A RESEARCH PERSPECTIVE ON USING GRAPHING CALCULATOR INTERVENTIONS TO IMPROVE MATHEMATICS ACHIEVEMENT

Jeremy Roschelle and Larry Gallagher, SRI International Contact: Jeremy.Roschelle@sri.com

SRI Report P11961.560 November 28, 2005

ABSTRACT

Educators and policy makers are looking for interventions to improve students' mathematics achievement, particularly in middle and high schools. A consistent association between frequent calculator use and higher average scores on the National Assessment of Educational Progress suggests that calculators may be a good basis for the design of technology-enhanced interventions at these grade levels. Consideration of experimental data from an inclusive set of high quality studies supports the idea that graphing calculator use can lead to student achievement gains, particular when calculators are allowed on the test.

We argue that allowing calculators on the test makes sense when the goal is to estimate students' performance on more complex, real-world tasks because technology tends to be allowed in the matching real-world context. It does not make sense when educators want to measure students' ability to perform mental or paper-and-pencil arithmetic computations, which is typically a goal in elementary school.

We emphasize that technology itself is not an intervention; technologies enable robust student learning gains only when integrated with curriculum, instruction, assessment, and teacher professional development. Because such integration must be performed locally to fit locally-controlled American schools, we call upon states and large districts to support the design of integrated, research-based interventions and to partner with universities and professional evaluators to determine which designs work best in their settings.

A common complaint about technology in education is that it is "oversold and underused" (Cuban, 2001). Indeed, a preponderance of research on technology use shows little or no effects when technology is used infrequently, poorly supported with professional development, badly aligned with assessment measures, etc (Means & Haertel, 2004). We have plenty of evidence to conclude that technology only produces effects when integrated with curriculum, instruction, assessment, and teacher professional development. The meaningful unit of adoption for schools to consider is a technology-enhanced (or technology-based) intervention, not a technology.

As schools work towards technology-enhanced interventions to address their concerns with mathematics achievement, why not start with technology that is little noticed and often well-used? While the mainstream media has focused much attention on school use of laptops and the Internet, students have quietly purchased over 25 million graphing calculators. In mathematics, graphing products are integrated in national and state standards (e.g., National Council of Teachers of Mathematics, 2000; Texas Legislature, 1998) and supported in some curricula. Best practices of instruction are well-documented (Burrill *et al.*, 2002) and teacher professional development offerings are widely available. What happens when technology is meaningfully integrated into classrooms?

For several years running, results on the National Assessment of Educational Progress (NAEP) have shown a correlation between frequent use of calculators and performance at the proficient and advanced levels. As the Nation's Report Card (National Center for Education Statistics, 2001) stated:

Student performance at grade 4 showed no significant relationship to teachers' reports of calculator use—regardless of its frequency, instruction provided, or the degree of restriction placed on its use. At grade 8, however, a mostly positive relationship was evident between students' average scores and teachers' reports on calculator use. Eighth-graders whose teachers reported that calculators were used almost every day scored highest. Weekly use was also associated with higher average scores than less frequent use. In addition, teachers who permitted unrestricted use of calculators and those who permitted calculator use on tests had eighth-graders with higher average scores than did teachers who did not indicate such use of calculators in their classrooms.

The *No Child Left Behind Act* (2001) requires that states set and meet challenging benchmarks for improving achievement for all subgroups of students; it is a national goal that all students achieve the "proficient" or "advanced" level of performance by 2014. Even though a correlation between frequent calculator use and performance at these levels is consistently present in NAEP datasets from 1996, 2000, 2003 and 2005, correlational evidence must be treated carefully. On the one hand, this correlation is consistent with the idea that technology use has a relationship to increased achievement. On the other hand, it may be that in those classrooms where calculators are allowed, smart math students are more likely to buy and use a calculator. Alternatively, it may be that teachers who are more skilled and confident of their skills are more likely to permit

or require calculator use, and it is their skill that actually translates to increased math proficiency in their students. Calculator use thus might be an indicator, not a cause, of math proficiency.

In this article, we take a deeper look. We consider a wider range of data sources and types of evidence that bear on the links between the conditions of calculator use and student achievement. An analysis across multiple experimental studies suggests that calculator use has a reliable, positive effect on achievement. This effect varies with testing policies. To tease out this factor, we consider some studies on the effects of allowing calculators on a test. Although the evidence is not conclusive, a reasonable argument can be made that allowing calculators on the test harms neither the students nor the validity of the tests. Further, tests are meant to predict real-life performance and most everyday situations allow people to use technology while doing mathematics.

