An Exploration of Students' Low Mathematics Achievement on the State Student Assessment Test, Part II

Thomas H. Fisher
and
Lyn Tabeling
Department of Education
State of Florida

Purpose

The purpose of this study was to review the results of the State Student Assessment Test, Part II in mathematics from 1977 to 1981 to determine the nature of student's greatest deficiencies in mathematics and to explore reasons for the trends in scores which have been seen.

The specific research questions to be addressed were:

1. What have been the most serious problem areas in mathematics as revealed by the State Student Assessment Test, Part II over the last four years?

2. To what extent have low student mathematics scores been caused by item content changes or changes in the psychometric properties of the test over time?

Perspectives on the Problem

The State Student Assessment Test, Part II (SSAT-II) was initiated by the Florida Legislature by the 1976 Educational Accountability Act (Section 229.57 F.S.). This law provided for a revised student testing program which would encompass grades three, five, eight, and eleven and would test all Florida public school children in the areas of reading, writing, and mathematics. Students in the eleventh grade would take a test of the application of these basic skills by solving problems of a practical nature (the SSAT-II). Passage of the SSAT-II was to have been a graduation requirement beginning with the class of 1979. A court injunction has delayed that requirement until the 1982-83 school year *(Debra P. vs. Turlington)*.

The SSAT-II test was created by the Student Assessment Section, Department of Education with the assistance of various contractors. After each assessment administration, the resultant data are released to local school districts. Detailed analyses are made available for school personnel, while other even more detailed analyses are retained by the Department. All such information is considered public, although the tests themselves are maintained under strict security.

Methods of Data Analysis

To determine the most serious problem areas in mathematics, student performance was reviewed for the thirteen skills tested on the SSAT-II from October 1977 through April 1981. The skills were ranked in order of decreasing student performance. The skills with consistently poor performance were designated as "high need" skills.

The test items for those skills were reviewed to determine consistent error patterns. The percent of students choosing each of the options for each item was examined. Items were compared to see if they contained similar stimulus numbers, required similar computational processes, and included an incorrect option obtained from performing the same type of error. All options were carefully scrutinized if they were chosen by 10% or more of the students. For the items used on more than one administration of the test, the percent of students choosing each option for the different administrations was reviewed.

For each of the "high need" skills, the most common types of items and error patterns were determined. On Attachments #2-7, a discussion of each of these types of items and errors is given. Each type of item is stated as a performance objective, and an example of a test item is presented. The options used in each example show the most common types of errors made for the situation presented in the item. The manner in which each option was obtained, and the approximate percent of students who might choose that option are given. The percents given do not necessarily add up to 100%. In most cases, it was possible to determine the types of responses that 85-90% of the students will choose consistently.

In considering the second research question, it was necessary to review the test construction procedures used by the Student Assessment Section to determine if the procedures have contributed to poor student performance on the "high need" skills. This was accomplished by considering the sources for the items, the item review procedures, and the manner in which successive forms of the test are prepared.

Further, selected statistics generated in the annual technical reviews of the testing program were analyzed. These data included test reliability data, basic performance data, and item discrimination data. The reliability data included test-retest coefficients, Brennan-Kane dependability coefficients, and KR-20 coefficients (where available). The item discrimination information was based on corrected point-biserial correlations between mastery of a test item and total mathematics score. These data were assembled into appropriate tables and graphic displays.

Results of the Analyses

Student Performance on the Math Test

In determining the skills which would be defined as "high need," the performance data from the SSAT-II for the years 1977 through 1981 were reviewed. Skills which consistently had performance rates below approximately 78 percent were selected for this study. The complete skill statements are given in Attachment #1. Briefly stated, these skills involve solving the following types of items:

Skill Number	*Short Title of Skill*
17	Elapsed Time
33	Comparison shopping
37	Linear measures
38	Area
39	Capacity
40	Weight

As a beginning point for the in-depth analysis of the data, the overall passing rates for all students were inspected. These data are shown in Table 1. As can be seen, the passing rates are higher now than they were in 1977. Overall passing rates have not increased since 1979; this may be due to the Federal Court's imposition of an injunction against the use of the test in determining who can get a high school diploma.

Table 1

Percent of Eleventh Grade Students
Passing SSAT-II Mathematics

Mo./Year	Total
10/77	64
10/78	74
10/79	78
10/80	78
4/81[a]	78

[a] Tenth grade students

The performance of students on the individual skills herein defined as "high need" was investigated next. These data are displayed in Table 2. The general pattern is one of increasing student achievement, although there is considerable fluctuation between the skills over time.

