Calculator Use on Stanford Series Mathematics Tests

Thomas E. Brooks, Ph.D.
Betsy J. Case, Ph.D.
Tracy Cerrillo, Ph.D.
Nancy Severance
Nathan Wall
Michael J. Young, Ph.D.

May 2003 (Revision 1, November 2003)
Abstract

Calculator use has, over the years, become increasingly integrated into mathematics instruction and testing. At the same time, a curricular emphasis on mathematics problem solving over computation has increased. As it relates to standardized testing today, the question always asked is: “Do students who use calculators during testing have a differential advantage over those who do not use calculators?” In light of the trend in schools away from computation and toward richer mathematics problem solving activities, the potential advantages or disadvantages to using calculators for computation purposes in mathematics applications and problem solving situations become increasingly insignificant.

Beginning in 1988, Pearson began to study the use of calculators and its effect on test performance in grades 4 through 12. These studies were conducted concurrently with the standardization of three different editions of the Stanford Achievement Test Series and the Otis-Lennon School Ability Test®.

For Stanford 8, small but significant differences were found in student performance when the results of calculator users were compared to those of non-users. For Stanford 9 and Stanford 10, no significant differences were found. The results confirm that performance differences between calculator users and non-users were not large enough to warrant the development of separate score conversion tables.
Background

Student use of calculators in grades 4 through 12 has been an important issue for mathematics educators since calculators became widely available to students in the late 1970s. Before that, mathematics was taught with much of the emphasis on manual computation. With the arrival of calculators and various reform initiatives, more mathematics educators began to introduce the use of calculators in the classroom and to introduce and increase the emphasis on applications and problem-solving processes. Being freed of the labors of manual computation, it was expected that students would be able to focus on problem solving (Poe, Johnson, and Barkanic, 2000).

A key factor influencing the use of calculators in the classroom was the National Council of Teachers of Mathematics (NCTM) recommendation that mathematics programs take full advantage of the power of calculators (NCTM, 1980 as quoted in Poe et al.).

More recently, NCTM recommended the integration of calculators into school mathematics programs at all levels (NCTM, 1998, reformatted Oct. 2002). NCTM (2002) explains its rationale as follows:

Research and experience support the potential for appropriate calculator use to enhance the learning and teaching of mathematics. Calculator use has been shown to enhance cognitive gains in areas that include number sense, conceptual development, and visualization. Such gains can empower and motivate all teachers and students to engage in richer problem-solving activities.

Appropriate instruction that includes calculators can extend students’ understanding of mathematics and will allow all students access to rich problem-solving experiences. Such instruction must develop students’ ability to know how and when to use a calculator. Skill in estimation, both numerical and graphical, and the ability to determine if a solution is reasonable are essential elements for the effective use of calculators.

Assessment and evaluation must be aligned with classroom uses of calculators. Instruments designed to assess students’ mathematical understanding and application must acknowledge students’ access to, and use of, calculators.
A Review of the Literature on Calculator Use

The findings of a meta-analysis of the effects of calculator use in high school mathematics programs were reported by Hembree and Dessart (1992) in Dessart, McRidder, and Ellington (1999). The meta-analysis analyzed the results from 88 studies focused on students’ achievement and attitude. Each of the 88 studies involved the comparison of one group using calculators and another group having no access to calculators. Hembree and Dessart concluded that the calculator use did not hinder the acquisition of students’ conceptual knowledge and that it significantly improved their attitude and self-concept concerning mathematics.

Also, as reported in Dessart et al. (1999), Smith (1997) conducted a meta-analysis extending the study of Hembree and Dessart (1992). Twenty-four studies conducted between 1984 and 1995 were analyzed. As in the Hembree and Dessart study, test results of students using calculators were compared to those of students not using calculators. Smith found that calculator use had a positive effect on increasing conceptual knowledge and increased students’ ability in problem solving and computation. The effect was evident through all grades and statistically significant for students in grades 3, 7 through 10, and 12.

In 1996, ACT, Inc. stated in a June 14 news release that the time was right to allow students to start using calculators on its College Entrance Examination, beginning in the fall of 1996. According to ACT, “The decision to allow calculators recognizes, in part, their widespread use in today’s classroom.”

