APPLIED MEASUREMENT IN EDUCATION, *15*(1), 95–112

Calculator Access, Use, and Type in Relation to Performance on the SAT I: Reasoning Test in Mathematics

Janice Dowd Scheuneman
Quality Assessment Services
Trenton, NJ

Wayne J. Camara
The College Board
New York, NY

Alicia S. Cascallar, Cathy Wendler, and Ida Lawrence
Educational Testing Service
Princeton, NJ

To evaluate the effects of calculator use on performance on the SAT I: Reasoning Test in Mathematics, questions about use of the calculator on the test were inserted into the answer sheets for the November 1996 and November 1997 administrations of the examination. Overall, nearly all of examinees indicated that they brought a calculator to the test and about two thirds reported using them on one third or more of the math items. Some group differences in the use of calculators were observed with girls using them more frequently than boys and Whites and Asian Americans using them more often than other racial or ethnic groups. Use of calculators was associated with higher test performance, but the more able students were more likely to have calculators and used them more often. The results were analyzed further using multiple regression and differential item functioning procedures. The degree of speededness on different degrees of calculator use was also examined. Overall, the effects of calculator use were found to be small, but detectable.

In 1980, the National Council of Teachers of Mathematics (NCTM; NCTM, 1980) called for the use of calculators throughout the mathematics curriculum including

Requests for reprints should be sent to Janice Dowd Scheuneman, Quality Assessment Services, P.O. Box 5964, Trenton, NJ 08638. E-mail: jds@aosi.com

standardized tests. These sentiments were formalized and more broadly promulgated with the release of national standards in mathematics (NCTM, 1989). After a panel of mathematics educators initially recommended against using calculators on the SAT in 1986, the College Board reconsidered this issue 4 years later and announced it would permit students to use a calculator when the revised SAT I was released in March 1994.

Although equity and cost concerns were raised by some, reactions in the mathematics and educational communities were largely quite positive (Rigol, 1993). Calculators were then used in an overwhelming majority of high school math courses by a majority of students in all types of schools. A survey of calculator usage (Maroney, 1990) revealed that over 70% of urban and rural schools permitted calculator use on homework and classwork. Over 70% of urban and 55% of rural students were also permitted to use calculators on some tests; and 65% and 55% of urban and rural schools, respectively, reported that 9 out of 10 college-bound students owned or had regular access to a calculator. The use of calculators on the SAT was therefore seen to increase the instructional relevance and validity of the test.

A series of studies were then undertaken to explore the feasibility and effects of calculator use before the release of the new SAT I. A field trial examined calculator effects on a new item type, student-produced responses (SPRs). Unlike multiple-choice items, the SPRs required students to write (or “grid-in”) the response on their answer sheet. Students using a calculator performed slightly better than students not using a calculator (mean difference of 0.12 on a 55-item test) with no interaction between gender and calculator use (Braswell & Jackson, 1991). A subsequent field trial found that (a) calculators had no effect on speededness; (b) calculator-sensitive items could be predicted by test developers, and thus could be removed from any test form prior to administration; (c) the type of calculator did not significantly affect test performance; and (d) gender and ethnicity had little effect on the size of calculator effects, but prior experience using calculators on tests appeared beneficial (Bridgeman, Harvey, & Braswell, 1995). Calculator use overall did appear to have a moderate effect for students at all ability levels. Overall, these differences were about 10 to 15 points on the SAT-Math compared to a standard deviation of about 100 points.

In June 1991, the College Board began to offer the SAT II Math Level II Subject test in both a calculator-required form (IIC) and a form that did not permit a calculator (II). Harvey, Jackson, and Faecher (1993) found that students choosing to take the calculator-required form did not differ substantially in preparation and background characteristics. Morgan (2000) reported that boys and Whites taking the Advanced Placement Calculus examinations were more likely to own their own calculators and use calculators with the most advanced features (e.g., computer algebra systems), and girls were more likely to use calculators more often in the classroom and on exams than boys.

Calculators can either increase or decrease the construct validity of items on a math test if calculator-sensitivity is not considered before constructing the test (Bridgeman et al., 1995). Researchers have reported significant calculator effects for items requiring complex computations (Hearn & Loyd, 1987; Loyd, 1991), but items that were more conceptual and required less complex computations were equally difficult with or without calculators. Gao (1997) found calculators had a negative impact on 10th-grade students on the PLAN Mathematics test and 12th-grade students on the ACT Assessment. The effect sizes were larger for 10th graders. Lower proportions of students using calculators in 10th grade than in 12th grade may explain some of these differences. In addition, differences between the calculator and no-calculator groups on the PLAN in the 10th grade were largest for African Americans and the mean difference between White students and African American students increased from 2.7 without calculators to 3.1 with calculators ($z = 2.058$, $p < .05$). Mean ACT math scores increased for virtually all groups when calculator use was permitted, regardless of gender, ethnicity, income, and high school grades and courses completed (Colton, 1997).