Table 2

Percent of Grade 11 Students Mastering Six "High Need"
Skills for Five Test Administrations

	Skills					
Year	17	33	37	38	39	40
10/77	36	43	69	47	70	47
10/78	39	50	75	37	68	53
10/79	70	47	76	60	77	45
10/80	65	43	76	51	69	45
4/81[a]	72	49	72	53	71	50

[a] Tenth grade

It is interesting to consider the performance of tenth grade students who took the SSAT-II for the first time in April 1981. Prior to April 1981, it was not known how tenth graders would perform on the SSAT-II. Some hypothesized that their performance would be less than that of eleventh graders because they were younger and the potential grade 10/11 dropouts were still in the group. Others thought that performance would be higher because tenth graders were at the end of the school year and would not be suffering the effects of summer forgetfulness.

The performance of tenth grade students did not differ much from that of the eleventh graders. The overall passage rate was identical to the grade eleven rate. Performance of tenth graders on the six "high need" skills was higher than that of eleventh graders on five of the six skills, although the differences were not great. (See Table 2)

Table 3 presents a comparison of the performance of all eleventh graders in October, 1977, 1978, and 1979, and twelfth grade students from those same test groups. That is, seniors, tested in April of 1979 were first tested in October, 1977. Performance of seniors is lower than that of eleventh grade overall performance. Since the majority of seniors tested were those who had failed the test one or more times, this would be expected. It should be noted that the twelfth grade results include students who were new to the state and had not taken the test before. This would lead one to conclude that the twelfth grade results were probably inflated.

Analysis of Student Deficiencies

Four of the six "high need" skills involve units of measure: #37 - linear units, #38 - area units, #39 - capacity units, and #40 - weight units. Solving items involving units of measure is clearly one of the greatest weaknesses that student have. There is not a skill involving units of measure on which students perform well.

Most of the test items for these four measurement skills involve a conversion of units. A conversion table is always given, and the conversions required are always within the same system (U. S. Customary or Metric). The item specifications for the measurement skills require that the conversions be very explicit. The stimulus narrative for each item must state the unit of the correct response if a conversion of units is required. Example 1 below is an example of an item requiring a conversion that is only implied by the options. This type of item would not be permitted on the SSAT-II. Example 2 below, is an example of an item requiring a conversion that is made explicit by the use of the unit of the correct response in the stimulus narrative.

Example 1: One object weighs 8 ounces and another object weighs 24 ounces. How much do they weigh together?

A. 32 pounds
B. 16 pounds
C. 3 pounds
D. 2 pounds

Table 3

Comparison of Grade 11 and Grade 12 Performance

Skill	Oct. 1977	April 1979
17	36%	22%
33	43%	24%
37	69%	41%
38	47%	27%
39	70%	46%
40	47%	13%
	Oct. 1978	April 1980
17	39%	40%
33	50%	17%
37	75%	46%
38	37%	17%
39	68%	44%
40	53%	15%
	Oct. 1979	April 1981
17	70%	54%
33	47%	53%
37	76%	42%
38	60%	26%
39	77%	41%
40	45%	16%

Example 2: One object weighs 8 ounces and another object weighs 24 ounces. How many pounds do they weigh together?

A. 32 pounds
B. 16 pounds
C. 3 pounds
D. 2 pounds

If an item requires conversion from one unit to another, students tend either not to convert or to use an incorrect conversion factor. Many items that do not require a conversion require the students to perform two operations. This increases the chance for error because students tend either to perform only one of the two operations or to perform an incorrect operation.

Items involving metric units were difficult to analyze. It is too speculative to make many generalizations about metric items without analyzing the students' written work or interviewing the students. It is possible to determine whether or not a student failed to make a necessary conversion, but if the student attempts to convert, it is impossible to determine whether the student chose the wrong conversion factor or whether the student made a computational error or an error in moving the decimal point when making the conversion.

Each of the measurement skills presents particular difficulties for students as indicated on Attachments 2-7. Unlike most of the other skills, many items in the linear measurement skill, #37, involve quantities that are expressed in two units; for example, a quantity may be expressed in both feet and inches. The use of two units to designate one quantity poses a unique problem for students. In an item requiring the student to find the difference in two quantities, both expressed in feet and inches, students will tend to subtract the smaller number of inches from the larger number of inches, regardless of the number of feet in each quantity. The use of two units to express one quantity also creates a situation where a student may try to borrow the usual "10," but, in this case, an error is made because the student is borrowing ten inches for one foot. In this situation, it is impossible to determine if the student actually tried to borrow ten, or if the student believes ten inches is equal to one foot.

The area skill, #38, poses unique problems because students tend to confuse perimeter and area. Many students also will err by adding instead of multiplying to find the area. It appears that more students will use an incorrect conversion factor in items for this skill than for any other skill (i.e., using three square feet for one square yard).

