Some Factors Contributing to Gender Differences in the Mathematics Performance of United States High School Students

Pamela L. Paek, Ph.D.

Charles A. Dana Center, The University of Texas at Austin, Austin, Texas, United States

Paper for presentation in the Gender and Mathematics Education Topic Study Group (TSG 32) at the eleventh International Congress on Mathematical Education, Monterrey, Mexico, July 6–13, 2008

Abstract

This paper discusses factors that contribute to **gender differences in mathematics performance.** We uncovered these factors through a study examining differences in the mathematics performance of 122 11th- and 12th-graders on a proctored administration of a retired SAT® I. Students were first administered the SAT items in paper-and-pencil form and then administered a subset of those items on the computer, which tracked and timed each step students used to solve the items. A subset of students was also interviewed for their feedback and impressions about solving problems online, a mode that forced them to show their steps.

Students' problem solving processes were analyzed using an online platform that timed and recorded the *actual steps students took*. In contrast to psychometric methods that model students' strategies by looking at different item types they answer correctly, this approach provides direct, empirical evidence of student thinking. Our results confirmed the outcomes from previous research (e.g., Gallagher 1990, 1992): females tended more than males to follow algorithms and verify their steps and the answer before moving on to the next problem. In the interviews, females articulated a greater need for verification in their work. When given an unlimited amount of time to solve a problem, females were more successful than males in achieving the solution and in attempting more problems (Paek, 2002). In contrast, on the timed paper-and-pencil assessment, the females performed worse than the males on these same items. Our findings indicate that timed assessments may inadvertently favor males over females: as seen through the online capturing of their steps, males tended to use fewer steps, and thus took less time to solve each item.

Introduction

A major challenge for teachers of mathematics is understanding what students know and what misconceptions deter them from solving problems correctly. We can infer that students who are higher achieving (as measured by grades and test scores) understand more than do students who are lower achieving, but that inference merely allows us to stratify them, not to deeply understand how they are engaging with the mathematics.

In this study, we used an online computer system to examine males' and females' problem solving processes on SAT[1] mathematics items in order to capture what happens between the presentation of an item and the student's final choice of an answer. We focused on SAT mathematics items because of the SAT's centrality and importance to college admission in the United States, and because research continues to confirm that males perform better than females on this assessment. It would appear that something in the SAT test-taking process must differ by gender, and this study sets out to investigate what those differences are.

The purposes of this study, then, are 1) to identify the strategies that high school students use when solving mathematics problems so as to better understand the processes they use and 2) to uncover some potential reasons females underperform in mathematics compared to males. This study enables a more detailed understanding of student test-taking behavior by providing a more authentic look at what students do before they choose a final answer. Ultimately, determining what enables higher-performing students to respond correctly can inform new ways of conceptualizing instruction.

Theoretical Framework

Gender Differences

Studies document that high school males tend to take more—and more advanced-level—mathematics classes than do females in the United States (American Association of University Women, 1992, 1998). The gender difference in scores on standardized tests increases with age, as shown by the Third International Mathematics and Science Study (TIMSS) and the United States' National Assessment of Educational Progress (NAEP). These two assessments were administered at the fourth-, eighth-, and twelfth-grade levels, and for each test, more males than females received top scores in mathematics, a trend most prominently seen in grade twelve (Braswell, Lutkus, Grigg, Santapau, Tay-lim, & Johnson, 2001; Mullis, Martin, Fierros, Goldberg, & Stemler, 2000).