Since the introduction of the calculator on the SAT in 1994, more tests have permitted calculator use. NCTM (1999) reaffirmed its support of calculator usage, recommending that "authors, publishers, and writers of assessment, evaluation, and mathematics competition instruments should integrate calculator applications into their published work" (p. 5). Today, the ACT, Advanced Placement Calculus examinations, National Assessment of Educational Progress, SAT II: Math Subject tests, and many state assessments in math permit or require calculator use for a portion or all of the test. Equity concerns, such as familiarity with calculators, access to calculators, and regular use of calculators in class and on tests have also decreased. Mullis, Dossey, Owen, and Phillips (1991) reported 98% of high school seniors owned or their family owned a calculator and over 80% of these students had at least a scientific calculator. A 1999 survey of high school mathematics curricula reported that 99.9% of schools either require or permit calculators for precalculus and trigonometry, with 95% requiring or permitting calculators in algebra and graphing calculators increasingly required for advanced courses (Dion et al., 2000).

Students are asked about their access to and use of a calculator when registering for the SAT I. Students are asked if they have regular access to a calculator, what type of calculator they most frequently use, and how often they use a calculator on tests in math and science courses. These questions do not specifically refer to calculator use on the SAT. In 1995, a substantial majority of students from each racial or ethnic and gender group reported some access to calculators (ranging from a low of 89% for Puerto Ricans to a high of 98.5% for girls). Relatively larger differences among these groups were found in the extent of access and the extent of calculator use on in-class tests. Differences of approximately 20% existed among students from different racial or ethnic groups who reported using a calculator almost everyday and using a calculator on every or almost every test. In November 1994, 87.4% of students brought a calculator with them to take the SAT and about 50% reported using it on one third or more of the questions.

This study collected data on calculator use with actual operational SAT Math tests. The purpose of this study was to (a) determine the extent to which students bring a calculator and use a calculator in taking the SAT I; (b) examine student performance on an operational admissions test by calculator use, extent of use, and type of calculator used; and (c) determine if calculator use, extent of use, and type of calculator further affect differences in performance for gender and racial and ethnic groups.

METHOD

All students taking the SAT I in domestic test centers in November 1996 and November 1997 were asked after completing the test to respond to a set of three questions about their use of a calculator. Answer sheets for these two administrations were modified to include these questions on the back page. Test center administrators were instructed to have students complete the SAT and close their test booklets. Students were then asked to read a disclosure statement describing the purpose of this study and to complete the three questions inquiring about calculator use during the test.

The Examination

The mathematics part of the SAT I is made up of three sections consisting of 35 standard multiple-choice items, 15 quantitative comparison items, and 10 SPR items. The questions cover arithmetic, algebra, and geometry topics that have been previously studied by most high school students in their mathematics courses. Emphasis is on reasoning, however, rather than mathematics knowledge. The mathematics score, together with the verbal score, are used by many colleges and universities for admissions purposes. Calculators were approved for use with the mathematics items beginning in March 1994.

The Sample

The full sample of juniors and seniors who took the examination at the regular Saturday administration included 241,743 examinees for the November 1996 administration and 253,576 for the November 1997 administration. For the data analyses, the sample was reduced to those who provided information about their gender and race or ethnicity and who answered at least the first of the three special questions on calculator use on the test answer sheet. This left a total of 202,391 examinees (84%) for 1996 and 215,034 (85%) for 1997. About 10% of this loss was accounted for by students who did not provide their racial or ethnic background. The remainder of the loss was due primarily to students who did not respond to the calculator questions, although more students replied to this question in 1997 than in 1996 (94.5% and 92.5%, respectively). About 1% did not respond to either question. Almost no students failed to respond to the gender question (17 students in 1996 and 15 in 1997).

The composition of the sample in 1996 and in 1997 was very similar. In both years, more than 55% of the sample were girls and about two thirds were White. African American examinees made up 11% to 12% of the samples and Hispanic American and Asian American examinees were 8% to 9% each. Native American examinees comprised about 1% of both samples.