In July 2001, the National Center for Educational Statistics (NCES) reported on certain aspects of the National Assessment of Educational Progress (NAEP) 2000 Mathematics Assessment, during which students responded to a questionnaire. Students were asked to report on their frequency of calculator use for classwork. These responses from students in grades 4, 8, and 12 were then compared to the average mathematics scores earned. The major finding was that:

At grade 4 more frequent calculator use was associated with lower scores, while at grades 8 and 12 the opposite was generally true—students who said they used calculators more often tended to score higher than their peers who reported using calculators less frequently (NCES, 2001).

The findings are depicted graphically on the following page in Figure 1.
Since the use of calculators in classrooms has become so widespread, there has been a
groundswell of interest in having calculators available to students taking standardized
tests. Although standardized norm-referenced tests have always been constructed so
that tests of mathematics problem solving or concepts could be successfully completed
using only relatively simple manual computations, it is desirable that students be able
to use calculators in testing situations where appropriate. But their introduction into
standardized testing raised issues about the extent to which students using calculators
may have an advantage over those who do not, and the extent that the availability of
calculators impinges on the needs to maintain standardized testing conditions.
Pearson takes great care to not give one group an advantage over another.

Accordingly, Pearson studied the effects of calculator use on the performance of
students on the mathematics tests of the last three editions of the *Stanford Achievement
Test* Series (Stanford 8, Stanford 9, and Stanford 10). The results, while illuminating
the effects of calculator use under standardized testing conditions, were also used to
determine if any adjustments to mathematics norms were warranted for students who
did or did not use calculators.

**Stanford 8: 1988 and 1991 Calculator Studies**

**Overview**

Pearson conducted the first calculator studies as part of the 1988 and 1991 National
Standardization Programs accompanying Stanford 8. The purpose was to investigate
possible differences in performance between students who used calculators while
taking the mathematics test and those who did not use calculators, and to determine if different derived score conversion tables (i.e., norms) would be required for these two testing conditions.

In 1988, a separate calculator mathematics sample was used to conduct a calculator study in conjunction with the Stanford 8 Spring National Standardization Program. Then in the spring and fall of 1991, the original two forms of Stanford 8 were re-standardized and two additional forms were standardized. An accompanying follow-up calculator study (again, involving a separate calculator mathematics sample) took place in conjunction with the Stanford 8 Spring National Standardization Program.

With the exception of the calculator studies, the national standardization samples to which Stanford 8 was administered in 1988 and 1991 were not permitted to use calculators while taking the mathematics subtests. Each calculator study involved the administration of one of the eight levels of the Stanford 8 mathematics test, covering grades 4 through 11/12*, together with the Otis-Lennon School Ability Test® Sixth Edition (OLSAT® 6).

The 1988 and 1991 calculator studies involved eight test levels of the Stanford 8 mathematics subtests, covering grades 4 through 11/12. The mathematics test for the elementary and middle grades was made up of three subtests: Mathematics Computation, Concepts of Number, and Mathematics Applications. Calculator use was not allowed on the Mathematics Computation subtest because students’ ability to perform manual computations was the construct being tested. The Concepts of Number subtest did not require any calculations, so the use of calculators was unnecessary. The Mathematics Applications subtest required students to perform calculations to solve problems. Students in the high school grades are assessed by a single Mathematics subtest, which also requires students to perform calculations to solve problems. Therefore, the 1988 and 1991 calculator studies involved only the Mathematics Applications subtest for students in the elementary and middle grades and the Mathematics subtest for students in the high school grades.

The approach taken for the 1988 and 1991 calculator studies was to match, by ability as measured by OLSAT 6 total score, the calculator mathematics samples to the national standardization samples. This would reveal the effect, if any, of calculator use on students’ performance on the Mathematics Applications or Mathematics subtests.

* The highest Stanford test level is designed to be administered to grade 11 or grade 12.
Results

Figure 2** presents a comparison between the spring 1988 national standardization sample and the 1988 calculator mathematics sample in terms of mean raw scores on the Mathematics Applications or Mathematics subtests, as applicable to the grade.

Figure 2. Mean raw scores for students in grades 4 through 11/12 taking the Stanford 8 Mathematics Applications and Mathematics subtest in 1988. The “Calculators Used” group is the 1988 calculator mathematics sample. The “Calculators Not Used” group is the 1988 national standardization sample.