Research on gender differences in mathematical learning and performance has established that—at least when looking at standardized test scores in the U.S. such as the SAT, NAEP, and TIMSS—males generally perform better than females in mathematics (American Association of University Women, 1992, 1998; Braswell, et al., 2001; Coley, 2001; Gallagher & Kaufman, 2005; Mullis, et al., 2000; Perie, Grigg, & Dion, 2005; Santapau, 2001). In an effort to

[1] The SAT, which long ago stood for Scholastic Aptitude Test, is also now known as the SAT® Reasoning Test. At the time this study was conducted (in the 2000-2001 academic year), the SAT was also known as the SAT®I.

explain sex differences in mathematics performance, researchers have investigated the problem solving processes used by males and females. For instance, Gallagher (1990, 1992) found that females' approaches to solving SAT mathematics problems were more algorithmic than the approaches taken by males. In Gallagher's study, males tended to use more innovative strategies—rather than algorithmic processes taught in classrooms—which accounted for their higher performance on a wider variety of items. U.S. females are more likely than males to dislike mathematics, and as a result, to use problem-solving strategies which led to lower mathematics performance, which in turn reinforced their negative attitudes toward mathematics (Gallagher & DeLisi, 1994).

Standardized Testing

Standardized assessments are widely used to measure student achievement. Many of these assessments are considered "high stakes" because they publicly rank the performance of schools and various student groups, and because individual students' future education and career options can be affected by these tests' results. In fact, high-stakes assessments are considered by some as gatekeepers for student admission to college or to specialized academic programs, since many programs and universities require students to achieve a minimum score to be considered for admission. As a result, standardized assessments are not reviewed favorably in some quarters, as reducing student knowledge into a single set of quantitative scores seems to deny the complexity of knowledge. Nonetheless, standardized assessments are the most common tools for evaluating K–12 student performance in the United States.

The SAT is a primary example of a high-stakes assessment, as it plays a central role in many U.S. college admissions processes. The SAT® I[2] consisted of a verbal and a quantitative component, each with three sections. Scores were reported in three numbers: a verbal score, a quantitative score, and a total score, which was the sum of the verbal and quantitative scores. The verbal and quantitative scores each had a minimum score of 200 and a maximum of 800, which meant the total score ranged from 400 to 1600.

Since the SAT is used to rank students applying to college, much research has been conducted to ensure the SAT items discriminate well, meaning that the items correctly discriminate between those students who show they have the knowledge and skills to answer the item correctly and those who do not (e.g., Angoff, 1974; Katz, Friedman, Bennett, & Berger, 1996; Wainer, 1983). Thus, a higher score on the SAT-I is a valid indication of the more able students (Dorans & Lawrence, 1987). There has been much debate, however, regarding the validity of performance on the SAT items as indicators of how students will perform in *college* (Atkinson, 2001; American Association of University Women, 1992, 1998; Geiser & Studley, 2001), mainly due to differential performance on the assessment by various subgroups.

[2] The SAT®I was used in this study. It differs from the currently administered SAT Reasoning test, which consists of three scores, each with a minimum score of 200 and a maximum of 800: critical reading, mathematics and writing. The SAT Reasoning Test replaced the SAT®I in 2005.

Problem Solving

When approaching mathematics problems, students rely on various resources and types of information (Chi & Glaser, 1985; Ericsson & Simon, 1993; Schoenfeld, 1988). These resources and types of information form the frameworks that students use to interpret and solve different items. Ideally, students are able to identify and interpret a problem sufficiently well to choose the correct framework for solving it, resulting in a correct response. In reality, however, many students sometimes do not know which framework to choose, or choose inappropriate frameworks, which results in inconsistent patterns of correct and incorrect responses (Marshall, 1995; Tatsuoka, 1993). Furthermore, choosing an appropriate framework does not necessarily always lead to a correct answer, because computational errors can also lead to an incorrect response (Tatsuoka, 1990). Thus, trying to explicate students' problem-solving processes requires that we look beyond their correct and incorrect answers and undertake instead a detailed, empirical investigation of how students organize information.