Information about the examinees was drawn from questions on the answer sheet, which included gender as well as the questions on calculator use, and from the student descriptive questionnaire (SDQ). The SDQ is completed by students when they register for the SAT I and includes academic information such as years of math, math courses taken, self-reported grade point average (GPA) in academic subjects, approximate grades in math; demographic information such as racial or ethnic group, mother's and father's education, and family income; and additional information about calculator use. These questions concern frequency of access to calculators, type of calculator used, and whether calculators are used on tests in the student's math and science courses.

Differential Item Functioning (DIF) Analyses

DIF analyses were performed to evaluate differences among groups defined according to their responses to the calculator-use questions on the answer sheets. Analyses were performed contrasting (a) those who indicated that they used a calculator (reference group) with those who did not (focal group), (b) those who used the calculator on most questions (reference group) with those who used it a few times or not at all (focal groups), and (c) those who used a scientific calculator (reference group) with those using a four-function calculator or a graphing calculator (focal groups). This was a total of five contrasts for both 1996 and 1997 data.

The method used for these analyses was the Mantel–Haenszel procedure (Holland & Thayer, 1988). The method compares the right and wrong responses of two groups on a given test item at each level of total score on the test and combines the statistics across levels to get an value for the item. This method has become widely accepted as a valid procedure for identifying DIF. (See e.g., Hambleton & Rogers, 1989; Raju, Bode, & Larsen, 1989.)

Regression Analyses

Hierarchical regression analyses were performed to determine if calculator usage had an independent contribution to performance beyond academic ability and mathematical training and demographic characteristics known to be associated with performance. The four clusters of variables used in the hierarchical analysis were ability and training, calculator variables, demographics, and one variable in self-perception of mathematics ability.

The academic variables consisted of (a) academic ability as measured by the overall GPA, (b) mathematical ability as measured by the grades in math classes, (c) academic year (junior or senior), (d) years of math taken, and (e) whether the examinee had taken or was currently taking precalculus or calculus courses in high school. Data were available for other mathematics courses taken, but they showed little association with performance in preliminary analyses and were not examined further. Related to the academic variables but considered separately in the regression analyses was a self-perception variable, the students' assessment of their own math ability. Also obtained from the SDQ, this variable came from a rating scale in which students classified themselves as in the top 10% of students, above average, average, or below average in math.

The demographic variables used in the analyses included gender, dummy variables representing race or ethnicity and socioeconomic status as measured by mother's and father's education and parental income.

The calculator variables included in the analyses were taken from the three questions on the answer sheet as well as the three questions from the SDQ. The type of calculator was coded as a dummy variable for the three commonly used types—four- function, scientific, and graphing. The type of calculator used from the SDQ was highly correlated with the type of calculator used on the SAT and seemed to be related to the other variables in a similar way. Consequently, it was not included in the analyses.

Because the sample sizes were so large, almost all variables were significant in the regression analyses. To identify the most important variables for inclusion in the hierarchical analyses, step-wise regression was performed within each of the three variable clusters to identify those variables that independently contributed at least 1.0% of the variance. Analyses were performed separately for both the 1996 and 1997 data.

RESULTS

The number and percentage of examinees responding to each of the calculator questions is shown for the total group in Table 1. Table 1 shows that most students, nearly 95%, brought calculators to the November administration of the examination in both years. This is substantially more than the 87% who brought them in November 1994, the only previous occasion such information was collected. The majority of students, however, used them for fewer than half the items. Scientific calculators were used most often, followed by graphing calculators. Results are

TABLE 1
Responses to Calculator Questions

	1996		1997	
Question	*Number*	*%*	*Number*	*%*
Brought calculator to the test?				
Yes	191,804	94.8	203,852	94.8
No	10,587	5.2	11,182	5.2
Used calculator on how many questions?				
None	1,733	0.9	1,694	0.8
A few	64,766	34.0	71,528	35.3
About a third	52,325	27.5	56,343	27.8
About half	40,845	21.4	42,177	20.8
Most	30,938	16.2	31,051	15.3
What type of calculator?				
Four-function	20,407	10.7	18,745	9.3
Scientific	101,072	53.1	101,886	50.3
Graphing	68,087	35.7	80,880	40.0
Other	948	0.5	874	0.4

Note. $N = 202{,}391$ in 1996; $N = 215{,}034$ in 1997.

very similar for 1996 and 1997, except that graphing calculators became more common in 1997, increasing from 35.7% to 40.0%.

