

Development and Analysis of a Mathematics Aptitude Test for Gifted Elementary School Students

Pamela Paek and Paul W. Holland
University of California, Berkeley

Patrick Suppes
Stanford University

This paper describes the development and preliminary analysis of a mathematical test targeted for high mathematical ability elementary school students, the Stanford Education Program for Gifted Youth (EPGY) Mathematical Aptitude Test (SEMAT). A version was administered to 248 students, 9-11 years old, in EPGY. The SEMAT was developed because no other satisfactory test was designed or normed for this population. Most standardized tests assess mathematics proficiency for the general population so that gifted students' scores cluster in the few top percentiles. The SEMAT discriminated among this extreme upper end. Item response theory determined proficiency estimates, which were then used as scores to predict various outcomes in EPGY. The SEMAT proved to be a strong predictor of acceleration in EPGY.

This paper describes the development and preliminary analysis of a mathematical test targeted for high mathematical ability elementary school students, the Stanford Education Program for Gifted Youth (EPGY) Mathematical Aptitude Test (SEMAT). This test was developed because other assessments were not satisfactory. There appear to be at least three problems with the existing tests that are typically used to assess young children with gifted mathematics ability: (a) standardized achievement tests rely on a cluster of high scores for gifted students, (b) intelligence tests assess only basic arithmetic skills orally, and (c) tests designed for older students measure students' exposure to mathematics, not their ability.

Standardized achievement tests are commonly used to assess gifted ability, as they are the one consistent measure administered to most students. Since these scores are easily accessible, some programs have nominated a cutoff percentile for gifted programs. For example, EPGY previously considered the 97th percentile as the cutoff score for student admission. The belief was that since achievement tests were normed for age and grade level, only students in the top 3% were truly gifted. These tests are designed for the appropriate age levels, however, they are targeted toward students of average ability so that the top-performing students' scores are clustered at the top of the scale. Thus, gifted students are not assessed for their full range of ability, and the distribution of their scores is limited to the top few percentiles of the general population. The main problem with

the use of achievement tests is the lack of ability to discriminate among the top performing students. The next step would be to have an individually administered intelligence test with a higher ceiling, so students are less limited and can take items beyond their academic level.

Intelligence tests have also been developed for group administration. Group administration saves time and cost, so many schools use these tests routinely as an overall screen of gifted ability. However, many of these tests are closely related to academic achievement; thus, these scores are highly correlated with standardized achievement scores. Another problem is the timing issue, where speed is considered important in identifying gifted ability. Similar to achievement tests, group intelligence tests are also normed to discriminate among the general population and may not be a good discriminator for high performing students. Since the recurring problem is the ceiling effect, where higher performing students' scores cluster in the top percentiles, it is best to find a tool with higher ceilings to show individual differences among this population.

The Wechsler Intelligence Scale for Children (WISC) and the Stanford-Binet Intelligence Scale are individually administered by a school psychologist and involve oral prompts in several different areas. One reason these tests are used is that the ceiling is higher than on group administered tests. These tests assess simple age-appropriate arithmetic skills in terms of the time a student needs to complete each test

item. They do not assess a student's ability to apply basic mathematics knowledge in insightful or innovative ways.

The WISC reports three IQ scores: verbal, performance, and full scale (Wechsler, 1991). These scales allow for identification of nonverbal giftedness (a high score on the performance scale), as well as verbal expression (a high score on the verbal scale). However, many gifted students have significantly different scores on these two scales, as well as on the subtests that comprise scores. Hollinger and Kosek (1986) showed that 85% of gifted students had significantly different subtest scores, which indicates a problem in interpreting the different scale scores. Another problem with the WISC is the lowered ceiling in the latest version, the WISC, Third Edition (WISC-III). The highest possible score on the WISC-III is 155. This new version has a lower ceiling, which causes yet another cluster of scores. The ceiling is higher than group standardized tests, but not high enough to discriminate among highly gifted students. In addition, the WISC-III scores for gifted students run about 9 points less than the revised, earlier version (WISC-R). The WISC-III test thus falls short of discriminating among high gifted ability.