Every mathematics problem contains a host of concepts that can be linked to a group of specific steps that must be followed to successfully solve the problem. When students choose to follow certain steps in a particular order, they are demonstrating a pathway for organizing information to solve that problem. By assembling detailed information on students' responses to multiple problems, researchers can trace the steps that students took to solve each problem and evaluate how well their approaches worked on different types of items—at the individual student level and across various groups of students (e.g., by classrooms). Most students refine their problem-solving strategies over time—which is consistent with models of skill acquisition (Ericsson, 2004)—gradually using fewer steps and eventually settling on a preferred approach (Stevens, Soller, Cooper, & Sprang, 2004; Stevens & Soller, 2005). Researchers can analyze student self-regulation and self-monitoring of strategies by investigating these steps, which will provide a better understanding of the complexity of student problem-solving performance (Hartley & Bendixen, 2001; Song & Hill, 2007).

IMMEX: Using Technology to Study Problem Solving

Given the myriad ways that students can organize information and regulate the steps they take to solve problems, the project of truly understanding what students are thinking and doing as they work problems can seem insurmountable. Advances in technology, however, have provided one way to get inside students' heads, as it now enables us to track in detail the actual steps students take to solve problems and the time they spend on each step (Hartley & Bendixen, 2001).

One pioneering technology that enables better understandings of student thinking is the Integrated (now Interactive) Multi-Media Exercises (IMMEX) program, which draws on both case-based (Schank, 1990) and production system (Newell & Simon, 1972) models of problem solving. Such a tool gives a more qualitative look at how students solve problems, since it captures in intricate detail the variety of approaches they can take. This captured information opens the door to a deeper understanding of the comprehensive nature of student thinking because rather than analyzing only a student's final answer, this tool allows researchers to look at each step that led to that answer.

IMMEX presents a problem to be solved in a *problem space*—that is, a space with a finite set of concepts, numbers, and equations that students must combine in order to create a solution path. Within the IMMEX problem space, various drop-down menus provide pathways the students can choose from. A problem space typically will not encompass all the combinations that students could use to solve problems, but it does provide an essential preliminary view for better estimating what they are able to do (Tatsuoka, 1990, 1995). Further, although IMMEX's simulated problem spaces are finite, they do provide enough different types of information that students with diverse math backgrounds could successfully solve the problems.

Working in IMMEX, students can assess the *problem structure*—the information needed to solve the problem—and then organize a mathematical *representation*—an arrangement of that information into a series of steps that solves the problem (Bennett, Morely, & Quardt, 1998). Most students want to arrive at an answer and will follow some process to produce one. If their chosen process leads to a wrong answer, they will probably try a different process if given a chance to try again. IMMEX software allows researchers to track all the steps—forward, backward, and sideways—that students take as they attempt to solve problems. IMMEX also records and displays the sequence of steps and the time spent on each one (Ericsson & Simon, 1993; Stevens, 1991).

Method

This study analyzed—and compared across gender—the problem-solving strategies of 122 high school students on a set of SAT mathematics problems administered both on paper and in IMMEX. The researcher also interviewed a subset of students for their feedback and impressions about solving problems online, a mode that forced them to show their steps.

Participants

A total of 122 high school students (72 females and 50 males) from a single high school in Orange County, California, participated in this study. Most students were enrolled in AP Chemistry and were either juniors or seniors.

Mathematics Grades in the Studied Population

Previous studies have shown that females receive better grades than males in the mathematics classroom (Dwyer & Johnson, 1997; Kenney-Benson, Pomerantz, Ryan, & Patrick, 2006; Pomerantz, Altermatt, & Saxon, 2002). In this study, we analyzed grades for participating students in Algebra I and Geometry in the first and second semesters to see if there were differences by gender. We selected these two courses because most SAT math items are based on students' knowledge of Algebra I and Geometry. As seen in Table 1, the females tended to have higher grades than the males in the two mathematics courses. In all four semesters, females outperformed males, although the differences are not statistically significant. In addition, males exhibited more variability in grades than females, as shown in the higher standard deviation (SD) for all four semesters.