The questions concerning calculator use were also considered separately by gender and by racial or ethnic group. The results were quite similar for the 2 years. The results by gender showed that about 2% more girls brought calculators to the exam than boys, 96% versus 94% in 1997. In general, girls used a calculator much more often than boys. In 1997, nearly 43% of girls reported using calculators on half or more of the items compared to about 27% of boys. On the other hand, more than 45% of boys reported use on few or no items. Girls more often than boys used scientific calculators (57% vs. 49%) and less often used graphing calculators (33% vs. 40%).

For ethnic groups, a higher percentage of Whites and Asian Americans brought calculators to the test, about 96% each in both years, and a lower percentage of African American and Hispanic American examinees, 88% and 90%, respectively, in 1997. Whites also used the calculators more often than the other groups with about 40% reporting use on half or more of the items. Hispanic Americans and African Americans used calculators somewhat less often, with only about 32% each reporting use on half or more of the items. For type of calculator, about 46% of Asian American students indicated that they used graphing calculators versus 23% for African Americans, 25% for Hispanic Americans, 29% for Native Americans, and 38% for Whites.

Table 2 shows the association of calculator use with performance on the SAT I mathematics. In general, those with calculators performed better than those with-

TABLE 2
Performance on SAT Math by Calculator Use

	1996		1997	
Question	*M*	*SD*	*M*	*SD*
Did you bring a calculator to the test?				
Yes	506	103	507	104
No	428	103	427	101
Used calculator on how many questions?				
None	454	126	471	124
A few	498	111	500	112
About a third	512	101	512	102
About half	513	97	513	99
Most	508	93	506	94
What type of calculator?				
Four-function	447	97	443	96
Scientific	484	95	481	95
Graphing	557	95	554	98
Total Group	502	104	502	106

Note. N = 202,391 in 1996; N = 215,034 in 1997.

out calculators. Students who used the calculator on one third to one half of the questions performed better than those who used it more or less often. Those students with graphing calculators performed much better than those with scientific calculators, a difference of 73 points in both years. Performance of those with four-function calculators was poorer still. Little difference in performance was observed between the results for 1996 and 1997. Although these results suggest that calculator use is related to performance, the calculator variables are also associated with other variables that may be producing this effect. For this reason, regression procedures were used to determine the independent effects of calculator use.

Regression Analyses

Regression analyses were performed using hierarchical procedures. In hierarchical regressions, the variables are logically grouped into categories. The variables within each category are entered into the regression models as a set. The sets of variables are then entered into the successive models in some logical order according to theory or some expectation about the data. This procedure tends to avoid some of the problems with step-wise regression. The initial step in this study was to reduce the overall number of variables to be considered.

Separately for each of the three groups or clusters of variables—academic, calculator, and demographic—regressions were run in a step-wise manner to identify the most important variables in each. Variables were retained for further analysis if they contributed at least 1.0% to the prediction of total scores on the mathematics sections. A total of 10 variables from these three clusters were retained for further analyses including four academic variables, two calculator variables, and four demographic variables. The variables retained and the regression results for each of the three separate sections and for math self-perception, including the unstandardized regression weight and the percentage of variance lost if that variable were removed from the regression, are shown in Table 3.

As expected, the set of four academic variables were clearly most related to scores, predicting about 45% of the variance in both years. The two calculator variables together predicted about 16% and the four demographic variables 18%. Somewhat surprisingly, the self-perception variable alone predicted nearly 40% of the variance in the math scores.

Because of the apparent relation of calculator use to ability, the next question of interest is how much additional variance in scores is predicted independently by the calculator variables after the academic variables have been taken into account. The results for the Stage 2 analyses, including both academic and calculator variables are shown in Table 4. These results show that the contribution of the calcula-

TABLE 3
Initial Model: Hierarchical Regression Analyses (Predicting Mathematics Scores Separately From Three Classes of Variables)

	1996		*1997*	
Variable	*Beta Wt.*	*% Var. Lost*	*Beta Wt.*	*% Var. Lost*
Academic				
Calculus in HS	.26	7.7	.28	8.4
Math grades	.23	2.8	.23	2.9
Precalculus in HS	.20	3.1	.20	3.0
GPA	.17	1.4	.16	1.2
% Variance accounted for	44.7		45.1	
Calculator				
Graphing calculator	.33	9.8	.33	9.5
Calculator access	.16	2.6	.18	3.0
% Variance accounted for	15.9		16.7	
Demographic				
Father education	.26	5.9	.26	5.7
African American	–.23	5.3	–.23	5.1
Sex	–.14	2.2	–.15	2.3
Hispanic American	–.11	1.2	–.11	1.3
% Variance accounted for	18.4		18.1	
Self-perception				
Math ability	—	—	—	—
% Variance accounted for	38.8		38.7	

Note. HS = high school; GPA = grade point average.