Due to the variety of items and the conversion factors used in the capacity skill, #39, generalizations were not made for errors made by students solving capacity items. Attachment 6 includes examples of the types of items which make it impossible to determine the exact errors that students make.

The weight skill, #40, appears to be particularly difficult because most of the situations presented in the items require the student to perform at least two steps, a conversion and another step. The extra step enables students to make additional mistakes by omitting the step, or by performing the wrong operation. Many students also will use 1,000 pounds for one ton, which doesn't seem all that unreasonable, since there are 1,000 grams in a kilogram.

The other two skills also create particular difficulties. Students appear to be attempting to solve problems in the elapsed time skill, #17, similar to all other subtraction problems, ignoring changes from a time with an a.m. label and one with a p.m. label. Even though some students may properly find the number of months elapsed between two dates, many students will ignore changes from one year to another when counting past December, and they will just subtract the two years given in the narrative.

The most blatant error made when students were asked to find the difference in unit prices for the comparison shopping skill, #33, was to find the difference in the prices given in the narrative even when the given prices are for different quantities. Since many items for this skill require the student to simply identify which good or service is the least expensive, and the responses are non-numerical, no generalizations could be made from these items about the type of errors that students are making.

Review of Test Construction Procedures

The test construction procedures used by the Student Assessment Section have not changed dramatically since the program began in 1976. Test items are constructed especially for the Florida tests through the use of commerical or university contractors. Contractors are selected by a competitive bidding process, although there have been some "sole source" contracts in prior years.

All test items are developed to meet stringent test specifications which were created to guide the work of the test item contractors. The specifications have had extensive input from Florida school district teachers and curriculum specialists. Over the last five years, the specifications have been made increasingly more detailed, thus increasing the consistency of the items which have been produced by contractors.

All test items undergo an extensive review process both external to the Department internally. All items are reviewed to make certain they meet the test specifications. Content validity is constantly checked by reviewers brought to Tallahassee for that purpose.

All test items are pilot tested on small groups of students and then field tested on a representative sample of students statewide. The items are substituted into the existing test only when they match the existing items on the basis of similar content and performance. Approximately 30 to 40 percent of the items are renewed in building a new form of the test.

These procedures have worked very well for the testing program. The procedures have been tightened over the years, and they have been applied evenly across all skills. Students' poor performance on the six "high need" skills cannot be attributed to inadequate test construction procedures.

Analysis of Test Properties

As a beginning step in reviewing the test results in terms of the psychometric properties, the number of items per skill was summarized. This was done in an attempt to see if there was any change which would be reflected in the test results. The number of items per skill has been either four or five and has not changed except for increasing the number of items for three skills from 1977 to 1978. The "high need" skills have a greater proportion of five-item sets than the other skills. However, test data do not indicate that poor performance is related to the use of five-item sets. Hence, differences in student performance are not related to the number of items per skill. (A student must correctly answer 3 out of 4 or 4 out of 5 items to master a skill.)

Test reliability for the SSAT-II has been reviewed several ways over the years. Tables 4 and 5 summarize the results of these investigations. The total mathematics test has reliability coefficients which are above 0.90 regardless of whether the coefficient is based on the traditional KR-20 or the newer Brennan-Kane index of dependability. For individual skills, KR-20 indices were reported in 1977. The values were in the 0.50's. Beginning in 1978, Brennan-Kane dependability coefficients were used. The values for the six skills under investigation herein ranged from a low of 0.40 to a high of 0.84. Test-retest reliability coefficients have been calculated and are shown in Table 5. They vary from 0.73 to 0.82.

The Brennan-Kane dependability data for the "high need" skills and the remaining skills were compared for 1978 and 1980. As can be seen from Tables 6 and 7, the "high need" skills have a slightly lower average dependability index, and the values of the indices do not materially change from 1978 to 1980. In 1979, test-retest information is available in addition to the B-K indices. This information is shown in Table 8. As can be seen, the mean value of the B-K index is similar to those reported in Tables 6 and 7. However, the test-retest indices demonstrate that the "high need" skills have a lower mean value. This is thought to be a reflection of the poorer student performance–students who do not have the skill mastered are more likely to guess, thus generating inconsistent mastery information at the skill level. It should be noted, however, that even for the "high need" skills the test-retest data indicate a classification consistency above 70 percent.

The B-K indices were compared to the skill mastery rates for 1978 and 1980. The resultant plot is shown in Figure I. As can be seen, the data for the two years are remarkedly similar. Generally, the higher the mastery rate, the higher the B-K dependability index, but the relationship is not terribly strong.