Table 1. Algebra I and Geometry grades in the study's student population

Mathematics Course	Gender	N	Mean	SD
Algebra I, first semester	Female	71	3.69	0.49
	Male	49	3.51	0.74
Algebra I, second semester	Female	71	3.64	0.58
	Male	49	3.50	0.67
Geometry, first semester	Female	71	3.64	0.57
	Male	49	3.61	0.62
Geometry, second semester	Female	71	3.62	0.58
	Male	49	3.53	0.66

Procedure

Participants first completed a retired SAT-I test from the 10 Real SATs (College Board, 1997) under standard SAT time constraints using paper and pencil. Next, students used IMMEX to solve 31 SAT-I mathematics problems that came from the same SAT-I test the students had completed for the first part of the study. Eight students also participated in a focus group to reflect on their problem solving strategies and to explain their thinking behind their steps.

Developing the IMMEX Problem Space

I created a specific problem space in IMMEX for solving each of the 31 mathematics items along with a method for using and coding the search path maps. The problem space included formulas, definitions of concepts, and a breakdown of the process for arriving at a solution. The problem space for this study was developed based on a formal task analysis the researcher had already conducted, for which students listed the steps they used to solve certain math items. Additionally, the researcher incorporated into the problem space common errors that students make in arithmetic, basic algebra, and geometry (Tatsuoka, 1990, 1995). These two elements (the formal task analysis and the incorporation of common errors) helped to determine the menus and submenus needed for the IMMEX platform in this study so that the majority of students could solve the problems using the information given (Mislevy, Yamamoto, & Anacker, 1991).

The problem space was the same for each of the 31 items in IMMEX, except for some of the submenus, which were changed to correspond with the proper substeps and the numbers and equations related to each problem. The problem space included the math concepts necessary for a correct solution as well as bugs and distractors, which were included to track where students made arithmetic errors or had misconceptions about the problem. Students could easily navigate through all the menus and submenus and still not be able to correctly solve a problem—to reach an accurate solution, they needed to know what kinds of information were pertinent and be able to order that information correctly. Within the IMMEX problem space, each problem was presented with five main menus. The problem and the five main menus were always at the top of the screen, even when students navigated through the submenus. Each main menu represented one of five math concepts: arithmetic, angles, area, perimeter, and solving equations. Clicking on one of these menus revealed a host of submenus also representing math concepts. Clicking on a submenu led to a series of equations and/or numbers. These equations/numbers were represented in expanded form, so that the student had to decide where to combine or collapse terms. The menu structure included shortcuts so that students could collapse several steps into one. For example, consider the equation $2x + 10 = 5 – 3x$. The traditional way of solving it would be first to move the numbers to one side by subtracting 10: $2x + 10 – 10 = 5 – 10 – 3x$. To get the variables on one side of the equal sign, the next move would be to add $3x$ to both sides: $2x + 3x = –5 – 3x + 3x$. Then both sides of the equation would be divided by 5: $5x/(–5) = –5/5$ to arrive at the final response of $x = –1$. In IMMEX, the menu shortcuts enabled students to collapse these three steps, computing the information in their heads so they could move to the answer in one step.

The IMMEX problem space also included common arithmetic errors students could make that were associated with the distractors offered on the paper-and-pencil test. In the problem above, for example, students could incorrectly subtract by $3x$ so that the response would be $2x – 3x = –5$, resulting in a final answer of $x = 1$. These types of mistakes were included as options in the submenus and the equation structures. The purpose of these incorrect paths was to document where students might go wrong in coming up with their final answer choice.

After completing all the work and arriving at an answer, students entered their responses after clicking the “solve problem” button. The study was structured to give students two chances to solve the problem so that they could reconsider each step they had taken and so that the researcher would have an opportunity see how students revised their steps to answer the problem correctly on the second try. The second chance also allowed students who might have made a simple arithmetic error to backtrack and correct it.