TABLE 4
Second Stage Model: Hierarchical Regression Analyses
(Predicting Mathematics Scores From Two Clusters of Variables)

	1996		*1997*	
Variable	*Beta Wt.*	*% Var. Lost*	*Beta Wt.*	*% Var. Lost*
Academic				
Calculus in HS	.23	6.6	.25	7.1
Math grades	.22	2.5	.22	2.5
Precalculus in HS	.16	2.0	.15	1.9
GPA	.15	1.1	.14	1.0
% Variance accoounted for	44.7		45.1	
Calculator				
Graphing calculator	.14	1.2	.15	1.5
Calculator access	.07	0.4	.07	0.4
% Variance total	46.8		47.5	
% Variance added	2.1		2.4	

Note. HS = high school; GPA = grade point average.

tor variables is reduced to a bit more than 2%. The independent contribution (percentage of variance lost when the variable is removed from the model) of the calculus and precalculus courses, however, was also reduced, suggesting that taking these courses is related to calculator use.

The next stage considers whether the differences in calculator use between the demographic groups might be the source of the remaining variance attributed to calculator use. The results are shown in Table 5. The demographic variables improve prediction, adding about 7% so that the Stage 3 model accounts for a total of around 54.5% of the variance in both years. The independent contribution of the calculator variables in this model is reduced, but not reduced to zero. Given the large sample sizes, such effects are highly significant. The loss in prediction from removing both calculator variables was 0.8 in 1996 and 1.0 in 1997.

The contrast between the effect sizes in the initial separate regressions and the three-stage analyses for the academic and demographic variables are also informative (comparing Table 5 and Table 3). When both academic and demographic variable are included in the regression model, the independent contributions of the demographic variables are reduced except for gender. The contribution of the calculus and math grades are also reduced, suggesting common variation among those academic and demographic variables.

Finally, the self-perceived mathematics ability variable was included in the regression. The results are shown in Table 6. This variable added more than 4.5% in both years, the largest independent contribution of any variable in this model. Although this variable might have been expected to further reduce the predictive effect of calculator use, it affected it very little. The loss if both calculator variables

TABLE 5
Third Stage Model: Hierarchical Regression Analyses
(Predicting Mathematics Scores From Three Clusters of Variables)

	1996			*1997*		
Variable	*Beta Weights*	*B Weights*	*% Var. Lost*	*Beta Weights*	*B Weights*	*% Var. Lost*
Academic						
Calculus in HS	.21	46.3	5.6	.22	49.7	6.1
Math grades	.19	24.9	1.7	.19	25.4	1.8
Precalculus in HS	.15	31.5	1.8	.14	30.7	1.7
GPA	.17	9.3	1.3	.16	27.4	1.2
% Variance accounted for		44.7			45.1	
Calculator						
Graphing calculator	.11	22.6	0.3	.12	24.9	0.6
Access	.06	6.3	0.3	.05	7.4	0.2
% Variance subtotal		46.8			47.5	
% Variance added		2.1			2.4	
Demographic						
Father education	.14	5.2	1.7	.13	4.9	1.5
African American	–.14	– 50.5	1.8	–.13	–47.4	1.7
Sex	–.16	– 33.8	2.4	–.16	–34.4	2.4
Hispanic American	–.07	– 29.9	0.5	–.07	– 29.3	0.5
% Variance total		54.3			54.6	
% Variance added		7.5			7.1	

Note. HS = high school; GPA = grade point average.

were removed from the regression was reduced to only 0.5% for 1996 and 0.8% for 1997.

Again it is of some interest to observe the affects of adding self-perceived math ability on the other variables. The independent contribution of the academic variables other than GPA was markedly reduced in this model, particularly math grades. Interestingly, the variance in mathematics scores predicted by the student's gender, which was largely unaffected by the inclusion of the academic and calculator variables, was essentially halved with the inclusion of self-perceived math ability.