In the 1988 study (Figure 2), the performance differences in grades 4, 5, and 6 were negligible. However, the differences in grades 7 and 8 were greater than one raw score point. In the high school samples, the differences were even greater, with the grade 9 difference being greater than one raw score point, and the difference at grade 11/12 being greater than four raw score points.

Figure 3 on the following page presents the same comparison for the 1991 calculator study. In both studies, the calculator mathematics sample consistently outperformed the national standardization sample on the Mathematics Applications or Mathematics subtests.

** The highest Stanford test level, described as 11/12 in the text, is shown as 11 on Figures 2, 3, and 4.
ASSESSMENT REPORT

Calculator Use on Stanford Series Mathematics Tests

Figure 3. Mean raw scores for students in grades 4 through 11/12 taking the Stanford 8 Mathematics Applications and Mathematics subtest in 1991. The “Calculators Used” group is the 1991 calculator mathematics sample. The “Calculators Not Used” group is the 1991 national standardization sample.

In the 1991 study (Figure 3), the differences were larger and more consistent, ranging from one and one-half to two and one-half points in grades 4 through 8, and from about two and one-half to more than six and one-half points in grades 9 through 11/12.

On the basis of these data, Pearson concluded that the effect of calculator use on performance was large enough to warrant an adjustment of scores, depending upon whether a student did or did not use a calculator on the Mathematics Applications or Mathematics subtest. These adjustments were made through the raw score to scale score conversion tables presented in the Stanford 8 spring and fall Multilevel Norms books.

Stanford 9: 1995 Calculator Study

Overview

Lenke (personal communication, 2002) summarized the calculator study conducted in 1995 as part of the Stanford 9 National Standardization Program. The study involved eight levels of the Stanford 9 mathematics subtests, which covered grades 4 through 11/12. Once again the purpose of the calculator study was to investigate possible differences in performance between students who used calculators and those who did not, and to determine if different score conversion tables would be required.
The Stanford 9 mathematics subtests were developed to align with the National Council of Teachers of Mathematics (NCTM) *Curriculum and Evaluation Standards for School Mathematics*. The Stanford 9 mathematics test for the elementary and middle grades is made up of two subtests—Mathematics: Procedures, and Mathematics: Problem Solving. Calculator use was not allowed on the Mathematics: Procedures subtest because students’ ability to perform manual computations was the construct being tested. The Mathematics: Problem Solving subtest requires students to perform calculations to solve problems. Students in the high school grades are assessed by a single Mathematics subtest, which also requires students to perform calculations to solve problems. Therefore, the 1995 calculator study involved only the Mathematics: Problem Solving subtest for students in the elementary and middle grades and the Mathematics subtest for students in the high school grades. Pearson provided four-function calculators for this research to the students in the calculator mathematics sample.

Each school district invited to participate in the Stanford 9 Spring National Standardization Program completed a questionnaire asking them to indicate, among other things, whether calculators were normally used in mathematics instruction. The calculator mathematics sample was selected from the group of districts indicating that calculators were regularly used. The participating districts were then provided with directions, materials, and calculators for the 1995 calculator study. Those districts not selected to be part of the calculator mathematics sample were instructed not to allow the use of calculators during the administration of Stanford 9. All students in the 1995 calculator mathematics sample and the National Standardization Program were administered Stanford 9 and the *Otis-Lennon School Ability Test*, Seventh Edition (OLSAT® 7).

As with the earlier studies, matching was done to statistically adjust for differences that might exist between the group using calculators and the group not using them. The 1995 Calculator Mathematics Sample was matched to the national standardization sample using OLSAT 7 results and the demographic characteristics of the 1995 national standardization sample.

In grades 4 through 8, approximately 1,000 students per grade used calculators when taking the Mathematics: Problem Solving subtest; in grades 9 through 12, about 1,000 students per grade took the Mathematics subtest using calculators. The remaining Stanford 9 and OLSAT 7 subtests were administered in the usual fashion.

**Results**

The impact of calculator use was investigated along two main analysis variables: (1) raw scores, and (2) item susceptibility.

**Raw Scores**

A comparison of the raw score distributions for the calculator mathematics sample and the national standardization sample revealed no consistent pattern of differences. The overall difference between students using calculators and students not using calculators across grades 4 through 12 and Forms S and T of Stanford 9 did not
ASSESSMENT REPORT
Calculator Use on Stanford Series Mathematics Tests

indicate any particular advantage or disadvantage for the group using calculators. In fact, only minor differences in mean raw scores were detected; sometimes these differences were in favor of the non-calculator group and sometimes in favor of the calculator group.