Results

Males scored about 20 points higher (*M* = 642.12, *SD* = 73.946) than did females (*M* = 621.43, *SD* = 69.372) on the paper-and-pencil SAT-I mathematics section, but this difference was not statistically significant. Note that the average SAT mathematics score in this study is more than 100 points above the national SAT mathematics average of 515 (College Board, 2007). This higher achievement may be due to the fact that at the time of the study, most (84%) of the participants were enrolled in at least one Advanced Placement course (mainly AP Chemistry).

I conducted analyses of the amount of work shown on the paper version of the SAT mathematics section and analyses of students’ problem solving on IMMEX: the number of steps students took to solve each problem correctly, the amount of time spent on each step and on each problem, and the number of attempts made at solving each problem. All these analyses revealed differences in females’ and males’ performances.

Test Booklet Analysis

Each student's test book was coded by the work they showed and the number of steps they used. The student markings on the test booklets provide insight into the different ways that students approached the problems. Since students do not mark their test books for every problem, only the items where they showed their work could be coded for analysis. Note that students were not asked to show their work; they were asked to take the test and use the materials they were given as they would in a live testing situation.

Females tended more often than males to show their work. Given that the SAT has three separate mathematics sections, we coded each section separately, as seen in Table 2. Totaling the items from all three sections, females showed their work on almost 34 of the 60 items, while males showed their work on 28 items.

Table 2. Work shown on the math section of the paper SAT

Showing Work	Gender	N	Mean	SD
Section 1	Female	69	14.22	5.67
	Male	48	11.38	6.05
Section 3	Female	67	14.60	5.51
	Male	48	12.77	6.60
Section 6	Female	69	4.93	2.76
	Male	47	4.19	2.65

IMMEX: Number of Steps Taken

Using the data IMMEX collected, the number and types of steps the students took to solve each problem were analyzed. In general, the fewer steps a student had taken in attempting to solve a problem, the more likely it was that the student had solved the problem correctly—as students who unsuccessfully completed a problem tended to take more irrelevant steps that were not helpful to solving the problem. In addition, the number of steps taken differed between males and females. On average, males took two fewer steps than did females did to solve a problem: males averaged four or fewer steps (M = 3.91, SD = 2.06), whereas females averaged six or more steps (M = 6.12, SD = 1.77). Even with these differences, females attempted more IMMEX items, took longer to solve each problem, and answered more problems correctly than males did; the reason for the higher success of females on these items appears to be that they verified their steps, not that they were inefficiently taking extra steps. On the paper-and-pencil SAT-I mathematics sections, however, females scored lower than did the males. The IMMEX test had no time constraints, so it may be that females performed better in the untimed situation than males did.

Informal interviews with the participants suggested that the females liked to be sure of their answers and would use any available information to verify them. They wanted their answers to be correct on the first attempt, and they took more steps and more time to ensure correctness. The interviews suggested that males, on the other hand, tended to be inclined toward an answer and would select it, knowing they had a second chance if they got it wrong. This method resulted in fewer steps and less time taken per problem. Males indicated they were also more likely to guess once they had eliminated some choices.

IMMEX: Amount of Time Spent per Step and per Problem

We also analyzed time spent per step and per problem. Females tended to take 2 s more per step (*M* = 19, *SD* = 21) than did males (*M* = 17, *SD* = 18), which resulted in females taking about 55 s more per problem (*M* = 2:06, *SD* = 2:51) than males (*M* = 1:11, *SD* = 1:45). These differences are statistically significant ($p < 0.01$). This difference in time spent per step, coupled with the fact that females took more steps than males did, may well help to explain why females tend to score lower on standardized mathematics tests: They are not able to complete as many problems.