The current version of the Stanford-Binet Intelligence Scale, Fourth Edition, (SB-4) reports four scores: verbal reasoning, abstract/visual reasoning, quantitative reasoning, and short-term memory (Thorndike, Hagen, & Sattler, 1986). The SB-4 has not been recommended for assessing gifted ability (Davis & Rimm, 1998). The upper limit has also been reduced, to a standard score of 164, so unlike the older version (Terman & Merrill, 1973), exceptional abilities (e.g., IQ of 180) will be neglected. Similar to the WISC-III, the fourth edition of the Stanford-Binet runs 13-14 points below the original version. These lower scores result in diminished discrimination among gifted students. The problem again is the ceiling effect, clustering top performing students into a group, rather than differentiating among unusually high gifted ability.

The Study of Mathematically Precocious Youth (Stanley, 1991) utilized the Scholastic Aptitude Test (SAT) to identify mathematically gifted seventh graders. The SAT is targeted for college-bound high school juniors and seniors and is neither designed for nor normed on the population of gifted students in the K-8 range. Using the SAT to discriminate among elementary-level gifted students is strongly affected by their exposure to high school mathematics—geometry and algebra. It is an inappropriate tool to

assess what younger students are able to do, given basic math skills.

The problems with these tests suggest a need to design a test specifically normed for the gifted population. To remedy this lack of an appropriate assessment instrument for young gifted mathematics students, the SEMAT was developed.

Rationale for Development of the SEMAT

EPGY has been in operation since 1990. It has courses in three subjects: mathematics, physics, and English. This article focuses on mathematics (future plans will include assessments in physics and English). The program is dedicated to providing bright students with an intensive mathematics curriculum. The kindergarten through eighth-grade (K-8) mathematics sequence provides a rigorous and complete mathematics curriculum at the elementary school level, with a strong emphasis on conceptual geometry and mathematical foundations.

The Stanford EPGY Mathematical Aptitude Test (SEMAT) was designed to be part of the admissions criteria for the EPGY mathematics courses. The initial form of this instrument was designed for 9- to 13-year-old students. Developing related instruments that are appropriate for younger students is in process. The SEMAT assesses mathematics ability with noncurriculum specific items that are challenging but age appropriate. These items do not require deep mathematical knowledge to solve. Instead, they require the ability to apply insightfully basic mathematics-related skills. All of the test questions are in a four-option, multiple-choice format; no item is part of the EPGY or traditional mathematics curriculum. The items take the form of puzzles, brainteasers, and similar challenging problems. The test has 32 questions, is untimed, and took a sample of EPGY students from 15 to 120 minutes to complete.

A sample of 248 current enrollees in the EPGY mathematics courses took the initial form of the SEMAT during April and May 1997. As part of the initial evaluation of the SEMAT, the data from these examinees was used to address three questions:

1. How well do percentiles from existing standardized tests correlate with SEMAT scores in this population of examinees?
2. How well does the SEMAT predict measures of success in the EPGY program?
3. How well does the SEMAT discriminate among students in the EPGY mathematics program?

Description of the SEMAT

Test Content and Development

The SEMAT consists of 32 items classified into three categories: arithmetic (12 items), geometry (11 items), and logic (9 items). Each item belongs to a single category. The arithmetic items consist of series, probability, and word problems. The geometry items include geometric shapes and spatial rotation problems. The logic items are deductive logic problems, written as word problems. The SEMAT was designed to assess a student's ability to solve mathematics-related problems and was specifically not intended to be individually administered, as are some aptitude and IQ tests.

To ensure that the SEMAT was of an appropriate level of difficulty, several EPGY math educators with extensive experience tutoring EPGY mathematics students reviewed and revised the test items in terms of difficulty and age appropriateness of the questions. They also assessed and approved the nonalgorithmic, noncurriculum specific nature of the items included in the test. Since they teach students in the nation's top 5% (based on standardized test percentiles), the educators' input on the overall difficulty of the test, different item types, and clarity of language was vital in targeting this test to this specific EPGY age and ability level.

