other cases are classified as non-LD.

10. All cases classified as LD are reviewed for the presence of reading and math t-scores of 50 or greater or WISC-R Full Scale t-scores of 32 or less. These cases are reclassified as non-LD.

The rule through step 9 classified 235 cases as LD on the basis of three or more discrepancies. Forty were added by scores on the Bender. Twenty five more were classified by three or more observations of unusual performance on the WISC-R in addition to two discrepancies on the test scores, and only 18 were added by all other combinations of symptoms. The application of this rule through step 10 produced the prevalence estimates presented in Table IV.

The prevalence values presented in this table give a clear indication that whatever factors are at work in these communities dividing the boys into adjudicated and non-adjudicated populations are also associated with the phenonmenon of LD as defined in this study. The data do not say that a boy who is LD is or will become an adjudicated delinquent. We found more LD boys who were not delinquent than those who were. It is not possible to calculate from these data the likelihood that a boy with LD will become an adjudicated delinquent. The determination of that likelihood would probably require a longitudinal study over a period of at least five years.

Another interesting observation from the data is this. Three percent of the non-delinquent sample were classified as LD on the basis of wide discrepancies within their performance on the tests or the presence of the other indicators, but had somehow been able to achieve atleast average scores for their age in reading and math, and were reclassified as non-LD by step ten. In contrast, only one delinquent, less than a third of a percent, showed such a pattern. Could it be that the academically successful boys had developed a coping style, in spite of a potential handicap, which provided a measure of school success and also kept them out of trouble?

Conclusions and Recommendations

The data presented herein represent a rational estimate of the prevalence of LD in the adjudicated and non-adjudicated populations which were represented in this study, in terms of the operational definition which was applied. Additional analyses of the data are recommended and planned. One analysis will examine further the interrelations of the assessment battery, including the markers, to address the question of whether or not the various tests are performing as expected in these populations. This analysis will undoubtedly produce a better understanding of the problem. Another analysis has been suggested. It specifies a redefinition of LD in order to rule out a larger portion of cases who show discrepant profiles but who have developed academic coping styles which permit them to achieve at approximately average levels. One simple step suggested for this analysis is to reduce the level of acceptable achievement to a score of 40, representing an average performance of about two years below level in both reading and math. Another suggestion is to increase the required magnitude of the discrepancy so that it represents intrapersonal inconsistency in performance congruent with three or even four years. Each of these analyses can be readily

done with the data in its present form, and each is quite likely to be undertaken. The trouble is that a decision about what is mild, moderate, or severe should be made in terms of the consequence of the decision to the individual, and should only be made if it will lead toward corrective action or toward recognition of unattainable expectations. LD is a phenomenon which cannot be legislated or regulated, desirable as that may be for funding purposes. It may only be discovered.

I am certain that alternative research design and additional data are necessary to make that discovery sufficiently unambiguous to allow us to properly serve the students who are so handicapped.

Table I

WISC Observations	Pronounced Characteristic	Occasionally Observed Characteristic	Not Observed
Block designs - perseverating in patterns	☐	☐	☐
Inability to perform on sequence test	☐	☐	☐
Inability to complete any math problem	☐	☐	☐
Inability to complete puzzles	☐	☐	☐

Observations	Pronounced Characteristic	Occasionally Observed Characteristic	Not Observed
Difficulty following oral directions	☐	☐	☐
Low Frustration Tolerance -- Early onset of fidgeting, inattentiveness	☐	☐	☐
Guarded response style (may be withdrawal, hostile response, evasive response)	☐	☐	☐
Repeated verbalization of inability to learn	☐	☐	☐
Gross motor difficulty -- unusual awkwardness	☐	☐	☐
Fine motor difficulty -- difficulty with handling pencil or similar tasks	☐	☐	☐
Manifestation of vision problems e.g.,squinting, holding books very closely, rotation of paper	☐	☐	☐
Manifestation of hearing problems e.g., favoring one ear, focusing on speakers lips	☐	☐	☐
Continuous rocking, tapping, drumming	☐	☐	☐
Difficulty in oral expression disjunctive sentences,inconsistent grammatical errors, long latency for common words	☐	☐	☐
Distractability	☐	☐	☐

TABLE C

CREIGHTON INSTITUTE LD/JD CROSS-SITE PROTOCOL CHECK: Summary

TOTAL: 42 cases

Site			Site I														Site II														Site III													
Case			1	2	3	4	5	6	7	8	9	10	11	12	13	14	1	2	3	4	5	6	7	8	9	10	11	12	13	14	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Judgment by Site	Site I	NON LD			T	T	T	T				T	T	T	T	T			X	X	X	X				X	X	X	X	X	O	O	X	X			O	O	X	X	X	X	X
		LD	T	T					T	T	T						X	X					X	X	X										O	O	X							
	Site II	NON LD											X						T	T	T	T				T	T	T	T	T		O					O				X	X	X	X
		LD	X	X	O	O	O	O	X	X	X	O		O	O	O	T	T					T	T	T						X		O	O	O	O		X	X	O				
	Site III	NON LD		O	X		X						X						X	X		X	O		O	X	X	X	X	X			T	T	T	T				T	T	T	T	T
		LD	X			O		O	X	X	X	O		O	O	O	X	X			O			X							T	T					T	T	T						
Agreement/ Disagreement				p	p	t	p	t				t		t	t	t					p		p		p						p	t	p	p	t	t	p	p	p	p				

Codes T = original decisions
X = agreement with original decisions
O = disagreement with original decisions
p = partial disagreement (by one site): 13
t = total disagreement (by both sites): 9

Prepared by: Sue Fesmire
Date: September 30, 1977

Table III

PROFILE CORRELATIONS AND ERRORS OF MEASUREMENT

	r_{tt}	V	P	AF	VC	AC	WOOD COCK	KEY MATH
V	.95		.604	.541	.903	.749	.696	.747
P	.90	7.41		.920	.599	.592	.481	.561
AF	.89	8.43	9.66		.562	.468	.447	.524
VC	.95	6.10	7.74	8.70		.547	.671	.691
AC	.82	9.07	10.23	11.00	9.33		.554	.659
WOOD COCK	.98	4.82	6.76	7.84	5.29	8.55		.685
KEY MATH	.96	5.61	7.35	8.35	6.02	9.02	4.63	

TABLE IV

LD/JD STUDY: ESTIMATES OF LD/JD PREVALENCE FOR THREE SITES

CATEGORY	RECORDS[a] REVIEWED	LEARNING DISABLED	
		N	%
PS	984	161	16
JD	397	127	32

[a]THESE ARE ALL CASES IN THIS SAMPLE WITH COMPLETE DATA.

PREPARED BY: M. L. TRAINOR
ETS PROJECT MANAGER

DATE: MARCH 1, 1978

References

Murray, C. A. , et al. The link between learning disabilities and juvenile delinquency - current theory and knowledge. American Institutes for Research (prepared for NIJJDP, LEAA). April, 1976.

Witkin, H.A., Dyk, R.B., Faterson, H.F., Goodenough, D.R. and Karp, S.A. Psychological differentiations. New York: John Wiley & Sons, 1974.

Barrows, T. S., Campbell, P. B., Slaughter B. A., and Trainor, M. L. Psycho-educational Diagnostic Services for Learning Disabled Youths. Research Procedures. Educational Testing Service, Princeton, N.J. 1977.

END