IMMEX: Number of Attempts Made

Students were given two opportunities to solve each IMMEX problem, so I could document the changes they made in their steps from the first to the second attempt. The majority (61%) of students solved the problems correctly on the first try, and an additional 23% answered the problems correctly on their second try. The steps students took on the second tries for both correct and incorrect answers were analyzed. Students who correctly solved a problem on the second try demonstrated an orderly process in which they deliberately retraced their steps to verify their answers and more systematically regulated their steps, indicating that these students knew what they were doing but had made a small error in computation at some point. Students who did not solve the problem correctly on the second try, on the other hand, showed less organization and planning in their process.

Observing the processes used by the participants gives an inside view of how they regulated their learning as they solved problems, and suggests reasons for the performance differences documented between females and males. The amount of time and number of steps to solve each problem varied between males and females, with females taking extra steps to verify their answers and therefore taking more overall steps per problem than males. This verification process resulted in females spending more time on each problem than did males. The extra steps and time, however, paid off in the females performing slightly higher than the males on the IMMEX problems.

Discussion

Males seemed more able than females to solve mathematics problems efficiently. Males tended to collapse or skip steps more often than did females, thus spending less time and taking fewer steps per problem. IMMEX analyses confirmed that females took more steps to validate their correct responses than did males, which, under timed conditions, limited the number of items to which the females could respond and resulted in the females correctly solving fewer problems than did the males. Females' lower scores on the math sections of the timed paper-and-pencil SAT-I seemed to be a result of the time pressure. In the untimed IMMEX situation, females outperformed males on the same items that they had been presented with in the timed tasks. In the IMMEX problem space, the females also attempted more items than did males. Females tended to follow algorithms, verifying their steps and the answer before completing a problem. By being more purposeful and verifying their responses, females better self-regulated their work, which is beneficial in the classroom but is not as productive in timed high-stakes testing.

Interestingly, females showed a greater need for verification in their work. An informal talk with a few students revealed greater frustration among females than among males with getting a wrong answer on the first try. Males felt that in the IMMEX system they still had a second chance and were consequently less concerned. Females may have been more concerned because they had already planned their work carefully and therefore would have to be even more meticulous to find out why their answers were incorrect.

This research shows the importance of understanding in detail the steps that students take when solving mathematics problems. Tracing students' steps allows researchers to better understand how students organize information when coming up with an answer. In the present study, tracing students' steps illuminated some of the reasons that females' math test scores are typically lower than males' scores. Finding out what knowledge students possess no longer needs to be surmised only from final answers and scores, as the use of IMMEX in this study demonstrates. Researchers can and must continue to probe the processes that students employ and the knowledge they bring to bear when confronted with a mathematics problem. The more deeply that educators and researchers can analyze student thinking, the better we can measure students' competence, knowledge, and abilities—and thus the better we can design tools and practices for teachers to teach them effectively.

Implications of this study include the following:

1) We need to find ways to increase females' metacognitive skills in mathematics so that when they participate in assessments, their performance reflects their actual understanding rather than their habitual approaches to problem solving.

2) We also need to explore gender differences in mathematics performance across countries, to see what aspects of the performance differences found in the U.S. are manifested internationally, and to see if there are cultures in which females' mathematics performance is not affected by these factors, so that we can learn from these cultures and close the gender performance gap for United States students.

References Cited

American Association of University Women. (1992). *The AAUW report: How schools shortchange girls*. Researched by the Wellesley College Center for Research on Women. Washington, DC: Author.

American Association of University Women . (1998). *Gender gaps: Where schools still fail our children*. Washington, DC: Author.

Angoff, W. H. (1974). *Criterion-referencing, norm-referencing, and the SAT* (RR-74-01). Princeton, NJ: Educational Testing Service.

Atkinson, R. C. (2001). Achievement versus aptitude tests in college admissions. *Issues in Science and Technology 18(2).* Retrieved March 2, 2008, from http://works.bepress.com/richard_atkinson/28.

Bennett, R. E., Morely, M., & Quardt, D. (1998). *Three response types for broadening the conception of mathematical problem solving in computerized-adaptive tests.* (RR 98-45-ONR). Princeton, NJ: Educational Testing Service.

