GAIN ANALYSIS IN PLACEMENT TEST VALIDATION: An Example of Using the Descriptive Tests of Mathematics Skills

Brent Bridgeman

The utility of raw gain analysis for validation of college placement tests is demonstrated with scores from the Descriptive Tests of Mathematics Skills. Gains during one semester for students in various freshman level mathematics courses were evaluated. Two junior colleges and three four-year colleges provided the data. In addition to total gains, differential score gains on particular item clusters and gains of students who received good or poor grades in their mathematics courses were evaluated. Results suggest that gain analysis is a useful tool for placement test validation.

Test validation sometimes is narrowly considered as simply investigating the correlation between scores on a test and scores on a criterion, or the relationship between test items and the content domain. Although these are important areas, they are not the only ways in which information relevant to test validity may be provided. As noted by Cronbach (1971), "validation examines the soundness of all interpretations of a test—descriptive and explanatory interpretations as well as situation-bound prediction" (p. 43). One method of test validation that has received relatively little emphasis is the analysis of raw score gains from the beginning of a course to the end. Although the basic data for this gain analysis are similar to those found in a classical pretest-treatment-posttest research design, the emphasis is quite different. In the classical design, the effectiveness of the treatment is the primary issue and the measuring instrument is presumed to be valid. In the test validation setting, the primary emphasis is on the sensitivity of a

Brent Bridgeman, Division of Educational Research and Evaluation, Educational Testing Service, Princeton, NJ.

Research in Higher Education Vol. 15, No. 2, 1981

0361-0365/81/060175-12$01.50

particular measuring instrument for assessing treatment gains that are presumed to exist.

If a test is accurately targeted to the content of a course (and if one assumes that instruction in that course is at least somewhat effective), then a valid test should be sensitive to gains in knowledge. Gain analysis is particularly relevant for tests that may be used to exempt students from college course requirements (Willingham, 1974). The correlation between test scores and grades is one traditional index of the effectiveness of placement tests (Hills, 1971). Although such correlational evidence may be adequate in some circumstances, either by itself or as part of a demonstration of an aptitude-treatment interaction, for some exemption decisions, it is clearly insufficient. Given the generally high correlation between verbal aptitude and mathematical aptitude (about .67 on the SAT [Donlon & Angoff, 1971]), it is possible to predict success in a calculus course from scores on a verbal aptitude measure. However, one still might be reluctant to exempt students from a precalculus course requirement on the basis of a verbal aptitude score, especially if the exemption implied some degree of mastery of the skills taught in the precalculus course. Verbal aptitude test scores would show practically no gain during a one-semester precalculus course, but scores on a test that was designed to accurately reflect the skills taught in that course should demonstrate significant gains.

In addition to consideration of the magnitude of the gain in a particular course, gain analysis can yield additional useful information by investigating differential score gain on particular clusters of items, by indicating differential gains made by students who got good or poor grades in the course, and by comparing gains made in one course with gains made in another course. In the current study the utility of gain analysis is demonstrated with results from a validation study of a new set of tests designed to assist in placement decisions for college mathematics courses, the Descriptive Tests of Mathematics Skills (DTMS). In the validity study, results from several different validation procedures were presented (Bridgeman, 1980), but only the results relevant to gain analysis are included here.

Gain analysis is proposed as a useful additional validation technique and not as a substitute for other approaches. Thus, for example, content validation is an important component in the validation of a placement test. A survey of faculty members should indicate that most items are related to course content and that there are not substantial areas of the course for which there are no items. If the test content were deemed not appropriate, large gains could not be expected. But if the content were rated as appropriate and gains were still not found, one might question either the accuracy of the faculty judges or the quality of the instruction. Similarly, if content were rated as inappropriate but large gains were observed, one would need to carefully investigate the possible reasons for the failure of the different sources of information to yield a consistent picture. In a multi-method validation, where conclusions from the various methods converge, there can be considerably more confidence in the conclusions than if only one method were attempted.

Although analysis of score gain is a potentially valuable method of empirically assessing test validity, such analyses must be interpreted cautiously, as there are a number of serious problems in assessing raw gains (Harris, 1963). Regression to the mean, for example, could produce spurious gains in a group that was placed in a remedial section on the basis of low test scores. However, since DTMS scores were not used for selection in the current study, the portion of the regression that was due to errors of measurement on that instrument was eliminated. Furthermore, given a highly reliable selection test and reasonably high correlations of pretest and posttests, regression effects should not be too large, and should be essentially nonexistent for middle-ability students who are self-selected into typical courses.