The conclusion one draws from the analysis of the test reliability information is that the low reliability for the "high need" skills is a reflection of the students' performance, not the cause of it. The test as a whole contains items which are either very easy or moderately easy. The former will result in stable skill classifications; the latter will result in some inconsistency. If the test had contained very hard items, the results would probably show consistency of classifications for the hard skills as well as the easy ones.

Test item discrimination is monitored by the tranditional point-biserial correlation coefficient corrected for attenuation. The data reflect the relationship between performance on the test items and performance on the total mathematics test. Generally, one would like to have positive, moderately strong correlations in a test. However, in the design of the SSAT-II, items were not eliminated solely on the basis of a low point-biserial correlation coefficient if the item measured an important dimension of a required minimum skill.

The corrected point-biserial coefficients are compared to the distribution of coefficients for the entire mathematics test in Table 9. As can be seen, the coefficients are all positive and most are in the range from 0.30 to 0.49. The coefficients for the six "high need" skills are generally higher than those of the entire mathematics test.

Table 4

Total Test Reliability

Year	KR-20	B-K[a]
1977	.92	NA
1978	.92	.93
1979	.94	.94
1980	.91	.92
1981[b]	NA	.93

[a]Brennan-Kane Dependability Index

[b]Tenth grade

Table 5

Individual Skill Reliability

	Skills					
Year	17	33	37	38	39	40
	KR-20					
1977	.36	.57	.58	.56	.49	.65
	B-K					
1978	.65	.53	.50	.65	.60	.78
1979	.41	.60	.48	.84	.61	.75
1980	.54	.50	.51	.77	.47	.71
1981	.53	.48	.40	.72	.55	.71
	r_{tt}					
1979	.76	.73	.78	.82	.74	.77

Table 6

Reliability of SSAT-II Skills,

1978 Data

High Need Skills		Remaining Skills	
Skill	B-K	Skill	B-K
17	.65	24	.88
33	.53	30	.63
37	.50	32	.52
38	.65	34	.71
39	.60	35	.69
40	.78	36	.48
		41	.84
mean	.62	mean	.68

Table 7

Reliability of SSAT-II Skills, 1980 Data

High Need Skills		Remaining Skills	
Skill	B-K	Skill	B-K
17	.54	24	.89
33	.50	30	.64
37	.51	32	.54
38	.77	34	.67
39	.47	35	.68
40	.71	36	.46
		41	.89
mean	.58	mean	.68

Table 8

Reliability of SSAT-II Skills, 1979 Data

High Need Skills			Remaining Skills		
Skill	B-K[a]	r_{tt}[b]	Skill	B-K[a]	r_{tt}[b]
17	.41	.76	24	.90	1.00
33	.60	.73	30	.70	.93
37	.48	.78	32	.54	.84
38	.84	.82	34	.54	.93
39	.61	.74	35	.69	.89
40	.75	.77	36	.35	.80
			41	.83	.99
mean	.62	.77	mean	.65	.91