Arithmetic and Geometry Items. In these two categories, most of the items on the SEMAT stem from a test designed in Chile by the late Rolando Chuaqui, who worked on an earlier version of this study. A copy of the Chilean test was translated and modified to create a partial, early version of the SEMAT in 1994. One hundred and seventeen students, 9 to 15 years old, were administered this early version of the test.

Based on these data, the majority of the arithmetic and geometry items were modified and expanded to create items for the SEMAT. The preliminary data was used to order the arithmetic and geometry items, from easy to hard, for the SEMAT. Figure 1 illustrates two examples similar to some of the SEMAT arithmetic and geometry items.

For comparison, arithmetic items from the SB-4 and WISC-III tests were analyzed. The items on the SB-4 and WISC-III consist of simple arithmetic tasks, in which the speed of response is assessed (Thorndike, Hagen, & Sattler, 1986; Wechsler, 1991). Geometry items for the SB-4 and WISC-III are in the form of puzzles that are manually assembled by the test taker and the speed of response assessed (Thorndike, Hagen, & Sattler, 1986; Wechsler, 1991). These types of items did not seem satisfactory for the SEMAT for two reasons: (a) the items for the SEMAT were to be more complex than the level of mathematics taught in the classroom, and (b) the SEMAT was to be administered in a paper-and-pencil or computer-delivered format rather than individually administered.

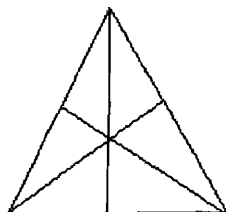
Logic Items. The logic section of the SEMAT was based on Hill (1960) regarding first grade students' logic abilities. Unlike Piaget's theory that most 6- to 8-year-olds have not yet developed deductive reasoning, Hill's research on 270 subjects showed that in oral form, young children can deduce conclusions based on sentences whose forms are drawn from sentential or first-order predicate logic. Since her work showed that the most difficult items were those containing multiple negations, the logic items created for the SEMAT included many negations. In addition, to increase their difficulty, the format of responses for the SEMAT logic items was designed to be four-option multiple-choice, instead of Hill's

Figure 1. Examples of SEMAT arithmetic and geometry items.

Example 1 (arithmetic): What number would come next? 54, 46, 39, 33, 28, ...

- (a) 26 (b) 25 (c) 24 (d) 23

Example 2 (geometry): What is the total number of triangles in the figure below?



- (a) 6 (b) 7 (c) 12 (d) 13

“yes/no” format. Figure 2 is an example similar to a SEMAT logic item.

Quantitative Data Analysis

The Data

The study was conducted in spring 1997 with current EPGY students ranging in age from 4 to 13, with most in the range of 9 to 11 years of age. Of the 350 EPGY students invited to participate in the study, 248 students responded and took the SEMAT. Of these, 246 students had their EPGY curriculum performance, advancement through courses, and demographic information like age, grade, and sex on file with EPGY. Only one-third of these 248 students had documented standardized test score percentiles that could be used in this study. The EPGY students in the study returned their tests either via regular mail or through electronic mail.

The data record for each student consisted of name, EPGY identification number, sex, school grade, age, EPGY grade placement (grade level in EPGY as of the time of testing), EPGY total score (percent correct for all exercises attempted as of the time of testing), total time (in minutes) to complete the SEMAT, number-right score on the SEMAT (unattempted items were scored as incorrect), number-right score for each of the three subscales (arithmetic, geometry, and logic), and the item responses to each of the 32 items.

Figure 2. Example of a SEMAT logic item.

Example 3 (logic):

No student who likes cake likes vanilla.

Hilary is a student who likes vanilla.