Inflated gain estimates also could result from students who find the answers to specific items from the pretest and remember these answers on the posttest. If, as a result of the pretest experience, the student finds out how to solve problems of the same type that caused difficulty, this should be considered as real learning; it would be a spurious gain only when learning a specific correct answer did not generalize to similar items in the same content domain. In the current study, this type of gain was reduced by not allowing the students to see the questions or their answers after the pretest administration. In addition, students knew that they would not be graded on their gains on the DTMS and, hence, were not motivated to memorize answers to specific questions. Nevertheless, some practice effect was undoubtedly present, and small gains must, therefore, be interpreted skeptically.

Gain analyses will be presented for courses in arithmetic, elementary algebra, and precalculus.

METHOD

Sample Selection

Criteria for selection of institutions for the comprehensive validity study were as follows:

1. Broad geographical distribution
2. Both two-year and four-year institutions
3. Range of selectivity from fairly selective to open admissions

Initial contacts were made by College Board field staff to institutions they thought might be willing to cooperate. Thus, the sample recruited represented institutions that were quite diverse on a number of dimensions, but it was not a random sample. Although 36 institutions provided some data for the comprehensive validity study (e.g., pretest DTMS scores plus grades, content analysis survey), most institutions were not asked to administer the DTMS as an end-of-course posttest. Only the five institutions that provided both pretest and posttest DTMS scores are represented in the current analysis. Two of these were two-year colleges; the other three were four-year colleges. More detailed descriptions of each institution are provided where the data from that institution are introduced.

Each institution was asked to administer the tests either to all freshmen or to all students enrolled in mathematics courses that were open to freshmen. The latter option was consistently chosen because it was logistically much simpler.

Test Description

The primary purpose of the Descriptive Tests of Mathematics Skills (DTMS) is "to assist colleges in the proper placement of admitted students within the sequence of mathematics courses offered by a given institution" (Jones, Note 1). There are four DTMS tests: a 35-item Arithmetic Skills test, a 35-item Elementary Algebra Skills test, a 30-item Intermediate Algebra Skills test, and a 30-item Functions and Graphs test. All tests are in a four-choice multiple choice format, and the testing time is 30 minutes for each test. Each test contains three or four descriptive clusters that represent groupings of similar items (e.g., operations with fractions, coordinate plane, and graphs). Complete test descriptions and technical data are available in the *Guide to the Use of the Descriptive Tests of Mathematics Skills* (College Board, 1979).

Procedures

For the validation study only, the four tests were printed in a single book accompanied by a single common answer sheet. Instructions on the test booklets asked students to take the two of the four tests that were most relevant to their high school mathematics preparation. Students with no more than one year of high school algebra were asked to take the Arithmetic Skills test and the Elementary Algebra Skills test; students with more than one year's study in algebra but no trigonometry were asked to take the Elementary Algebra Skills test and the Intermediate Algebra Skills test; students with more than one year of algebra and at least one-half a semester of trigonometry were asked to take the Intermediate Algebra Skills test and the Functions and Graphs test. The front of the test answer sheet requested some basic background information from each student including sex, birth date, number of semesters in various high school mathematics courses, and the grades achieved in these courses.

Pretesting was at the beginning of the fall semester. At the end of the fall semester, the participating colleges were asked to supply final grades in the mathematics courses for all students who took the fall tests. In addition, the DTMS testing was repeated.

RESULTS AND DISCUSSION

Arithmetic Courses

Pretest and posttest DTMS scores were available for arithmetic courses from two institutions. The course in a two-year community college in a large southwestern city (College LL) was described as a course in prealgebra mathematics "designed to develop an understanding of fundamental operations using whole numbers, fractions, decimals, and percentages and to strengthen basic skills in mathematics . . . [it] includes an introduction to algebra." Students in this course generally had less than one year of algebra in high school. Of the 44 students included in the analysis for College LL, 35 reported a grade in high school general mathematics and 29 reported an algebra grade. On a 4.0 point scale (i.e., A = 4.0, B = 3.0, etc.) the average general mathematics grade was 2.1 and the average algebra grade was 1.8. Slightly over 90 percent of the students were white, and the sexes were equally represented.

The other course with a major arithmetic component was designed for special admissions students at a large midwestern university (College L). The first half of this course was a review of basic arithmetic skills; the second half introduced elementary algebra concepts. Students in this course generally had one year of high school algebra. For the 77 (out of 109) students who reported high school algebra grades, the mean was 2.2 on a 4.0 point scale. About 57 percent of the students in this group were female and about 60 percent were white.

Pretest and posttest means and standard deviations and gains for the Arithmetic Skills test and the Elementary Algebra Skills test are presented in Table 1.