[a]Brennan-Kane dependability index

[b]Test-retest percent of agreement index

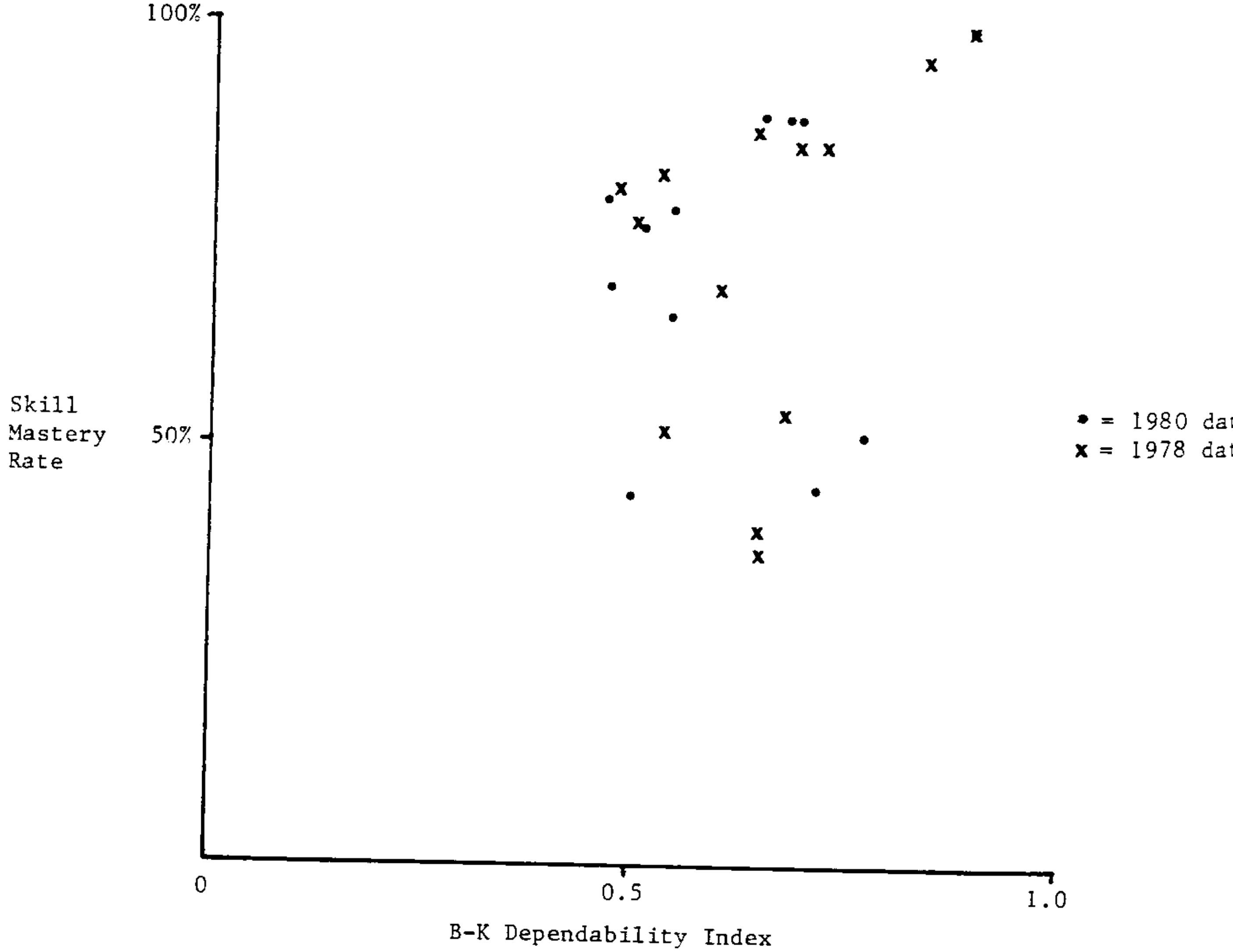

Figure I. Relationship between Brennan-Kane Dependability Indices and skill mastery rates for all skills based on 1978 and 1980 data.

