The College Board

Research Notes

Office of Research and Development RN-16, May 2002

Calculator Use and the SAT® I Math

Janice Scheuneman and Wayne J. Camara

INTRODUCTION

Calculators were first used on the SAT® with the introduction of the SAT I: Reasoning Test at the March 1994 administration. This action followed the 1980 recommendation by the National Council of Teachers of Mathematics that calculators be approved for use in the classroom throughout the mathematics curriculum including standardized testing. By 1994, calculators were being used by an overwhelming majority of high school math students in all types of schools. A more recent survey in 1999 indicated that calculators were permitted or required for nearly all high school math courses. Today, calculators are allowed and even required by numerous testing programs including the ACT, Advanced Placement Calculus examinations, National Assessment of Educational Progress (NAEP), SAT II: Math Subject Tests, and many state assessments.

In addition, calculator use on tests has been supported by research studies considering its effects on test performance. Most often the effects of calculator use have been studied by comparing performance of students on calculator and noncalculator versions of a test. Generally, these studies have shown modest increases in performance associated with calculator use. The increases, however, can be reduced if efforts are made to decrease the calculator sensitivity of items. Generally, calculators make the most difference on items requiring complex computations and little difference on items that are conceptual or require less complex computations. The type of calculator used has appeared to make little difference in performance. The results of these studies have also reduced concerns about equity issues, with little association found between gender and racial/ethnic group and calculator use.

KEYWORDS:

Calculators
SAT I Math

Missing from these studies is a clear association of test performance and actual use of calculators. Where calculators are permitted, students may or may not have brought calculators to the test and may or may not have used them in responding to the test items. This omission was addressed at the November administration of the SAT in both 1996 and 1997 when questions addressing calculator use were included on the answer sheet, permitting scores and item responses to be matched to reports of calculator use on the test. This study uses those data in a series of analyses to examine the relationship of student performance with calculator use.

DESIGN OF STUDY

The answer sheets of the November 1996 and November 1997 administrations included three questions concerning calculator use:

- Did you bring a calculator to the test? (Yes, No)
- If yes, on how many questions did you use your calculator? (None, A few, About a third, About half, Most)
- What type of calculator did you use? (Four-function, Scientific, Graphing, Other)

The sample consisted of 202,391 examinees in 1996 and 215,034 in 1997 who supplied information concerning their gender and ethnicity. Results for the two administrations were nearly the same, so that only the 1997 results will be reported here.

Information about the examinees was drawn from the material 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 in academic subjects, approximate grades in math, racial/ethnic group, mother's and father's education, family income, and additional information about calculator use.

RESULTS

Table 1 shows that most students, nearly 95 percent, brought calculators to the November administration of the examination. This is substantially more than the 87 percent 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.

The questions concerning calculator use were also considered separately by gender and by racial/ethnic group. In general, girls used a calculator much more often than boys. Nearly 43 percent of girls reported using calculators on half or more of the items compared to about 27 percent of boys. On the other hand, more than 45 percent of boys reported use on few or no items. Girls more often than boys used scientific calculators (57 versus 49 percent) and less often used graphing calculators (33 versus 40 percent).

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

In general, those with calculators performed better than those without calculators (see Table 2). Students who used the calculator on a third to a 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. Performance of those with four-function calculators was poorer still. Although these results imply that calculator use is related to performance, the calculator variables are also associated with other variables that may be responsible for producing this effect. For example, students in more advanced math courses would be expected to use a graphing calculator more frequently than other students, and they

TABLE 1
RESPONSES TO CALCULATOR QUESTIONS

Response	Number	Percent
Brought calculator to test?		
Yes	203,852	94.8
No	11,182	5.2
Used calculator on how many questions?		
None	1,694	0.8
A few	71,528	35.3
About a third	56,343	27.8
About half	42,177	20.8
Most	31,051	15.3
What type of calculator?		
Four-function	18,745	9.3
Scientific	101,886	50.3
Graphing	80,880	40.0
Other	874	0.4

TABLE 2
PERFORMANCE ON SAT MATH
BY CALCULATOR USE

Response	Mean	SD
Brought calculator to test?		
Yes	507	104
No	427	101
Used calculator on how many questions?		
None	471	124
A few	500	112
About a third	512	102
About half	513	99
Most	506	94
What type of calculator?		
Four-function	443	96
Scientific	481	95
Graphing	554	98
Other	502	106

would also be expected to have higher math scores on standardized tests. Regression procedures were used to determine the independent effects of calculator use after these other variables were taken into account.

Regression Analyses

Regression analyses were performed using hierarchical procedures to predict math scores. In hierarchical regressions, the variables are logically grouped into categories. The variables within each category are entered into the regression models as a set in some logical order according to theory or some expectation about the data. For this study, four categories of variables were considered. These categories in the order in which they were considered were:

- Academic background variables
- Calculator variables
- Demographic variables
- Self-perception of math ability

Step-wise regression analyses were performed first to reduce the number of variables to be included in the regression models. Variables that independently contributed at least one percent of the variance in the math scores were retained for later analyses.