Students in both colleges apparently gained in the kinds of skills assessed by the Arithmetic Skills test. Although the end-of-course arithmetic skill level was nearly identical in the two colleges, students in College LL showed more gain because they started at a lower level. Comparison of

TABLE 1. Gains on the Arithmetic Skills and Elementary Algebra Skills Tests for Remedial Arithmetic Courses.

	College LL			College L		
Test	N	M	SD	N	M	SD
Arithmetic Skills: pre		22.45	5.56		25.44	5.41
Arithmetic Skills: post	44	27.84	4.71	109	27.50	4.93
Gain		5.39			2.06	
Elementary Algebra: pre		10.28	4.98		12.88	5.30
Elementary Algebra: post	40	10.80	6.52	108	16.68	6.78
Gain		.52			3.80	

these DTMS gains with gains on the Test of Standard Written English (TSWE) (Breland, 1977) suggests that the arithmetic skills assessed with DTMS may be more susceptible to change in a single course than are the writing skills assessed by TSWE. Of four schools in Breland's report, the largest gain was about .35 of a within-group standard deviation and the smallest gain was about .15 of a within-group standard deviation as compared with the DTMS gain of from .4 to a full standard deviation.

Analysis of the gain in each cluster of the Arithmetic Skills test, presented in Table 2, reveals an interesting pattern. Although gains are consistently higher in College LL (as was indicated on the total sum gains), the pattern of gains is remarkably consistent in the two institutions. Cluster A (operations with whole numbers) contains nine items, and scores in both schools are at ceiling levels even on the pretest, as reflected by the high means and reduced standard deviations relative to the other cluster scores. Thus, mean gains in this cluster are necessarily small. The largest gains are found in Cluster B (10 items on operations with fractions) and Cluster C (10 items on operations with decimals and percents). The small gains in Cluster D (six items on simple applications involving computation) may again reflect some test ceiling problems (though not as severe as in Cluster A) but may also reflect greater difficulty in teaching these skills in a single semester course. Results of the cluster analysis are useful not only for providing evidence that the test may be more valid for assessing certain specific skills than others, but also for the feedback that can be provided to faculty members. Weak areas of the curriculum can be identified so that meaningful course modifications can be made.

TABLE 2. Gains on Each of Four Clusters on the Arithmetic Skills Test.

	Cluster A	Cluster B	Cluster C	Cluster D
		College LL (N = 44)		
Pretest Mean (SD)	8.05 (1.01)	5.09 (2.82)	5.41 (2.12)	3.91 (1.70)
Posttest Mean (SD)	8.59 (.62)	7.93 (1.97)	7.11 (2.09)	4.20 (1.53)
Gain	.54	2.84	1.70	.29
		College L (N = 109)		
Pretest Mean (SD)	8.28 (.92)	7.33 (2.50)	5.62 (2.33)	4.22 (1.44)
Posttest Mean (SD)	8.30 (.89)	8.31 (2.22)	6.59 (2.14)	4.30 (1.37)
Gain	.02	.98	.97	.08

The gains on the Elementary Algebra Skills test (see Table 1) indicate that the half-semester introduction to elementary algebra concepts at College L was considerably more effective than the "introduction to algebra" of unspecified duration at College LL. However, this should not be construed as a criticism of College LL, where limited resources were effectively used in improving the more basic arithmetic skills. This small gain also suggests that using the same form for pre- and posttests need not necessarily result in large gains, thus increasing the credibility of the other gains reported.

Willingham (1974) notes that analysis of score gains for students who earned different grades in a course may be one way of demonstrating the extent to which "the test confirms the teacher's judgments—and vice versa" (p. 164). Score gains for each final course grade in College LL are presented in Table 3. Final grades were not available from College L.

Contrary to findings in a previous study that related test score gains to college grades (Feldman & Kane, 1973), gains in the current experiment were not clearly related to course grades. However, it is important to note a critical difference between the studies. Students in the calculus course described by Feldman and Kane had little or no direct instruction in calculus before the course began, and pretest scores of students who eventually got As or Bs were little different from those of students who got Ds or Es. On the other hand, students in the current study presumably studied arithmetic skills for many years in secondary schools before they took the pretest in college. Posttest scores are largely reflective of these preexisting differences rather than indicating differential score gains as the result of instruction. Correlational evidence paints just about the same picture; the correlation of pretest scores with grades is nearly as high as the correlation of posttest scores with grades (.59 versus .64). Grading standards at most institutions are designed to reward final status and not gain.

TABLE 3. Gains on the Arithmetic Skills Test by Course Grades at College LL.