DIF Analyses

The DIF analyses addressed the question of whether calculator use affected examinee performance on individual math items. Analyses contrasted the performance of those examinees with different responses on each of the three calculator questions while controlling for differences in math scores between the groups being compared. The SAT I math items were developed to avoid sensitivity to calcu-

TABLE 6
Final Model: Hierarchical Regression Analyses
(Predicting Mathematics Scores From Four Classes of Variables)

	1996			1997		
Variable	*Beta Weights*	*B Weights*	*% Var. Lost*	*Beta Weights*	*B Weights*	*% Var. Lost*
Academic						
Calculus in HS	.14	31.1	3.5	.15	34.4	3.9
Math grades	.06	7.9	0.1	.06	8.2	0.2
Precalculus in HS	.12	24.3	1.1	.11	24.2	1.0
GPA	.14	7.8	0.9	.13	22.9	0.8
Calculator						
Graphing calculator	.09	18.9	0.2	.10	20.9	0.4
Access	.05	5.5	0.2	.05	6.6	0.2
Demographic						
Father education	.13	4.7	1.5	.12	4.4	1.2
African American	–.14	– 51.4	1.9	–.14	– 49.1	1.8
Sex	–.11	– 24.1	1.2	–.11	– 24.1	1.1
Hispanic American	–.07	– 30.4	0.5	–.07	– 29.8	0.5
Self-perception						
Math ability	.31	40.6	4.6	.31	41.3	4.7
% Variance total		58.9			59.3	
% Variance added		4.6			4.7	

Note. HS = high school; GPA = grade point average.

lator use, so few items were expected to be identified in the analyses. The groups contrasted were (a) brought calculator to test or not, (b) used calculator on most items or on no items, (c) used calculator on most items or on few items, (d) used scientific calculator or graphing calculator, and (e) used scientific calculator or four-function calculator. The results are shown in Table 7. Note that the scientific versus graphing calculator is not shown in this table because no items in either year were found to favor one of these groups over the other.

Often the same items tended to favor one group over another group in the different contrasts. All of the items identified in any of the other contrasts were also identified in the analysis contrasting those using the calculator on no items and those using it on most. Somewhat unexpectedly, this contrast also showed a number of items that favored the group that did not use the calculator. A total of five unique items were identified in 1996, and all were multiple-choice items. In 1997, the five unique items favoring frequent use of the calculator were comprised of one multiple-choice item, one quantitative comparison item, and three SPR items. Four items favoring no calculator use included one each of multiple choice and quantitative comparisons and two SPR items. Nothing in these results suggests an interaction between calculator use and item type.

TABLE 7
Significant Findings in Analyses to Detect
(Differential Item Functioning)

	Number of Items Identified	
Analysis	*1996*	*1997*
Did you bring a calculator to the test?		
Favor calculator use	3	3
On how many items did you use calculator?		
Favor most items	4	5
Favor no items	1	4
What type of calculator did you use?		
Favor scientific over four function	1	2
Total number of items identified	5	9

Examination of the items identified as favoring frequent calculator use showed that these items required either computations (as in finding the area of a geometric figure) or the use of fractions, exponents, or positive and negative signs. The items favoring nonuse of the calculator tended to be reasoning items that included numeric values, but required manipulations for which a calculator was unlikely to be of assistance. Students accustomed to using calculators on most items may have tried to compute an answer from the number provided rather than to think out what the problem actually required. Figure 1 shows examples of the items identified.

Although the items identified as favoring scientific calculators were also identified as favoring calculator use generally, they are of particular interest because of possible differences in the functions available on the two calculator types. The four items identified in the analyses are shown in Figure 2. Two of these items involve manipulation of fractions. One concerns exponents or scientific notation. Why the area problem should favor scientific calculators is unclear.

Speededness

The completion rate for math sections were compared for examinees differing in frequency of computer use. The math items appear in three separately timed sections of the SAT I, two sections of 25 items each with a 30-min time limit and one section of 10 items with a 15-min time limit. The 10-item section and one of the 25-item sections contain only multiple-choice items. The remaining section consists of 15 quantitative comparison items and 10 SPR items.

Of the comparisons, the most interesting results came from the question concerning frequency of use. The rates of completion were lower for groups using the calculator more frequently. The decrease in completion rate across the five differ-

Favors Calculator Use

In Italy, when one dollar was approximately equal to 1,900 lire, a certain pair of shoes cost 60,000 lire. Of the following, which is the best approximation of the cost of these shoes, in dollars?

(A) \$20
(B) \$30
(C) \$60
(D) \$120
(E) \$300

If $a + b = -3$, then $2(a + b)(a + b) =$

(A) 18
(B) 9
(C) -6
(D) -9
(E) -18

Favors No Calculator Use

Points X and Y are the endpoints of a line segment, and the length of the segment is less than 25. There are five other points on the line segment, R, S, T, U, and V, which are located at 1, 3, 6, 10, and 13, respectively, from point X. Which of the points could be the midpoint of XY?

(A) R
(B) S
(C) T
(D) U
(E) V

If the ratio of f to g is 2 to 3 and the ratio of g to h is 1 to 5, what is the ratio of h to f?