Four academic background variables were retained for the analyses: grades in math classes, grade-point average in all academic subjects, whether the student was taking or had taken calculus, and whether the student was taking or had taken precalculus or trigonometry. Together these four variables accounted for 45 percent of the variance in math scores.

The majority of the calculator variables contributed little to the variation in math scores. Two variables, however, were retained for the analysis: whether the student used a graphing calculator on the test and the frequency of use of a calculator outside the testing situation (from the SDQ), referred to as calculator access. These two variables together accounted for 17 percent of the variance in the math test scores.

Four demographic variables were also retained for the analyses: father's education, gender, African American, and Hispanic American. Being female, African American, or Hispanic American was associated with lower test scores. Together the four variables accounted for 18 percent of the variance in math scores.

Finally, the fourth category consisted of only a single self-perception variable, the students' assessment of their own math ability. This variable came from a rating scale in which students classified themselves as in the top ten percent of students, above average, average, or below average in math. This single variable was found to predict 39 percent of the variance in math test scores, an interesting finding by itself.

Table 3 summarizes results of the regression analyses. The first column provides the percent of variance in test scores accounted for by only the variables in that category. The following columns show the independent contribution of the categories, that is, the unique variance accounted for by each set of variables, under each of the hierarchical regression models. Model I consisted of only the academic variables.

Model II consisted of both the academic and calculator variables. The calculator variables added 2.4 percent to the percent of variance accounted for by academic variables alone. The independent contribution of the academic variables, however, was reduced. This happens because of common variation between the calculator and

TABLE 3
REGRESSION ANALYSES

Variables	% Variance Alone	Model I	Model II	Model III	Model IV
Academic	45.1	45.1	30.8	27.8	9.0
Calculator	16.7		2.4	1.0	0.8
Demographic	18.1			7.1	5.5
Self-perception	38.7				4.7
% Variance accounted for		45.1	47.5	54.6	59.3

academic variables. Examination of the results for the individual variables suggests that this is primarily because students taking the advanced math courses were more likely to be using graphing calculators.

Model III adds the demographic variables to the others, accounting for an additional 7.1 percent of variance in math test scores. The reduction in the independent contribution of the academic and calculator variables was relatively small, and no particular pattern among the individual variables was apparent.

Finally, Model IV adds the self-perception variable. This variable accounts for an additional 4.7 percent of variance in test scores, bringing the total variance accounted for to nearly 60 percent. Interestingly, it markedly reduces the independent contribution of the academic variables. This probably means that academic background and experience in math strongly influence students' perception of their math ability or vice versa. The reduction of independent contribution of the demographic variables appears to be largely due to reduction of the contribution of the gender variable. Math self-perception seems to be accounting for some of the gender difference in the math test scores but had little effect on the contribution of the calculator variables. Even after accounting for all the other variables, a small contribution of the calculator variables remained.

Differential Item Functioning Analyses

A different approach to the question of the effects of computer use on performance is to evaluate performance at the item level. Do some of the math items favor the use of calculators? Differential item functioning (DIF) procedures were used to address this question. These procedures contrasted the performance of groups defined by their responses to the three calculator questions: (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. Since test developers attempted to avoid calculator sensitive material on the SAT I, few items were expected to be identified.

The method used for the analyses was the Mantel-Haenszel (M-H) procedure. 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 a value for the item. In effect, this controls for overall score differences between groups. The M-H procedure has become widely accepted as a valid method for identifying DIF.

A summary of the results for these analyses is provided in Table 4. No items were identified when contrasting scientific and graphing calculators, so this line has been omitted from the table. The most important contrast turned out to be that between using calculators on no items and most items. The items identified in the other contrasts were also identified by this contrast, so that the five items in 1996 and the nine items in 1997 were the only unique items with significant results.

Somewhat unexpectedly, some of the items identified with DIF were found to favor those who did not use calculators on the test. Examination of the items found to favor 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 much assistance. Students accustomed to using calculators on most items may have tried to compute an answer from the numbers provided rather than to think out what the problem actually

TABLE 4
SIGNIFICANT FINDINGS IN ANALYSES TO DETECT DIFFERENTIAL ITEM FUNCTIONING

Analysis	Number Items Identified	
	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 (or few) items	1	4
What type of calculator did you use?		
Favor scientific over four- function	1	2
Total number items identified	5	9

TABLE 5
EXAMPLES OF ITEMS IDENTIFIED IN THE DIF ANALYSIS 1996 AND 1997

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

required. Table 5 shows examples of the items identified.