For mathematics in high school (and some of middle school), a strong research-based case can be made for further exploring intensive use of graphing calculators to increase achievement. Because studies do not conclusively identify the complete intervention package (which must include curriculum, instruction, and teacher professional development), the next step is to assemble and systematically study interventions to target particular learning needs. Because local variation and control of curriculum is the norm in America, it makes sense for this step to be taken by states and larger districts. States and districts could partner with local universities and professional evaluators to generate their own data on which *interventions* (which would be locally-meaningful integrations of graphing technology with curriculum, professional development, instruction and assessment) produce the best achievement gains in their local settings.

Over the past two decades, hundreds of research studies have investigated the impact of calculator technology on instruction, learning, and assessment. Synthesizing this literature into a simple message is difficult, given the diversity of settings, grade levels, applications, sample sizes, and assessed outcomes examined in the research. Even secondary syntheses of the literature vary in scope and quality. If researchers are allowed to pick and choose studies willy-nilly, almost any conclusion can be supported. Thus only literature reviews that have transparent study inclusion criteria, well-defined variables, and rigorous designs should be considered in high-stakes policy decisions.

Secondly, when many studies are available which meet inclusion criteria and share variables, a better estimate of the "true" effect possible with an intervention can be obtained by looking across studies. The technical process for doing this correctly is termed "meta-analysis." A meta-analysis is method of estimating the true effect size due to an intervention by statistically aggregating results across multiple, independent studies¹. Rather than look at individual studies here, we have chosen to focus on interpreting the best meta-analysis we could find.

Of the literature reviews we evaluated (Burrill *et al.*, 2002; Hembree & Dessart, 1986; Texas Instruments Incorporated, 2002, 2003, among others), the meta-analysis by Aimee J. Ellington (2003) was the most thorough, up-to-date, and transparent. In particular, her focus on more recent innovations such as the graphing calculator is highly relevant to policy-makers. Furthermore, she analyzed her primary sources along several dimensions, including:

- Mathematical skills: Which sorts of skills (operational, conceptual, or problem-solving) are seen changing as a result of calculator use?
- Test instruments: How do results from standardized tests compare with teachergenerated tests?
- Grade Range: Does the effect of calculator use vary by grade range (elementary, middle, and high school)?
- Calculator type: Do results differ when a scientific versus a graphing calculator is used?
- Student ability: Do calculators improve achievement for low, middle, and high ability students?

This meta-analysis provides more reliable evidence of calculator use *causing* changes in achievement levels for two reasons. First, 44 of the 54 studies Ellington synthesized are randomized experiments. By actually manipulating who had exposure to the experimental calculator treatment in a random fashion, one is justified in attributing resulting differences in outcomes to the effect of the calculator use. This goes well beyond

_

¹ For a readable introduction to the practice of meta-analysis, see Lipsey, M., & Wilson, D. (2001). Practical meta-analysis. Thousand Oaks, CA: Sage.

correlational studies such as NAEP. When students are randomly assigned condition, the charge that only "special" kinds of students can benefit from calculator use becomes moot; high-achieving and other groups of students are just as likely to be assigned to an intervention condition.

Second, no single study—no matter how well designed—should be used as the sole estimate of an instructional effect. Sometimes studies go badly for reasons beyond the control of the experimenter (teachers becoming ill, for example) and the resulting effects are diminished. Other studies may have fortuitous coincidences boosting the observed effect (for example, all of the calculator-using teachers turn out to have strong technology backgrounds). By combining the results of several studies, we can estimate a true overall effect. Further, this ensures that we focus on the reported effects that have been shown to be replicable over multiple studies.

We carefully analyzed Ellington's tables and graphed the data to more directly address policy-makers' concerns.

Ellington's Findings

The Ellington meta-analysis reports results in terms of "effect sizes." This is a technical term that reflects the difference between two experimental groups in terms of the score standard deviations. An easier-to-interpret reporting metric is the "relative percentile" statistic. This number, on a scale from 1 to 99, indicates the percentile rank an average student (at the 50th percentile) would be expected to obtain as a result of the experimental treatment. A number greater than 50 indicates that the student has progressed relative to control students, while a number lower than 50 indicates that a student has gotten worse. If no number is reported, either there were no studies testing this particular combination of circumstances or the effects could not be distinguished.