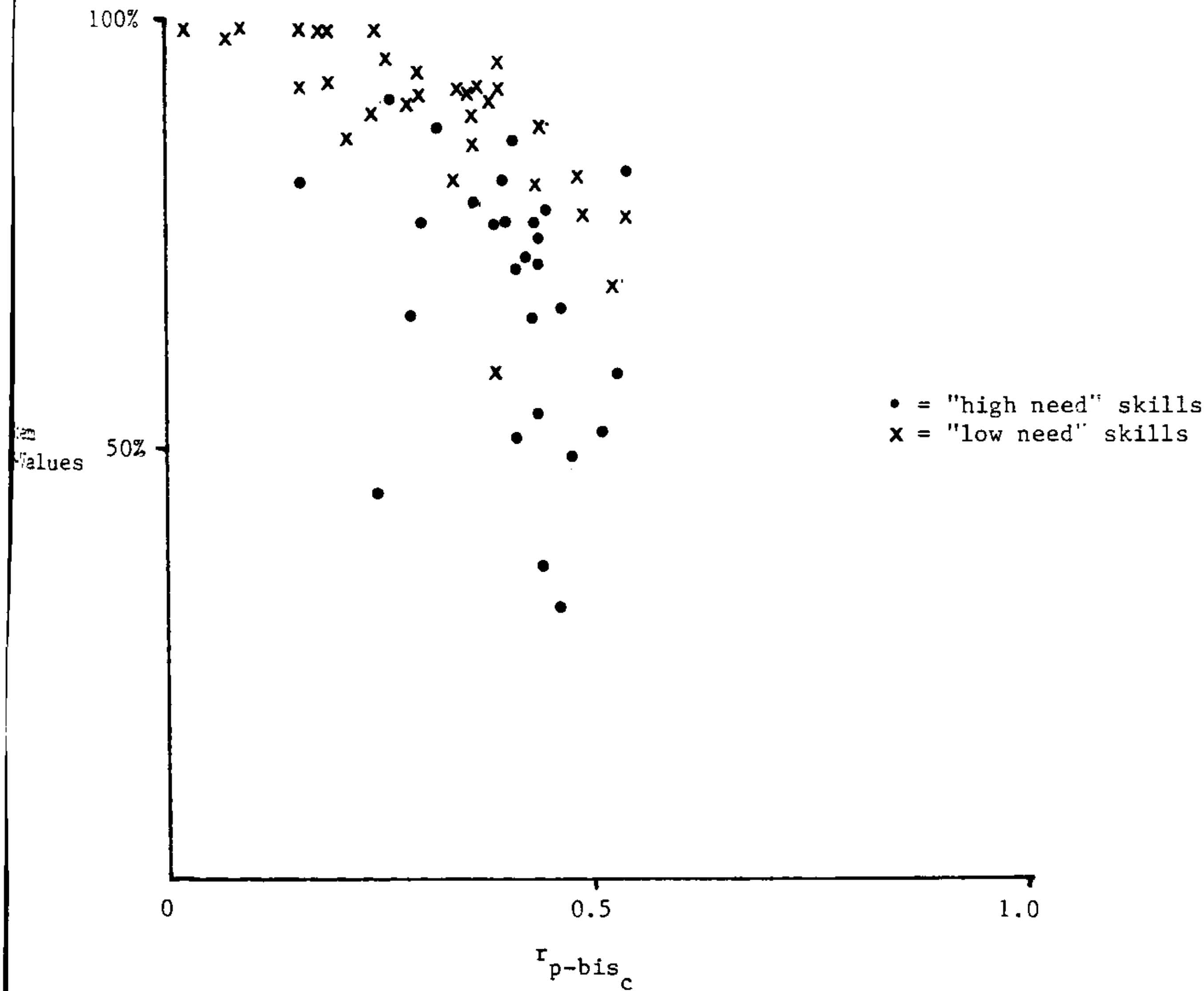

Figure II. Relationship between Corrected Point Biserial Coefficients of Correlation and item p-values for 1980 data.

SSAT-II Mathematics
"High Need" Skill Statements

Skill Number	Skill Statement
17	The student will determine the elapsed time between two events stated in seconds, minutes, hours, days, weeks, months, or years.
33	The student will determine the solution to real world problems involving comparison shopping.
37	The student will solve a problem related to length, width, or height using metric or customary units up to kilometers and miles, conversion within the system.
38	The student will solve a problem involving the area of a rectangular region using metric or customary units.
39	The student will solve a problem involving capacity using units given in a table (milliliters, liters, teaspoons, cups, pints, quarts, gallons), conversion within the system.
40	The student will solve a problem involving weight using units given in a table (milligrams, grams, kilograms, metric tons, ounces, pounds, tons), conversion within the system.

Skill #17 - Elapsed Time

Solving elapsed time problems requires algorithms that are not used in any of the other math skills on the SSAT-II. Although a comparison could be made between this skill and the four measurement skills (since time units are really just another form of measurement units), the process used to correctly solve an elapsed time problem usually involves a counting procedure. Yet students err by attempting to work the problems similar to a regular subtraction problem. This is shown in the three examples given.

In the first example, the most common error is made by writing the "larger number" above the "smaller number" and subtracting as in a problem with a decimal point, ignoring the difference in the a.m. and p.m. designations. If the time with the p.m. label is on the hour and "smaller" than the time with the a.m. label, even more students will make this error.

The performance on the second type of item varies greatly depending on whether the time elapsed passes through December. If it does, then about 55% of the students will simply subtract the two years given, ignoring the counting past December, as shown in type two, options C and D below. Also, for this type of item, which date is stated first in the narrative affects the students' performance. The students tend to start counting the months with the month that is stated first in the narrative, which in many cases will give the student the complementary number of months instead of the correct number. The performance also changed if a response that is one month different from the correct response was used used as an option. About 10% of the students will count one month too many or one month too few.

The third example given creates an additional problem by the use of the word "through." Again, students attempt to work the problem as a subtraction problem, simply subtracting the numbers given. Performance is probably low either due to the students not reading the problem carefully, or the students not understanding the meaning of the word "through," or both.

Common Types of Items

1. Given an initial and final time on the same day, one in the a.m. and the other in the p.m., identify the time elapsed.
   Example: A clean-up committee started work at 6:45 a.m. They finished at 2:15 p.m. How long were they working?
   - 15% A. 4 hours 30 minutes (subtracting small from large)
   - 75% B. 7 hours 30 minutes (correct response)

2. Given an initial and final date, identify the time elapsed.
   Example: Mark was born on May 30, 1968. Cindy was born on August 30, 1965. How much older is Cindy?
   - 10% A. 2 years 3 months (complementary number of months)
   - 15% B. 2 years 9 months (correct response)
   - 20% C. 3 years 3 months (complementary number of months and one year too many)
   - 35% D. 3 years 9 months (one year too many)

3. Given an elapsed time situation using the word "through," identify the amount of time elapsed.

Example: Carrie was on a special diet. She began the diet on April 5. She had to stay on the diet through April 12. How many days was she on the diet?