Therefore,

- (A) Hilary likes cake (B) Hilary does not like cake
- (C) Hilary likes vanilla cake (D) none of the above

The program ConQuest (Wu, Adams, & Wilson, 1997) was used to compute proficiency scores for each test-taker in this sample. Pooled data for all ages were used in ConQuest to calibrate all of the items and to obtain four proficiency estimates for each student—an overall proficiency for the entire test and separate proficiencies for each of the three subscales. These proficiency estimates were all in the “logit” scale and range from -3 to 3. Using the Rasch model, ConQuest scales both the test questions and the individual examinees onto the same logit scale. Logits are calculated by using the Rasch (or one parameter logistic) model, which calculates the probability of a correct response using the following equations: $P(X_{ni} = \text{correct}) = [\exp(\theta_n - \delta_i)] / [1 + \exp(\theta_n - \delta_i)]$, where θ_n is the proficiency of individual n, and δ_i characterizes the difficulty of item i. The formula to calculate logits is the log-odds: $\log [p(x_{ni} = \text{correct}) / P(x_{ni} = \text{incorrect})] = (\theta_n - \delta_i)$. The term *logit* is a standard reference in statistics to log-odds. Following standard terminology, proficiencies are referred to as “thetas.” The overall theta is an estimated proficiency based on the number of correctly answered items for the whole test. The thetas for the subscales are based on the number of correctly answered items in each subscale. Table 1 lists the means of some of these variables by age and sex. In this table “age 9” means “age 9 or less,” and “age 11” means “age 11 or more.”

As expected, older students’ theta scores were higher than those of younger students; however, this trend was clearer for the boys than for the girls. For 10-year-olds, the performance of girls and boys was about the same, but at the other ages the boys scored higher than the girls. It is difficult to interpret these results, since no attempt was made to make the samples in this study representative of any population. T-tests indicated that none of the differences in Table 1 were statistically significant at the 0.05 level.

Table 1
Means of Some Variables by Age and Sex

Age	Sex	N	Overall SEMAT Theta Score	EPGY Grade Placement	EPGY Percent Correct Score	Time to Complete Test (min)	Total SEMAT Number Right	Arithmetic Number Right	Geometry Number Right	Logic Number Right
9	M	66	-0.25	6.03	90.02	30.53	19.35	7.15	7.35	4.85
	F	29	-0.43	5.49	90.85	25.10	18.03	6.69	6.69	4.66
10	M	40	-0.01	7.55	87.13	26.03	21.38	7.62	8.51	5.24
	F	27	0.02	6.70	88.68	29.08	21.92	7.88	9.07	4.96
11	M	57	0.18	7.89	87.08	21.15	23.51	8.43	9.09	5.98
	F	27	0.02	7.07	85.41	20.17	21.96	7.96	8.11	5.89

In most cases in Table 1, the values for boys were slightly higher than for girls, but *t*-tests (Table 2) indicate that the differences between males and females were not significant for the overall SEMAT theta score, the time for test completion, EPGY Score (the percent correct in the EPGY curriculum), or the entering norm-referenced test (NRT) scores. A *t*-test of the male-female difference in EPGY grade placement showed that this difference in Table 2 was statistically significant ($p = 0.008$). This finding indicated that among boys and girls of the same age, boys were further along in the EPGY courses than the girls were at the time they took the SEMAT. Results discussing EPGY grade placement suggested that boys were more accelerated than girls who were similar in age and grade.

Predictive Validity

Because the SEMAT is to serve as an admissions tool for the EPGY courses, its value was investigated as a predictor of several quantities that were available in this initial study: (a) NRT percentile scores (not available on all students), (b) EPGY percent correct score, and (c) EPGY grade placement. NRT scores and the EPGY percent correct score are natural outcome measures, but they have some problems for this use. EPGY grade placement can be interpreted as a proxy for rate of progress through the EPGY course because, with few exceptions, students are initially placed at their current grade level when entering an EPGY course. Students then progress through the course at their own speed. More precise measures of their rate of progress through the EPGY course were not available at the time this study was designed but will be investigated in the future.

Norm-Referenced Tests

The correlation between the NRT percentiles and the SEMAT theta scores is 0.20 for the 88 students with available NRT percentiles ($p = 0.062$). The main reason these NRT percentile scores are not strongly correlated with the SEMAT is the restriction of range on the NRT percentile scores—with a few exceptions they were all above 95% (see Figure 3). In addition, there are several different standardized tests represented in the NRT percentile scores, so that these percentiles do not refer to the same tests or the same norming populations, thus adding to the difficulty of predicting them. Finally, the SEMAT specializes its assessment in mathematics aptitude, while achievement tests often measure a battery of skills, which are not necessarily related to mathematical aptitude.