The first effect Ellington examined was the use of calculators on tests of mathematics achievement. She grouped all of the experimental studies into two groups: (1) those that used a calculator as an intervention but did not allow its use on the test and (2) those that did allow calculator use on the test. As can be seen in Figure 1, when calculators were not allowed on the test, only tests of operational skills showed an improvement from learning with calculators (relative percentile: 57). This is an important finding, as one important worry is that calculator use might diminish a student's ability to perform operations without a calculator. This summary of studies suggests otherwise.

When calculators are allowed on the test, all four examined types of abilities were found to improve overall, at relative percentiles above 60. Thus, learning with a calculator contributes broadly to student achievement as measured on tests that allow calculator use.

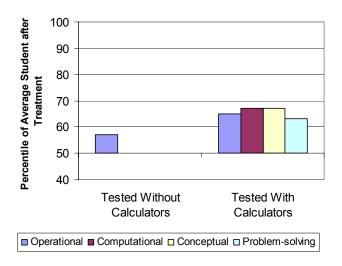


Figure 1. Skill Percentiles by User of Calculators in Testing

The additional results reported below only pertain to studies that allowed calculator use on the final test.

One may wonder whether these studies are biased by using "home grown" tests that may subtly favor calculator-based instruction. Fortunately, nine of the 25 studies testing operational skills used standardized test instruments and, as Figure 2 shows, standardized tests favor the calculator-using conditions with a relative percentile of 63. When experimenter-defined tests are used, as expected, the effects on operations tests are larger and experimenter-defined tests of conceptual skills show an improvement as well.

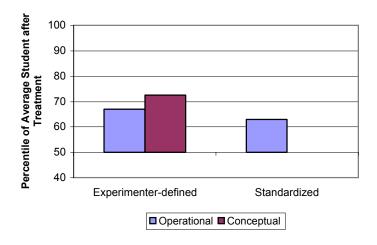


Figure 2. Skill Percentiles by Type of Test (tested with calculators)

Another concern among policy makers is that elementary school students would suffer harm from use of calculators in instruction. Unfortunately, only one study in Ellington's meta-analysis examined operational skills in elementary grades but it did show a statistically significant effect, and was deemed strong enough to report. The two studies examining conceptual growth in elementary school had inconclusive results. At the middle

5

and high school levels, however, there are significant positive effects of calculator inclusion in tests of both operational and conceptual skills (Figure 3).

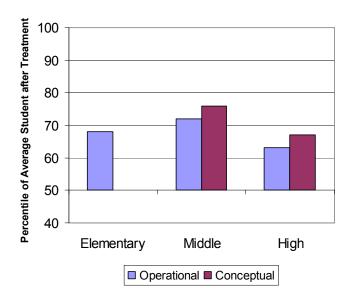


Figure 3. Skill Percentiles by Grade Range (tested with calculators)

Does the type of calculator used matter? Positive results in operational and problem-solving skills were found with both basic/scientific calculators and graphing calculators (Figure 4). Overall, improvement in conceptual understanding were only found in studies that employed graphing calculators.

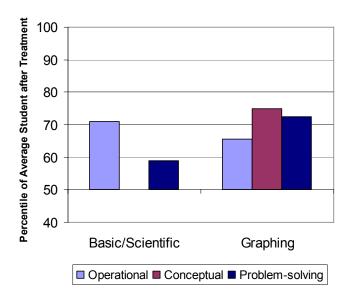


Figure 4. Skill Percentiles by Calculator Type (tested with calculators)

Last, Ellington examined the ability range of students studied. Perhaps calculators only help particular segments of the student body? The evidence is incomplete. Only two

studies purported to specifically examine low-ability students and the results were inconclusive. Similarly, only a handful of studies specifically examined high ability students, and of those, only operational skills showed reliable enough improvement to warrant reporting. The vast majority of studies examined mixed-ability youth and, as reported above, the results were consistently positive for tests of operational, conceptual, and problem-solving skills.

Based on the results of this meta-analysis, we believe the following conclusions are supported by high-quality research studies:

- An inclusive collection of well-designed experimental and quasi-experimental studies shows reliable positive effects of calculator use on a variety of mathematical skills.
 - Some of this effect is likely due to advantages gained while using a calculator on test (see the 6 studies that used "test only" designs).
 - Some of this effect is also likely due to the use of calculators in instruction (see the 15 studies that did not allow calculator use on tests).
 - We do not have sufficient information (not is it logically advisable) to distinguish out the "pure" effect of calculator introduction from the combination of calculator use with curricular and pedagogical innovation.
- 2. There were no groupings of individual studies where there were reliable negative effects of calculator use (e.g., there were no findings of harm).