(A) 15 to 2
(B) 10 to 3
(C) 5 to 2
(D) 5 to 3
(E) 6 to 5

FIGURE 1 Examples of items identified in the differential item functioning analysis 1996 and 1997.

ent responses to the frequency question from no items to most items was nearly linear. In other words, the more examinees used calculators, the less likely they were to finish. The percentage of students completing each of the three sections as well as the percentage completing all three sections is provided in Table 8.

Not surprisingly, the 10-item multiple-choice section has the highest completion rates and the section with the SPRs has the lowest. More interesting is that the

The positive distance between k and 1/8 is the same as the positive difference between ½ and 1/3. Which of the following could be the value of k?

(A) 1/7
(B) 1/6
(C) 7/24
(D) 23/24
(E) 4/3

A rectangular-shape field has a perimeter of 400 feet and a width of 80 feet. What is the area of the field in square feet?

Student Produced Responses

If $(3 \times 10^n) + (5 \times 10^4) = (5.03 \times 10^4)$, what is the value of n?

Student Produced Responses

If x divided by one-half is 50, what is the value of x?

Student Produced Responses

FIGURE 2 Items favoring the use of a scientific calculator 1996 and 1997.

difference between calculator use on no items to most items is the largest on the 25-item multiple-choice section in which there is a difference of 14%. The change for the section with the SPRs is less than 5%.

The completion rate on the SAT, however, is not necessarily due to speededness of the sections. The SAT I is formula scored with a small penalty for incorrect responses on multiple choice and quantitative comparison items. This provides an incentive for people to omit responses rather than guess. Items at the end of sections tend to be more difficult, and therefore, more likely to be omitted. On the other hand, those examinees using calculators more often (i.e., one half or one

TABLE 8
Percentage of Examinees Completing the Math Sections
(by Frequency of Calculator Use)

	Number of Items Using Calculator				
Section	*None*	*A Few*	*A Third*	*Half*	*Most*
25 MC items	68.9	67.1	64.1	59.1	54.9
25 QC & SPR items	53.7	54.4	52.1	49.8	48.9
10 MC items	85.2	82.5	81.3	79.9	77.6
All sections	41.0	38.9	35.4	32.0	29.3

Note: MC = multiple choice; QC = quantitative comparisons; SPR = student produced response.

third of the time) also tend to be more able than those using them infrequently, making the lack of time a plausible explanation for at least the former group of examinees.

SUMMARY AND CONCLUSIONS

The results of this study showed that nearly all the examinees reported bringing calculators to the test and about 65% used them on one third or more of the items. This is an increase in both the proportion of examinees with calculators and the extent of calculator use from earlier studies. The majority of examinees brought scientific calculators, although the percentage bringing graphing calculators was substantial and increased between the 2 years of the study. It should be noted that 1997 was the 4th year in which calculators were used on the SAT I. Further changes in calculator use may have occurred in the years since these data were collected.

Some differences in calculator use were observed between groups. Girls were more likely to bring calculators than boys and they generally used them on more items. Similarly, White and Asian American students were more likely than African American and Hispanic American students to bring calculators to the test and White students used them on more items than the other groups and African American and Hispanic American students used them on fewer items. In addition, Asian American and White students were more likely to use graphing calculators than Hispanic American or African American students.

Although calculator presence, frequency of use, and calculator type were each associated with test scores, this relation appears more likely to have been the result of able students using calculators differently than less able students. Students who brought calculators to the test and used them on one third or more of the questions performed better than other students did. Those using scientific calculators performed better than those using four-function calculators, and those using graphing calculators performed better than the others. This finding is generally consistent with results from Morgan (2000) illustrating that among advanced placement students, higher ability students are more likely to use graphing calculators with advanced features.

The preliminary regression analyses showed, however, that after the graphing calculator variable was entered into the model, the only calculator variable that accounted for more than 1% of the variance in SAT scores was the one related to calculator access. The percentage of variance in test scores accounted for by calculator access and type of calculator was further reduced when other variables representing academic achievement, student demographics, and self-perceived math ability were included in the regression model. Nevertheless a small percentage of variance was still accounted for by the two calculator variables even in the final model.

The DIF analyses also showed more differences in performance associated with frequency of calculator use than with the other calculator variables. In addition, the DIF analyses contrasting groups defined by calculator frequency showed items both favoring frequent use and those favoring little use of calculators. Items favoring calculator use generally included some type of computation or required keeping track of the sign or decimal point in arriving at the answer. Hearn and Lloyd (1987) similarly found that items on the General Educational Development Mathematics tests requiring computation became differentially easier with calculators and items where the answer was in fraction form versus decimal form were more difficult when calculators were used. The small number of items favoring those not using a calculator often included numbers in item stems, but computations were not an important part of the solution.