Figure 3 is a scatterplot of the NRT percentiles and SEMAT theta values. It shows the strong ceiling effect for the NRT percentiles and the weak relationship between them and the SEMAT thetas.

EPGY Percent Correct Score

The correlation between EPGY percent correct score and the SEMAT theta scores was 0.10 ($p = 0.055$). Figure 4 shows the scatterplot of these two variables. The EPGY percent correct score is constrained to be quite high, due to the mastery conditions required for students to progress through the courses. Students are, with a few exceptions, required to get at least a score of 80% in order to progress to the next stage of the curriculum. This usually results in EPGY percent correct scores between 80-100%. This property of the scores limits their utility as a criterion to be predicted and only weakly reflects real differences in performance among the students in the EPGY course. Essentially, everyone achieves the same EPGY percent

Table 2
T-tests (Male vs. Female)

Variable	Sex	N	Mean	SD	SE	t	p
EPGY Grade Placement	M	163	7.24	2.04	0.16	2.66	0.008
	F	83	6.62	1.56	0.17		
EPGY Percent Correct	M	160	88.23	7.68	0.61	0.42	0.676
	F	82	87.77	8.36	0.92		
Time to Complete SEMAT	M	150	25.26	15.89	1.87	-1.05	0.300
	F	72	24.83	12.03	2.83		
SEMAT Theta Score	M	163	-0.04	0.53	0.04	1.28	0.200
	F	83	-0.13	0.53	0.06		
NRT Percentiles	M	60	98.20	2.06	0.27	0.11	0.306
	F	28	97.40	4.06	0.77		

Figure 3. NRT percentile scores vs. SEMAT thetas.

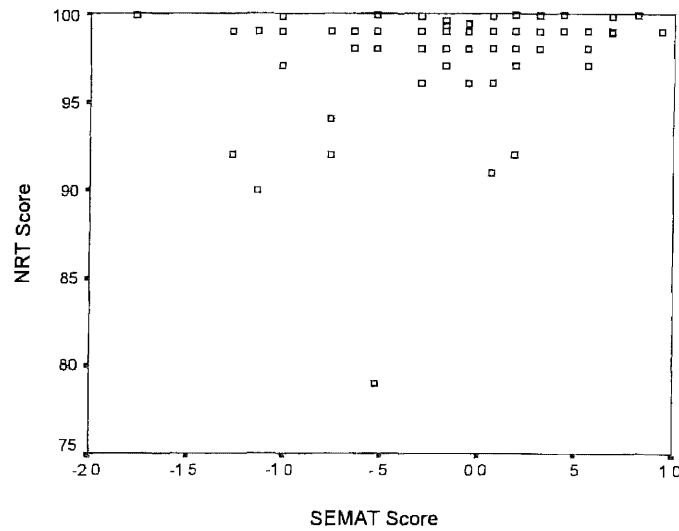
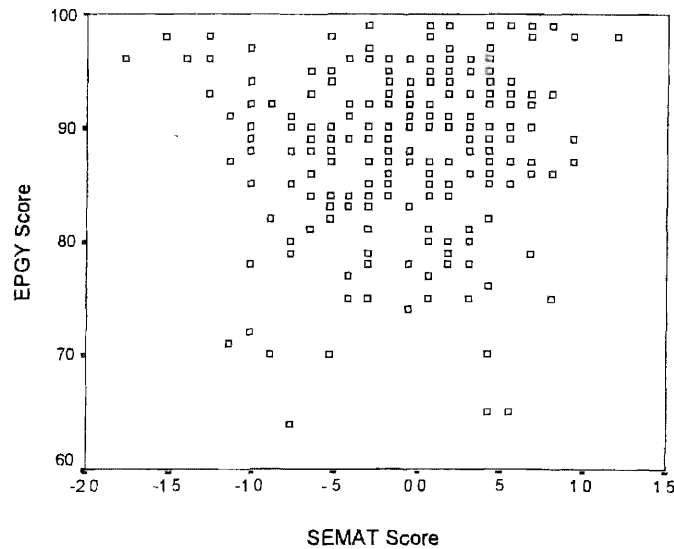


Figure 4. EPGY percent correct score vs. SEMAT thetas.



correct score. What distinguishes students' mastery in the curriculum is the speed with which they move through it, given where they started.