Comments on Testing Policy

The most daunting conceptual question of the Ellington meta-analysis concerns all of the special groupings of studies (Figure 2 through 4) where calculators were allowed on the experimental test. As Ellington puts it,

When treatment groups had access to calculators during posttreatment evaluations, the studies were used to evaluate the calculator's role in the extension of student mathematical skill abilities after treatment was concluded. (Ellington, 2003, p. 437).

That is, these studies really ask the question of what students are capable of doing mathematically with technological support, which is a subtly different question than asking about pure mathematics knowledge.

There are three possible explanations for the improved outcomes when calculators are used on tests:

- 1. Having a calculator available on the test produces an artificially high reading; the students' underlying proficiency is actually lower than that reported by the test.
- 2. The alignment between the instructional and testing situation raises scores; students' proficiency is more fairly tested when they can use everyday technology on a test.

3. The focus of instruction changes when calculators are going to be incorporated into test taking and this influences the effect size.

Ellington provides some evidence supporting the first proposition. Six of the studies reviewed were designed to examine the effects of calculator use on test taking alone, without any special calculator-based instruction. There was an overall significant effect, with a relative percentile of 61. On the other hand, the impact on test performance cannot be the entire story, as Ellington also cites the 15 studies that combine to show an effect when calculators are *not* allowed on the posttest.

In favor of considering the second possibility, we note that in "real-life" situations workers, scientists, and citizens are almost always allowed to use a calculator. Since, the everyday performance context for mathematics is one in which technology is allowed, it makes some sense for an assessment policy to follow suit. Further, a study by Russell & Haney (1997) shows that students who are accustomed to writing with a computer score much better when they are allowed to write with a computer on the test. Writing, like solving mathematical problems at the proficient or advanced level, requires an integration of skills. When the goal of testing is to determine how students will perform on more complex tasks in more realistic settings, allowing students to use familiar technology may result in a more accurate estimate of their capabilities.

In favor of considering the third possibility, it is well known that there are several canonical approaches to making tests either calculator-sensitive (making problems that are less 'textbook' and more 'real world' so that calculators are essential) or calculator-neutral (reframing questions so that the task is generative—'give an example of...'; requiring solutions in step-by-step, symbolic form). Each of these requires a different instructional strategy and corresponding student practice. It could be that changes in instruction that occur when teachers frequently use calculators causes increased scores on tests.

When considering testing policy, educators will want to consider what is known about how students who use calculators perform on national and international comparison tests, such as the National Assessment of Educational Progress (NAEP) and the Trends in Mathematics and Science Study (TIMSS). Loveless and Coughlan (2004) have argued argue that at the 4th grade level, calculators may harm student learning of arithmetic operations and artificially increase test scores. Certainly, this should caution policy makers against overusing calculators while students are still learning arithmetic. At the 8th grade and higher, levels at which a broader range of concepts and skills are important, research shows consistently higher NAEP achievement scores associated with self-reports of calculator use (National Center for Education Statistics, 2001; Sowder *et al.*, in press). Further, we found that the association between frequent calculator use and high performance is holds when one looks only at low SES students.

When we analyzed these results across five different states (CA, MA, MI, OH, and TX), we found that the level of calculator use does vary across states. For example, more than 80% of students report using a calculator at least weekly in Michigan, while only around 50% of students report such frequent use in Ohio and Texas. Nonetheless, in each of these states, students using a calculator frequently for math class outscored their non-

calculator using peers. While NAEP data cannot be used for causal attributions, these results serve an existence proof that use of calculators on school-based tests does not prevent students from performing at the proficient and advanced level on a later test.

From the TIMSS data, analysts have shown how the relationship between calculator use and achievement varies by country. For example, while US data shows a positive association between calculator use and achievement, in Japan the results are neutral or significantly negative. The authors of one report suspect that this is due to differences in the testing environments and cultural norms between the two countries. In particular, the high stakes university examinations in Japan do not allow calculators, while in the US calculators may be required (Lennex *et al.*, 2000). This reinforces our notion that the calculator is never a stand-alone, uniform intervention; rather, it interacts with curricular, pedagogical, and cultural factors to produce an effect.