Grade in Course	*N*	Arithmetic Skills Pretest		Arithmetic Skills Posttest		*Gain*
		M	*SD*	*M*	*SD*	
A	17	26.0	6.2	31.3	3.7	5.3
B	17	21.5	3.4	26.6	3.2	5.1
C	6	19.0	2.7	25.8	1.7	6.8
D	4	16.5	4.7	21.3	6.5	4.8
F	0	—	—	—	—	—

But these data are a good reminder that C students may work just as hard and benefit just as much from instruction as A students.

Elementary Algebra Courses

Two colleges with courses in elementary algebra reported data for the gain analysis. One college was the large midwestern university that also provided scores for the above arithmetic analysis (College L). Special admissions students who scored above a minimal level on a locally developed placement test were placed in the elementary algebra course rather than the developmental mathematics course described above. On the average, students in this course had about one year of algebra in high school. Thirteen of the 16 students reported grades in high school algebra, and the average grade was 2.38 (out of 4.0).

The other institution reporting data for the analysis was a two-year public community college in a large southeastern metropolitan area (College JJ). The sample was about 60 percent female and 80 percent white. On the average, students had one year of high school algebra. Grades in high school algebra were reported by 102 of the 132 students; the average grade reported was 2.02.

Pretest and posttest means and standard deviations for the Elementary Algebra Skills test are presented in Table 4.

Substantial gains were evident in both courses. In College L, pretest scores were close to the level that could be achieved by random guessing (random guess level = number of items [35] / number of choices per item [4] = 8.75). The increased mean and variance in the posttest scores suggest that they provide more meaningful measurement. In both colleges, the posttest scores were still quite low relative to the maximum possible score of 35. This might be due to the fact that no instruction was provided for many of the skills assessed on the test (suggesting that much of the El-

TABLE 4. Gains on the Elementary Algebra Skills Test for Elementary Algebra Students.

	College L			College JJ		
	N	*M*	*SD*	*N*	*M*	*SD*
Elementary Algebra (pre)	16	9.13	3.91	132	11.95	5.12
Elementary Algebra (post)		14.13	6.28		19.42	6.38
Gain		5.00			7.47	

ementary Algebra Skills test was not a valid measure for these courses), or it might be that the test provides a valid indication of the fact that many students failed to learn a number of the skills taught. The analysis of gains by course grade in Table 5 helps to resolve these conflicting interpretations.

Since A students answered an average of 26.9 items correctly, it seems likely that most of the skills on the Elementary Algebra Skills test were taught at some level, but that many of these skills were learned by only the best students. The high correlation between scores on the Elementary Algebra Skills posttest and a composite score reflecting the average of class tests and points given for classwork ($r = .78$) also suggests that the Elementary Algebra Skills test measures algebra skills that are relevant to the specific course.

Unlike the gains on the Arithmetic Skills test, gains on the Elementary Algebra Skills test are systematically related to course grades achieved. Nevertheless, pretest scores of students who ended up having to repeat the course were substantially below pretest scores of students who eventually got As.

TABLE 5. Gains on the Elementary Algebra Skills Test by Course Grades at College JJ.

Grade in Course	*N*	Elementary Algebra Skills Pretest *M*	Elementary Algebra Skills Pretest *SD*	Elementary Algebra Skills Posttest *M*	Elementary Algebra Skills Posttest *SD*	*Gain*
A	14	15.9	4.5	26.9	3.3	11.0
B	31	14.8	5.9	23.8	3.4	9.0
C	28	13.0	3.9	21.5	3.7	8.5
R*	58	8.9	3.5	14.2	5.0	5.3

*No Ds or Fs given in remedial courses; grade of R means that course must be repeated.

TABLE 6. Gains on the Intermediate Algebra Skills and Functions and Graphs Tests for Precalculus Students in Two Colleges.

	College E			College T		
	N	*M*	*SD*	*N*	*M*	*SD*
Intermediate Algebra Skills (pre)	35	18.69	4.90	127	19.98	4.33
Intermediate Algebra Skills (post)		24.43	3.42		24.78	3.75
Gain		5.74			4.80	
Functions & Graphs (pre)	34	14.94	4.15	81	16.16	5.09
Functions & Graphs (post)		23.50	3.54		24.53	3.39
Gain		8.56			8.37	

Precalculus Courses

Pretest and posttest DTMS scores were available for precalculus courses from two institutions. At College E, a four-year public college in a small eastern town, the course was entitled "Precalculus Mathematics." Most of the students reported that they had two years of high school algebra. The average algebra grade was 3.3 The sample was 56 percent male and 89 percent white. The second college reporting scores was a large private university in a major western city (College T). The course was entitled "Introductory College Mathematics" and is listed as a prerequisite for the Calculus I course, although students also could qualify for Calculus I with high school courses in trigonometry and analytic geometry. The course content included sets, functions (including exponential, logarithmic, and trigonometric functions), graphing, systems of linear equations, and analytic geometry. Most students in the course reportedly had two years of high school algebra, with an average grade of 3.4 reported. The sample was 57 percent male and 56 percent white.