EPGY Grade Placement

The correlation between EPGY grade placement and the SEMAT theta score was 0.49 ($p < .01$). This strong correlation is interesting because EPGY grade placement indicates where the student was in the

EPGY curriculum at the time of the SEMAT test. Because students are initially placed in the EPGY curriculum at their current grade level in school, their grade placement at the time of the test indicates how far they had progressed in the curriculum at that point in time and is a better measure of their ability to master the EPGY material than is their EPGY percent correct score. Figure 5 shows the scatter plot of these two variables, showing their positive relationship.

In order to study the predictive relationship between EPGY grade placement and SEMAT, a more detailed multiple regression analysis was calculated. To interpret the EPGY grade placement as a measure of a student's *acceleration* in the EPGY curriculum, it is necessary to hold fixed the *initial* EPGY grade placement of the student. This calculation was done by including the student's school grade level as a predictor in the equation, as well as the SEMAT theta scores. In addition, the student's age, ethnicity, sex, and the time it took to complete the SEMAT were also included as predictors. The results of this multiple regression analysis are listed in Table 3.

This regression model accounts for 39% of the variance of EPGY grade placement. The results of this analysis indicate that all of the predictors have a significant relationship with EPGY grade placement except ethnicity (white, nonwhite) and time to complete the SEMAT. Holding the other variables fixed, younger students had higher EPGY grade placements than older ones, boys had higher EPGY grade placements than girls, and students with higher SEMAT scores had higher EPGY grade placements than those with lower SEMAT scores.

Examining the regression coefficients, the "B"s in Table 3, there was a logit difference in the SEMAT theta score corresponding to a 1.14 grade difference in

Figure 5. EPGY grade placement vs. SEMAT thetas.

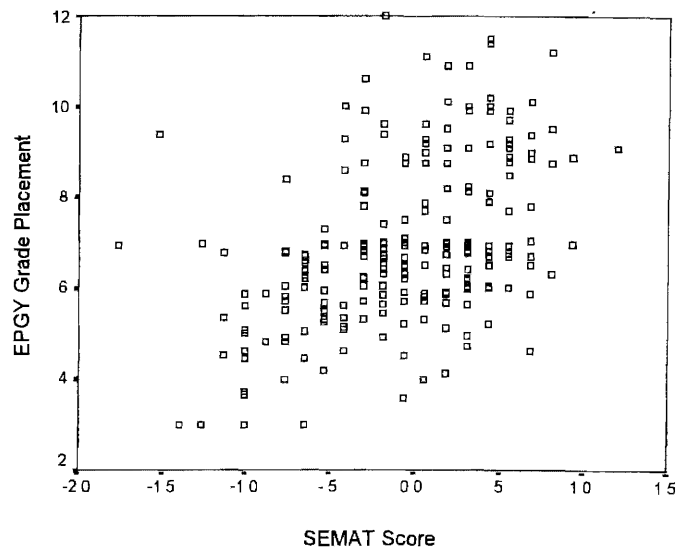


Table 3
Predicting EPGY Grade Placement from Several Variables

<i>R</i>	<i>R</i> ²	Adjusted <i>R</i> ²	<i>SE</i> of Estimate	<i>R</i> ² Change	<i>F</i> Change	<i>df</i> ₁	<i>df</i> ₂
.626	.392	.375	1.509	.392	23.1	6	215
Predictors	<i>B</i> ^a	<i>SE</i>	Beta	<i>t</i> -value	Significance		
Constant	8.411	1.524					
Age	-0.534	0.207	-0.267	-2.58	0.011		
Ethnicity	-0.409	0.265	-0.083	-1.54	0.124		
Female ^b	-0.531	0.216	-0.132	-2.46	0.015		
School Grade	1.060	0.184	0.607	5.75	0.000		
Time to Complete Test	0.059	0.005	0.062	1.13	0.260		
SEMAT Theta	1.14	0.212	0.317	5.41	0.000		

^aRefers to the regression coefficients.