Similarly, there is no clear association between calculator use and overall mathematics achievement by country. Some high-scoring countries such as Belgium and Singapore show broad access to calculators in mathematics education, while other high-scoring countries do not (e.g., Korea and Japan). Similarly, calculator-using countries can be found among the lowest scorers, as well as non-calculator-using countries. "Although on average internationally the relationship is unclear, in most of the countries where emphasis on calculator use was high, there was a positive association between calculator use and mathematics achievement." (Mullis *et al.*, 2001, p. 225).

These data suggest that the existing body of research cannot be used to determine whether calculators should be allowed or required on tests. Policy-makers should make this decision based upon how they want to interpret test data. If the point is to measure whether students have mastered basic computations that every student should be able to perform in their head or with paper and pencil, calculators should not be allowed on the test (Loveless & Coughlan, 2004). If the point is to estimate how students will perform in more complex, more realistic situations in which technology is ordinarily allowed, technology should be allowed on the test. A state could reasonably decide on different policies at different levels of schooling. At the lower grades, states may be more concerned that a student who cannot perform mental and paper-and-pencil arithmetic does not have the prerequisite capabilities for further study. At the upper grades, states may be more concerned with anticipating how students will do on the job, on their taxes, and in science, medical and engineering studies—all environments where technology is ordinarily available.

The Next Step: Local Intervention Design and Research

As described above, the general drift of the evidence is positive, but not yet in the form of "usable knowledge" (Lagemann, 2002) that schools could employ with confidence about the degree to which it will address their particular student achievement concerns. Rather than awaiting perfect evidence to arrive in a neat package, it would be quite practical for states and districts to partner with local universities and professional evaluators to generate their own data on which interventions (which would be locally-meaningful integrations of graphing technology with curriculum, professional development, etc.) produce the best student achievement gains in their local settings. The basic capabilities of graphing calculators can be aligned with school subject matter and assessments, making detection of effects easier. The cost of the hardware is relatively low, which can make equitable participation more likely and can enable schools to "do things right:" devoting more resources to curricular integration, professional development, and alignment of assessments. Although there are appropriate cautions against intensive calculator use before students master numbers and operations, there are few big controversies that would result from broader experimentation in the upper grades graphing calculators cannot browse the web for pornography, distract students with email and instant messages, etc. Through an enlightened policy of adoption and research, a state or large district could produce a volume of focused, school-based research studies on which technology-enhanced interventions produce the achievement gains they seek in mathematics and science.

We suggest that an enlightened policy would have several components.

First, a state or district should choose to experiment with intensive graphing calculator use at the grade levels in which their primary concern is students' achievement in more realistic, complex situations. Because technology is typically available in such situations when they occur in everyday life, there is no reason to prohibit technology on the test. Allowing technology on the test will both yield a more accurate prediction of student performance in the real world and make detecting which interventions work possible with smaller samples. Further, it will signal to teachers and students that there is no reason not to take full advantage of the technology to support learning.

Second, states and districts should support the design of interventions that meet specific local needs and conditions. Research on the use of graphing calculators "in general" makes little sense when so much is determined by state standards, district curriculum choices, and the commitments of localities to particular instructional practices and assessment regimes. Further, the best interventions are likely to use technology in concert with some changes in related conditions and practices of teaching and learning, not just as an add-on to an otherwise unchanged classroom experience (Honey et al., 2000).

Fortunately, an ample research base suggests what kinds of changes can be beneficial when integrating technology into the mathematics classroom (English *et al.*, 2002; Kaput, 1992; Masalski, 2005). For example, teachers may use calculators to reduce the need for laborious calculations or to enable students to work with graphs and tables more quickly. Students may be able to complete more problems or solve more difficult ones than they could with pencil and paper. The time gained can allow students to try different

approaches to problem solving. Further, it can create time for classroom discussions about problem solving. In addition, calculators are more than a tool for speed and elimination of tedious calculation. Teachers can exploit the representational features, like graphing and exploration of larger data sets, to make mathematics more meaningful and interesting. At the same time, a well-designed intervention should take note of some precautions. Using estimation skills to check the reasonableness of answers has always been important, but it becomes even more so when calculators are used for a series of complex calculations. Understanding order of operations becomes essential and can be pedagogically explored. Scales on graphs need to be considered carefully. Teachers and school leaders will need support to work through these possibilities and design a integrated use of the capabilities of technology that makes sense given local needs, conditions, practices, and goals.