Data concerning rate of completion showed a near-linear relation between percentage of items completed and frequency of calculator use, with those using the calculators less often being more likely to complete the examination. This finding may account for the slight drop off in mean scores for those examinees using the calculator on most items as compared to those using it on one third or one half of the items. Speededness with higher frequencies of calculator use was also postulated by Bridgeman and Potenza (1998) based on their results.

A final point of interest concerns the effects of graphing calculators. We were unable to detect any variable to add to the regression analyses that would reduce the contribution of graphing calculators to zero. One possible cause might be that characteristics of particular items favor graphing calculators in ways that DIF analyses fail to detect. Another possibility is that students accustomed to using graphing calculators actually use different metacognitive strategies that may increase their performance on specific items. Differences in students' approach to problems is an area where further investigation would be required. If such differences are found, teachers might then consider how differential strategies could be used in teaching.

Finally, for those who work with students who are preparing to take standardized math tests such as the SAT or ACT, the advice to students is to make sure that they understand the intent of the question before using the calculator. They should learn to be selective about the items on which the calculator is used. The calculator should be used as an aide; using it on all items may take too much time.

The results of this study reflect the increasing use of calculators in mathematics education and assessment. Students do bring and use calculators in taking the SAT and many other large testing programs, and the calculator is increasingly viewed as an integral tool in teaching and the assessment experience.

ACKNOWLEDGMENT

Alicia S. Cascallar is now at Assessment Group, Inc., Falls Church, VA.

REFERENCES

Braswell, B., & Jackson, C. (1991, April). *Feasibility study of calculator use on the SAT.* Paper presented at the annual meeting of the National Council of Measurement in Education, Chicago.

Bridgeman, B., Harvey, A., & Braswell, J. (1995). Effects of calculator use on scores on a test of mathematical reasoning. *Journal of Educational Measurement, 32,* 323–340.

Bridgeman, B., & Potenza, M. (1998, April). *Effects of an on-screen versus bring-your-own calculator policy on performance on the computerized SAT I: Reasoning test in mathematics.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.

Colton, D. A. (1997, March). *Monitoring calculator implementation for the ACT and PLAN.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Dion, G., Harvey, A., Jackson, C. Klag, P. Liu, J., & Wright, C (2000). *SAT program calculator use survey.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Gao, X. (1997, March). *Examining calculator effects on subgroup mathematics performance.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Hambleton, R. K., & Rogers, H. J. (1989). Detecting potentially biased test items: Comparison of IRT area and Mantel-Haenszel methods. *Applied Measurement in Education, 2,* 313–334.

Harvey, A. L., Jackson, C. P., & Faecher, K. E. (1993). *Findings from the first administration of the ATP Mathematics Level II Achievement test with calculators.* Paper presented at the annual meeting of the National Council on Measurement in Education, Atlanta, GA.

Hearn, D. L., & Loyd, B. H. (1987, April). *Use of calculators on standardized math tests: Effects on performance and the potential for bias.* Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.

Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129–145). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Loyd, B. H. (1991). Mathematics test performance: The effects of item type and calculator use. *Applied Measurement in Education, 4,* 11–22.

Maroney, J. M. (1990). *Calculator survey report: The use of calculators in urban and rural schools.* New York and Princeton, NJ: College Board and Educational Testing Services.

Morgan, R. (2000, April). *The impact of the use of graphing calculators on the performance of students taking the Advanced Placement calculus examinations.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Mullis, I. V. A., Dossey, J. A., Owen, E. H., & Phillips, G. W. (1991). *The state of mathematics achievement: NAEP's 1990 assessment of the national and the trial assessment of the states.* Washington, DC: National Center for Education Statistics.

National Council of Teachers of Mathematics. (1980). *An agenda for action: Recommendations for school mathematics in the 1980's.* Reston, VA: Author.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (1999). *Principles and standards for school mathematics.* Reston, VA: Author.

Raju, N. S., Bode, R. K., & Larsen, V. S. (1989). An empirical assessment of the Mantel-Haenszel statistic for studying differential item performance. *Applied Measurement in Education, 2,* 1–13.

Rigol, G. (1993, April). *Balancing educational, administrative and equity interests in the development of a calculator policy for national testing programs.* Paper presented at the annual meeting of the National Council on Measurement in Education, Atlanta, GA.