^bFemale refers to the coding of sex as a dichotomous variable.

EPGY grade placement. In other words, a logit higher on the SEMAT predicts that the student will be over one grade higher in the EPGY curriculum beyond students with a logit lower theta, even though they started out in the same place in the curriculum. Girls were about half an EPGY grade below boys the same age who were in the same grade in school and who scored the same as they did on the SEMAT.

Reliability

ConQuest produces a Cronbach alpha reliability, which is the same as the KR-20 when all items are dichotomously scored (Wu, Adams, & Wilson, 1997). The KR-20 value is 0.79 for the SEMAT. This value is typical for a test of this length. For instance, the reliability of the Graduate Management Aptitude Test sections, containing 20 to 25 items, ranges from 0.63 to 0.89.

Item Response Modeling

The SEMAT was fit to the Rasch Model. In such analyses, an item characteristic curve describes the relationship between an underlying, unobserved (latent) "ability" or "proficiency" variable and the probability that the student will give a correct response on that item. In the Rasch model, the only characteristic of each item that is measured (i.e., is an item parameter) is the item difficulty. Both unidimensional and multidimensional analyses of the SEMAT were performed.

Unidimensional Analysis

In the unidimensional analysis, all 32 items were regarded as measuring a single dimension, mathematics proficiency. Using the diagnostics computed by ConQuest, none of the 32 items would be regarded as poorly fit by the Rasch model. For the 32 items, the mean number correct score was 19.82, with standard deviation equal to 4.93. The mean theta score is arbitrarily set equal to the value of 0.7 logits. In Figure 4, Items 1-12 are arithmetic, Items 13-23 are geometry, and Items 24-32 are logic. Although this is a unidimensional analysis, the items can be grouped within each category, and the results are fairly well ordered sets of items, from easiest to hardest.

Items 1-23 were originally ordered based on their "percent corrects" from the 1994 EPGY arithmetic and geometric test items, and this order held up fairly well in this dataset. For example, in the arithmetic items, the order of the difficulty from 1-7 was almost exactly the original order of difficulty. Similarly, for the geometry items, Item 13 was the easiest and Item 23 was the most difficult. The logic items were not as

well ordered in the test: for example, Item 31 was the easiest item in the section.

The right side of the map in Figure 6 shows the difficulty for each of the items on the test, according to the students who attempted the item and answered it correctly. An "X" indicates a student's proficiency estimate. As seen in the histogram, the distribution of scores was slightly negatively skewed, (skewness = -0.41).

Multidimensional Analyses

Correlations between the several different SEMAT theta scores, using the unidimensional and multidimensional results from ConQuest, are listed in Table 4. All of the correlations were significantly different from 0 ($p < 0.0005$).

Similarly to the overall SEMAT theta, analyses of these three subscales showed that within grade and sex, there were positive and significant trends in predicting EPGY Grade Placement. Since these subscales do not predict more than the overall theta score, the three subscale scores are used only for reporting purposes to students.

Conclusions and Future Analyses

Returning to the three evaluation questions, the answers may be summarized as follows:

1. There is not a strong relationship between percentiles from standardized test scores and the SEMAT scores. This result is primarily due to the restricted range of the standardized test percentiles for the population of gifted students.
2. The SEMAT does a good job of predicting how well students progress in the EPGY mathematics curriculum, based on the sample. Acceleration through the EPGY mathematics program is regarded as good evidence of mathematical ability, because the students in the program were required to sequentially master concepts before continuing to a higher level of the curriculum (Suppes & Zanotti, 1996).