Item Susceptibility
Mathematics content experts determined in advance of the study those Stanford 9 mathematics items most susceptible to being influenced by the use of a calculator. Investigation of item statistics (post-calculator study) for these items revealed no discernible performance differences between the calculator and non-calculator groups. In fact, the pattern appeared to be random with respect to which group scored better on the items. This random pattern was consistent with the conclusion that calculator use had no effect on group performance.

Based on the results of the 1995 Stanford 9 calculator study, Pearson determined that no significant differences existed between the calculator and non-calculator groups. As such, it was not a matter of concern if students used calculators or not on the Mathematics: Problem Solving subtest at grades 4 through 8 or on the Mathematics subtest at grades 9 through 11/12.

Because of these results, Pearson did not publish separate Stanford 9 score conversion tables for students using calculators on the Mathematics: Problem Solving or Mathematics subtests. Therefore, it was deemed appropriate to use the single published set of Stanford 9 norms for all students, calculator users and non-users alike.

Stanford 10: 2002 Calculator Study

Overview
For the 2002 Stanford 10 calculator study, Pearson repeated the research strategy described above to determine if the use of calculators on the Mathematics Problem Solving or Mathematics subtests had a significant differential impact on student performance, thereby requiring the development of separate score conversion tables. We conducted another study for two major reasons.

First, because the use of calculators in K through 12 instruction has grown dramatically since the 1995 study, Pearson wanted to again examine the relationship between calculator use and achievement, especially since a new edition of Stanford was being standardized.

Second, because Stanford 10 has flexible administration times (e.g., no strict time limits) and current mathematics instructional practices continue to include more problem solving and processing, the use of calculators is not nearly as important as it was earlier since speed of computation is no longer a parameter of the construct being measured.

The National Standardization Program for Stanford 10 took place during the spring and fall of 2002. Approximately 250,000 students from 650 school districts across the nation participated in spring standardization. Fall standardization involved
approximately 110,000 students. All students were also administered the Otis-Lennon School Ability Test®, Eighth Edition (OLSAT® 8).

During the spring and fall 2002 standardization programs, approximately 1,000 students per grade taking the Mathematics Problem Solving subtest at the Intermediate 1 (grade 4) through Advanced 2 (grade 8) levels used calculators. In addition, approximately 1,000 students per grade taking the Mathematics subtest at the TASK 1 (grade 9) through TASK 3 (grade 11/12) levels used calculators. As with the Stanford 9 study, the students in the Stanford 10 calculator mathematics sample were selected from districts that reported that their students had access to calculators in their daily instructional environment. Matching was again done using OLSAT 8 results and the demographic characteristics of the 2002 national standardization sample.

Results

The impact of calculator use in the 2002 Stanford 10 calculator study was investigated by examining differences in mean raw scores for the calculator mathematics sample and the national standardization sample. Figure 4 shows the mean raw scores for each group. The analysis controlled for demographic variables such as gender, ethnicity, urbanicity, region of the country, and ability level. As can be seen in Figure 4, the differences in performance on Stanford 10 at all grades studied were very small. The insignificance of these differences was confirmed when the raw score to scaled score conversion tables prepared for the separate samples of calculator and non-calculator users were found to be nearly identical.

![Figure 4. Mean raw scores for students in grades 4 through 11/12 taking the Stanford 10 Mathematics Problem Solving and Mathematics subtests in 2002. The “Calculators Used” group is the 2002 calculator mathematics sample. The “Calculators Not Used” group is the 2002 national standardization sample.](image-url)
Based on this investigation, Pearson has determined that no significant differences exist between the calculator and non-calculator groups on Stanford 10. Therefore, we have not published a separate set of Stanford 10 score conversion tables for students who use calculators.

**Conclusion**

The results of Pearson’s calculator studies conducted over the past 15 years confirm that performance differences between calculator users and non-users on the Stanford Achievement Test Series have decreased since the first study was conducted in 1988. As was the case for Stanford 9, observed score differences between calculator users and non-users on the latest edition of the Stanford Series—Stanford 10—are not large enough to warrant the development of separate score conversion tables.

**References**


Calculators to be used on ACT college-entrance test. (1996, June 14). *News from ACT* (news release).