Third, we suggest that states and districts should seek to generate additional research results that are worth sharing. This requires some extra attention to documenting all aspects of the intervention carefully, so other schools can replicate all elements of the design (e.g., including the professional development). It further requires introducing an experimental methodology, which can be as simple as delaying the introduction of the intervention for a randomly selected group of half the teachers. A simple rationale can be "we can't afford to train everyone in one year, so we'll randomly select half of the classrooms to participate in the first year and the rest will join in next year." For the trouble of this complication (which is probably best undertaken in a partnership with a local university or professional evaluator), a local experiment can generate results that can contribute to further refinement of the scientific knowledge base on "what works." While it is obviously somewhat more expensive to do so, it is more sensible use of public funds than using the same funds to spread an untested intervention throughout a state or large district.

Acknowledgements

We thank Texas Instruments for their support in producing this article under SRI project P11961 with a charge to seek the best evidence and report it clearly and fairly. Any opinions found within are those of the authors.

References

- Burrill, G., Allison, J., Breaux, G., Kastberg, S., Leatham, K., & Sanchez, W. (2002). *Handheld graphing technology in secondary mathematics: Research findings and implications for classroom practice*. Dallas, TX: Texas Instruments.
- Cuban, L. (2001). Oversold and underused: Computers in the classroom. Cambridge: Harvard University Press.
- Ellington, A. J. (2003). A meta-analysis of the effects of calculators on students' achievement and attitude levels in precollege mathematics classes. *Journal for Research in Mathematics Education*, 34(5), 433-463
- English, L. D., Bussi, M. B., Jones, G. A., Lesh, R. A., & Tirosh, D. (Eds.). (2002). *Handbook of international research in mathematics education*. Mahway, NJ: Lawrence Earlbaum Associates.
- Hembree, R., & Dessart, D. J. (1986). Effects of hand-held calculators in pre-college mathematics education: A meta-analysis. *Journal for Research in Mathematics Education*, 17(2), 83-99.
- Honey, M., Kulp, K. M., & Carrigg, F. (2000). Perspectives on technology and education research: Lessons from the past and present. *Journal of Educational Computing Research*, 23(1), 5-14.
- Kaput, J. (1992). Technology and mathematics education. In D. Grouws (Ed.), *A handbook of research on mathematics teaching and learning* (pp. 515-556). New York: Macmillan.
- Lagemann, E. C. (2002). *Usable knowledge in education: A memorandum for the Spencer Foundation board of directors*. Chicago, IL: Spencer Foundation.
- Lennex, L., Mittag, K. C., Tarr, J. E., & Uekawa, K. (2000). A comparison of calculator use in eighth-grade mathematics classrooms in the United States, Japan, and Portugal: Results from the third international mathematics and science study. *School Science and Mathematics*, 100(3), 139-ff.
- Loveless, T., & Coughlan, J. (2004). The arithmetic gap. Educational Leadership, 61(5), 55-59.
- Masalski, W. J. (Ed.). (2005). Technology-supported mathematics learning environments: Sixty-seventh yearbook of the National Council of Teachers of Mathematics. Reston, VA: National Council of Teachers of Mathematics.
- Means, B., & Haertel, G. (Eds.). (2004). *Using technology evaluation to enhance student learning*. New York: Teachers College Press.
- Mullis, I. V. S., Martin, M. O., Gonzalez, E., O'Conner, K. M., Chrostowski, S. J., Gregory, K. D., et al. (2001). *Mathematics benchmarking report: Timss 1999—eighth grade*. Boston: International Study Center, Boston College.
- National Center for Education Statistics. (2001). *The nation's report card: Mathematics 2000*. (No. NCES 2001-571). Washington DC: U.S. Department of Education.
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.
- Russell, M., & Haney, W. (1997). Testing writing on computers: An experiment comparing student performance on tests conducted via computer and via paper-and-pencil. *Educational Policy Analysis*, 5(3).
- Sowder, J. T., Wearne, D., Martin, G., & Strutchens, M. (in press). What grade 8 students know about mathematics: Changes over a decade. In P. Kloosterman & F. Lester (Eds.), *NAEP mathematics assessments: Interpretive analyses and materials development*. Reston, VA: NCTM.
- Texas Instruments Incorporated. (2002). Handheld technology and student achievement: A collection of publications (An Independent Study Conducted by The McKenzie Group). Dallas, TX: Texas Instruments.
- Texas Instruments Incorporated. (2003). *Using handheld graphing technology in secondary mathematics:* What scientifically based research has to say. Dallas, TX: Texas Instruments.
- Texas Legislature. (1998). Texas essential knowledge and skills for mathematics. In *Texas education code* 28.002 (Vol. 111.32). Austin, TX.