Table 4
Correlations Among Dimensions

	Arithmetic	Geometry	Logic	Overall
Arithmetic	1.00	0.76	0.88	0.98
Geometry		1.00	0.68	0.97
Logic			1.00	0.96
Overall				1.00

3. As seen from the histogram of theta values in Figure 6, the SEMAT spread out the gifted students tested and avoided the ceiling effects that occur with tests targeted toward average students.

However, if the SEMAT were used to predict success in EPGY in terms of EPGY grade placement, the data suggest that it would systematically tend to overpredict girls' performance and underpredict boys' performance. This phenomenon needs further investigation. At present, one can only hypothesize about the reasons for this finding. One suggestion is that girls are less likely to move on to the next part of the curriculum unless they feel they have mastered it, whereas boys move on to the next stage as soon as they can. If these ideas are correct, then EPGY grade placement will tend to have somewhat different meanings for boys than girls. This finding could suggest that separate norms for boys and girls should be created, but without further analysis, such action is not justifiable.

Since the SEMAT is currently being administered to students outside of EPGY to see how well it predicts various measures of performance for the general as well as the gifted student population, differential item functioning analyses comparing these and other groups of students will be performed. Task analyses of the items to get a better understanding of how students solve these problems will also be investigated.

Two other EPGY mathematical aptitude tests are under development to assess students ages 7- to 8-years-old and ages 6-years-old and below. These tests will be administered both to potential EPGY students and to public school populations. The 7- to 8-year old version will be initially tested within the EPGY population in both pencil and paper and World Wide Web forms. It is currently being administered to the general population. The six-years-old-and-below test will be made on a CD to allow the inclusion of audio prompts and will be administered to all interested students in that age range (projected date: summer 1999). After further analyses of these three tests within the coming years, more evidence of their predictive value and use in the population of young gifted students will be known.

References

- Adams, R., & Khoo, S. (1994). *QUEST: The interactive test analysis system*. Camberwell Victoria: The Australian Council for Educational Research Ltd.
- Davis, G. A., & Rimm, S. B. (1998). *Education of the gifted and talented* (4th ed). Boston: Allyn and Bacon.
- Fraenkel, J. R., & Wallen, N. R. (1993). *How to design and evaluate research in education* (2nd ed.). San Francisco: Mc-Graw Hill, Inc.
- Gregory, R. J. (1996). *Psychological testing. history, principles, and applications* (2nd ed.). Needham Heights: Allyn and Bacon.
- Hill, S. (1960). *A study of logical abilities of children*. Unpublished doctoral dissertation, Stanford University.
- Hollinger, C. L., & Kosek, S. (1986). Beyond the use of full scale IQ scores. *Gifted Child Quarterly*, 30, 74-77.
- Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence: an essay on the construction of formal operational structures*. New York: Basic Books.
- Stanley, J. C. (1991). An academic model for educating the mathematically gifted, *Gifted Child Quarterly*, 35, 67-70.
- Stillinger, C. (1994). *The performance of "EPGY supplementary test,"* Unpublished manuscript, Stanford University.
- Suppes, P., & Zanotti, M. (1996). Mastery learning of elementary mathematics: Theory and data. In P. Suppes and M. Zanotti (Eds.), *Foundations of probability with applications*. New York: Cambridge University Press.
- Terman, L.M., & Merrill, M.A. (1973). *Stanford-Binet intelligence scale: Manual for the third revision, form L-M*. Chicago: The Riverside Publishing Company.
- Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *The Stanford-Binet intelligence scale* (4th ed.). Chicago: The Riverside Publishing Company.
- Wechsler, D. (1991). *WISC-III: Wechsler intelligence scale for children* (3rd ed.). New York: The Psychological Corporation.
- Wu, M., Adams, R., & Wilson, M. R. (1997). *ConQuest: Generalised item response modelling software, draft release 2*, Camberwell Victoria: Australian Council for Educational Research.

Editor's Note: Correspondence concerning this article should be addressed to Pamela Paek, University of California, Berkeley; 3659 Tolman Hall; Berkeley, CA 94720-1670. Electronic mail may be sent via Internet to ppaek@uclink4.berkeley.edu