Pretest and posttest means and standard deviations and gains for both the Intermediate Algebra Skills test and the Functions and Graphs test are presented in Table 6. On both tests, gains were remarkably similar in the two colleges. The gain of about two standard deviation units on the Functions and Graphs test was especially striking. This gain is considerably

TABLE 7. Gains on the Intermediate Algebra Skills Test for Business Math Students at College E.

	Intermediate Algebra Skills			Functions & Graphs		
	N	*M*	*SD*	*N*	*M*	*SD*
Pretest	11	18.45	6.88	8	13.63	3.81
Posttest		20.27	6.93		15.63	5.37
Gain		1.82			2.00	

larger than the gain on the Arithmetic Skills test. A plausible explanation of this difference in gains is related to the proportion of the course content that is totally new to the student. Because most students have years of experience with basic arithmetic concepts, the benefit of one additional course could not be expected to be very large. But in a precalculus course where many new concepts are being introduced for the first time, large gains may be reasonably anticipated.

The magnitude of these gains can be better evaluated by comparing them with gains of students with similar pretest scores on the Intermediate Algebra Skills test and the Functions and Graphs test who were enrolled in a math course that did not emphasize instruction in algebra or in functions and graphs, although some of these concepts may have been included in the other courses. Specifically, gains on the Intermediate Algebra Skills test and on the Functions and Graphs test for students from College E who enrolled in a course in business mathematics were evaluated. As indicated in Table 7, gains were considerably smaller in these courses. Although very little confidence can be placed in these results because of the very small sample sizes, the analysis of gains in these courses is included here as an illustration of the kind of comparative analysis that may be useful in local validation studies.

CONCLUSIONS

The gain analyses presented above yield a considerable amount of information about the test itself and about the courses. The DTMS appear to be sensitive to instructional gains in college courses in arithmetic, elementary algebra, and precalculus mathematics. They are probably poorer at assessing gains in other skill areas for which they were not specifically designed (e.g., business math). Large gains are sometimes, but not always, associated with high course grades. Analysis of cluster scores indicates that certain clusters of items within a test might be more sensitive to gains in particular courses which emphasize those skills. Although within-school sample sizes were frequently quite small, the replicability of the findings across institutions permits considerable confidence in the results. Gain analysis seems to be a useful technique in the validation of college placement tests.

ACKNOWLEDGMENTS

This report is based upon research supported by the College Entrance Examination Board. Researchers are encouraged to express freely their professional judgment in the conduct of such projects; therefore, points of view or opinions stated do not necessarily represent official College Board position or policy.

REFERENCE NOTE

1. Jones, C. O. *A proposal for the development of the Descriptive Tests of Mathematics Skills,* unpublished manuscript, 1977.

REFERENCES

Breland, H. M. *A study of college English placement and the Test of Standard Written English* (RDR-76-77). Princeton, N.J.: Educational Testing Service, 1977.

Bridgeman, B. *Validation of the Descriptive Tests of Mathematics Skills* (RDR-79-80-2). Princeton, N.J.: Educational Testing Service, 1980.

College Entrance Examination Board. *Guide to the use of the Descriptive Tests of Mathematics Skills*. New York: CEEB, 1979.

Cronbach, L. J. Test validation. In R. L. Thorndike (Ed.), *Educational Measurement*. Washington: American Council on Education, 1971.

Donlon, T. F., & Angoff, W. H. The Scholastic Aptitude Test. In W. H. Angoff (Ed.), *The College Board Admissions Testing Program: A Technical Report on Research and Developmental Activities Relating to the Scholastic Aptitude Test and Achievement Tests*. New York: College Entrance Examination Board, 1971.

Feldman, S., & Kane, M. *The effect of relevant University of Illinois courses on CLEP scores* (Report No. 343). Urbana, Ill.: University of Illinois, Office of Instructional Resources, 1973.

Harris, C. W. (Ed.). *Problems in measuring change*. Madison: University of Wisconsin Press, 1963.

Hills, J. R. Use of measurement in selection and placement. In R. L. Thorndike (Ed.), *Educational Measurement* (2nd ed.). Washington: American Council on Education, 1971.

Willingham, W. W. *College placement and exemption*. New York: College Entrance Examination Board, 1974.

Received June 13, 1980