50% A. 7 days (one day short)

50% B. 8 days (correct response)

Skill #33 - Comparison Shopping

Almost every comparison shopping skill item requires three steps. The reasoning required is much more difficult, and more steps are involved to solve items for this skill than the reasoning and steps required to solve items used to test any of the other math skills. At least two, and in many items three, different prices for a good or service are given. The prices are generally for different quantities, requiring the students to first determine prices for the same quantity, so that the prices may be compared.

Many of the items in the item pool are like type one below. All the student is required to identify is the name of the good that is least expensive. No error patterns could be determined for any items like this example with its non-numerical responses. Each student's work would have to be examined to determine exactly where the error is made.

In the second example, about 40% of the students are subtracting the two prices given in the narrative, ignoring the fact that the prices given are for two different quantities of the good. The students are not realizing that the prices given are for different quantities, and they are not realizing that the item requires them to find the difference in the prices for one unit. The same error is made if the term "per" is used instead of the expression "for one."

Common Types of Items

1. Given two different costs of a good, for two different quantities, identify the least expensive buy per unit.
   Example: Brand A costs 60¢ for 12 ounces and Brand B costs 80¢ for 16 ounces. Which brand cost the least for one ounce?
   - A. Brand A
   - B. Brand B
   - 60% C. Both brands cost the same per ounce.

2. Given two different costs of a good, for two different quantities, identify the difference in the unit costs.
   Example: At Brown's Grocery, 2 pounds of a certain meat cost $4.50. Kelly's Grocery is selling the same type of meat at 3 pounds for $6.60. How much would you save on one pound at Kelly's Grocery?
   - 40% A. $2.10 (subtract the prices given ignoring the difference in the given quantities)
   - 50% B. $ .05 (correct response)

Skill #37 - Linear Units

Student performance on the linear measurement skill is better than on any of the other three measurement skills. The first type of item shown below is a very common type of item in the item pool. In this type of item, when a student is confronted with a situation requiring a regrouping of units, 20-30% of the students will simply subtract the "smaller number" of units from the "larger number" of units, ignoring the fact that the feet and inches must be dealt with as one quantity and cannot be separated, as done to obtain option A in the first example. Another 15% of the students use an incorrect conversion factor, ten inches for one foot. The number of students responding correctly will increase if one of the amounts is a whole number of feet.

The computational process in the second type of item is relatively easy, and no conversions are required. Yet, if two steps are necessary to solve the problem, about 20% of the students will perform only one of the two steps.

Common Types of Items

1. Given two amounts expressed in feet and inches, identify the difference, using one conversion factor.
   Example: Pat and Kim are on a track team. Pat's broad jump record is 6 feet 3 inches. Kim's record is 5 feet 10 inches. How much farther is Pat's record?

20-30%	A.	1 foot 7 inches	(subtracting small from large)
50-60%	B.	5 inches	(correct response)
15%	C.	3 inches	(wrong conversion factor or borrowing incorrectly, using ten inches for one foot)

2. Given distances traveled, identify the total distance, using the same unit as given in the narrative.
   Example: Sam traveled 12 kilometers from Carson City to Mill Town, then 16 kilometers from Mill Town to Youngstown. He returned to Carson City by the same route. How many kilometers did he travel altogether?

20%	A.	28 kilometers	(one step only)
75%	B.	56 kilometers	(correct response)

Skill #38 - Area Units

The area skill is one of the easiest skills to determine errors made by students. Items for this skill are all very similar. Unlike the other three measurement skills, all items for the area skill involve exactly the same step, multiplying two amounts to find the area. The four most common types of items are given.

Finding the sum of the two dimensions given and finding the perimeter of the rectangle account for the most common errors made by students in the first two types of items. Fewer students will find the perimeter, 5-10%, when the situation involves square-foot tiles, as in the second example, instead of simply finding the area, as in the first example where 10-15% found the perimeter. About 10% of the students will consistently add to find the area in these first two types.

The third type requires an additional operation (finding the area of four walls), which causes about 40% of the students who are able to find the area of a rectangle correctly to miss the item. The vocabulary is very simple, so overlooking the second step is probably due to the students reading the item too quickly or not carefully enough. Another 15% of the students realize the additional operation is required, yet they are not able to find the area correctly (as shown by option A).

The fourth type of item is one of the most difficult types of items used for any of the skills. Many students fail to realize a conversion is required. For this type of item, the percent of students using an incorrect conversion factor is greater than for any of the other measurement skills.