Speededness

Another issue of interest is whether calculator use affects completion of the test. The rates of completion were lower for groups using the calculator more frequently. The more frequently examinees used calculators, the less likely they were to finish the test. Figure 1 illustrates the percent of students completing each of the three sections: (1) 10-item multiple choice; (2) 25-item multiple choice; and (3) 15 quantitative comparison items and 10 student produced response items.

Not surprisingly, the 10-item multiple-choice section has the highest completion rates and the section with the student produced responses has the lowest. More interesting is that the difference between calculator use from no items to most items is the largest on the 25 item multiple-choice section where there is a difference of 14 percent.

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 (on a third or more of the items) also tend to be more able than those using them less frequently. Hence, lack of time rather than lack of ability to deal with the more

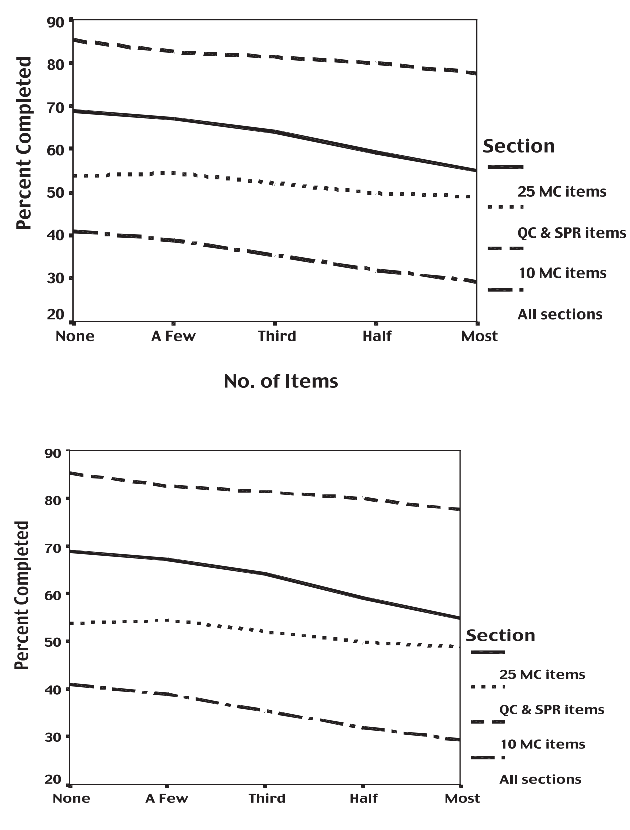

Figure 1. Percent students completing math sections by frequency of calculator use.

difficult items at the end of a section seems a more plausible explanation.

CONCLUSIONS

This study showed that almost all of the students brought calculators to the administration of the SAT I test in November of 1996 and 1997 and continue to do so today. The number who actually used the calculators on the test and the extent to which they used them varied across students. Performance on the math sections of the exam was associated with the extent of calculator use, with those using calculators on about a third to a half of the items averaging higher scores than those using calculators more or less frequently. Performance differences associated with type of calculator used were also observed.

These relationships, however, appear more likely to have been the result of able students using calculators differently than less able students rather than calculator use per se. Of the calculator variables, only the use of a graphing calculator on the test and frequency of use of calculators outside of the testing situation were found to be independently associated with prediction of math scores. Addition of other variables into a series of regression models reduced this relationship considerably, but some small independent contribution of these calculator variables to variance in score remained.

It is unclear why the effects of the calculator variables did not essentially disappear with the inclusion of the other variables. Some minimal effect of calculator use outside the testing situation and hence familiarity and comfort with the calculator seems a plausible explanation for that variable. The use of graphing calculators may affect the way that students approach and think about problems that is advantageous in taking the test. Possible differences in students' approaches to problems is an area where further investigation would be of interest. If such differences are found, teachers might then consider how the differential strategies could be taught.

The frequency of use of calculators on the test was found to be associated with DIF and with speededness of the test. In the DIF analyses, items favoring both use of the calculator on most items and use on few or no items were found. As might be expected, calculator use was favored with items requiring computations while nonuse of the calculator was favored on items that emphasized reasoning with little computation required. Although the majority of students did complete the individual test sections, those using calculators more frequently were clearly finishing less often, with the largest effect on the multiple-choice items.

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 the current 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.

The authors are Janice Scheuneman, an independent consultant in assessment, and Wayne J. Camara, vice president of research and development at the College Board.

For a more complete description of these issues and this study, see Calculator Access, Use, and Type in Relation to Performance on the SAT I: Reasoning Test in Mathematics *by J.D. Scheuneman, W.J. Camara, A.S. Cascallar, C. Wendler, and I. Lawrence (*Applied Measurement in Education, 15*, pp. 95-112).*

**For more information or additional copies of this report, please write to:
Office of Research, The College Board, 45 Columbus Avenue, New York, NY 10023-6992,
or visit the Research Library on our Web site at: www.collegeboard.com.**

5/02