Common Types of Items

1. Given the dimensions of a rectangle, identify the area, using the same unit as given in the narrative.
   Example: Bob painted a wall that is 15 feet long and 12 feet high. How many square feet did he cover?
   - 10% A. 27 square feet (sum)
   - 10-15% B. 54 square feet (perimeter)
   - 70-75% C. 180 square feet (correct response)

2. Given the dimensions of a rectangle in feet, identify the number of square-foot tiles needed to cover that area.
   Example: Ms. Smith is going to put new tiles on her kitchen floor. Each tile is one square foot. Her floor is 12 feet long by 8 feet wide. How many tiles does she need to cover the floor?
   - 10% A. 20 tiles (sum)
   - 5-10% B. 40 tiles (perimeter)
   - 75% C. 96 tiles (correct response)

3. Given the dimensions of a rectangle, identify the area of more than one of those rectangles, using the same unit as given in the narrative.

   Example: Jill painted all four walls in a room. Each wall is 15 feet long and 10 feet high. How many square feet did she paint?

   15% A. 100 square feet (sum multiplied by four)
   40% B. 150 square feet (area of one wall)
   35% C. 600 square feet (correct response)

4. Given the dimensions of a rectangle, identify the area, using one conversion factor.

   Example: Mike is going to put fertilizer on his garden. His garden is 12 feet wide and 18 feet long. How many square yards is he going to cover with fertilizer?

   20% A. 24 square yards (correct response)
   25% B. 72 square yards (wrong conversion factor, using three feet for one square yard)
   40% C. 216 square yards (not converting)

Skill #39 - Capacity Units

The errors made on the items for the capacity skill are very difficult to analyze without seeing the actual computational procedure that each student made to arrive at a particular response. Since the permissible U.S. Customary units all involve equivalencies with conversion factors of two or four (2 cups = 1 pint, 2 pints = 1 quart, and 4 quarts = 1 gallon) and many items require two conversions, it is impossible to determine whether the student chose the wrong conversion factor, or performed only one of the two necessary conversions. Some items also make it difficult to analyze the results because a step other than the conversion is required; for example, doubling a capacity or dividing the capacity into equal parts.

The following types of items illustrate the difficulties in making generalizations about students' weaknesses. In the first example, it cannot be determined whether a student obtained option C (24 quarts) by multiplying six by two twice, first to determine Kerry has 12 gallons, then to convert the 12 gallons to 24 quarts, using the wrong conversion factor; or if the student just multiplied six by four to convert the six gallons to 24 quarts, using the correct conversion factor, but performing only one step instead of two. The second type of item creates similar difficulties for analyzing the errors made to obtain each option.

Common Types of Items

1. Given a capacity, identify an equivalent capacity, using one conversion factor.
   Example: Kerry has two 6-gallon containers of a cleaner. How many quarts does he have?
   - A. 3 quarts
   - B. 12 quarts
   - C. 24 quarts
   - 65-75% D. 48 quarts (correct response)

2. Given a capacity, identify an equivalent capacity, using two conversion factors.
   Example: Pat is using a recipe that requires 8 cups of milk. How many quarts is this?
   - 65-75% A. 2 quarts (correct response)
   - B. 4 quarts
   - C. 16 quarts
   - D. 32 quarts

Skill #40 - Weight Units

The weight skill is the most difficult measurement skill for students. Most test items for this skill involve an operation as well as a conversion of units. The following two types of items create the most consistent error patterns made by students for items in this skill.

In the first type of item, performance is especially low because one of the stimulus numbers is a factor of the other stimulus number. If one number is not a factor of the other number, dividing instead of multiplying is not as common a response. Note that the percent of students not converting is 35-45%; this includes the students who made the errors shown in both options A and C.

The second type of item shows the only common use of an incorrect conversion factor that could be identified for this skill, using 1,000 pounds for one ton. Items of this type which involve a mixed number with a fractional component of one-half have a very similar percent of students answering the item correctly. The fraction does not appear to create additional difficulties for the students.

Common Types of Items

1. Given the weight of an object, identify the weight of a specified number of those objects, using one conversion factor.
   Example: One object weighs 6 ounces. How many pounds would 24 of these objects weigh?
   - 20% A. 4 pounds (incorrect operation and not converting)
   - 55-60% B. 9 pounds (correct response)
   - 15-25% C. 144 pounds (not converting)

2. Given a situation involving a weight limit, identify how much an object is over or under the weight limit, using one conversion factor.
   Example: An overseas freight company will not carry any trucks that weigh over 3 tons. Mr. Kennon's truck weighs 8,000 pounds. How many pounds over the weight limit is his truck?
   - 60-70% A. 2,000 pounds (correct response)
   - 10-20% B. 5,000 pounds (wrong conversion factor, using 1,000 pounds for one ton)