Braswell, J. S., Lutkus, A. D., Grigg, W. S., Santapau, S. L., Tay-lim, B.S.H., & Johnson, M. S. (2001). *The nation's report card: Mathematics 2000*. Washington, DC: National Center for Education Statistics.

Chi, M. T. H., & Glaser, R. (1985). *Problem-solving ability*. Report no. LRDC-1985/6. Pittsburgh, PA: University of Pittsburgh Learning Research and Development Center.

Coley, R. (2001). *Differences in the gender gap: Comparisons across racial/ethnic groups in education and work*. Princeton, NJ: Educational Testing Service, Policy Information Report.

College Board (1997). *10 Real SATs.* Forrester Center, WV: Author.

College Board. (2007). *2007 College-bound seniors total group profile report*. New York, NY: Author.

Dorans, N. J., & Lawrence, I. M. (1987). *The internal construct validity of the SAT* (RR-87-36). Princeton, NJ: Educational Testing Service.

Dwyer, C., & Johnson, L. (1997). Grades, accomplishments, and correlates. In W. Willingham & N. Cole (Eds.), *Gender and fair assessment* (pp. 127–156). Mahwah, NJ: Erlbaum.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data.* Revised edition. Cambridge, MA: MIT Press.

Ericsson, K.A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine 79(10)*, 70–81.

Gallagher, A. M. (1990). *Sex differences in the performance of high-scoring examinees on the SAT-M* (RR 90-27). Princeton, NJ: Educational Testing Service.

Gallagher, A. M. (1992). *Strategy use on multiple-choice and free-response items: An analysis of sex differences among high-scoring examinees on the SAT-M* (RR 92-33). Princeton, NJ: Educational Testing Service.

Gallagher, A. M., & De Lisi, R. (1994). Gender differences in the Scholastic Aptitude Test-Mathematics problem solving among high-ability students. *Journal of Educational Psychology, 86(2)*. 204–211.

Gallagher, A. M., & Kaufman, J. C. (2005). *Gender differences in mathematics*. New York: Cambridge University Press.

Geiser, S., & Studley, R. (2001). *UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California*. Oakland, CA: University of California Office of the President.

Hartley, K., & Bendixen, L.D. (2001). Educational research in the Internet age: Examining the role of individual characteristics. *Educational Researcher, 30*(9), 22–26.

Katz, I. R., Friedman, D. E., Bennett, R. E., & Berger, A. E. (1996). *Differences in strategies used to solve stem-equivalent constructed-response and multiple-choice SAT-mathematics items* (RR-96-20). Princeton, NJ: Educational Testing Service.

Kenney-Benson, G.A., Pomerantz, E. M., Ryan, A.M., & Patrick, H. (2006). Sex Differences in Math Performance: The Role of Children's Approach to Schoolwork. *Developmental Psychology, 42(1)*, 11–26.

Marshall, S. P. (1995). *Schemas in problem solving*. New York, NY: Cambridge University Press.

Mislevy, R. J., Yamamoto, K., & Anacker, S. (1991). *Toward test theory for assessing student understanding* (RR 91-32-ONR). Princeton, NJ: Educational Testing Service.

Mullis, I. V. S., Martin, M. O., Fierros, E. G., Goldberg, A. L., & Stemler, S. E. (2000). *Gender differences in achievement*. Chestnut Hill, MA: TIMMS International Study Center.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice Hall.

Paek, P. L. (2002). Problem solving strategies and metacognitive skills on SAT mathematics items (Doctoral dissertation, University of California, Berkeley, 2002). *Dissertation Abstracts International, 63*(09), 3139.

Perie, M., Grigg, W. S., & Dion, G. S. (2005). *The nation's report card: Mathematics 2005*. Washington, DC: National Center for Education Statistics.

Pomerantz, E. M., Altermatt, E. R., & Saxon, J. L. (2002). Making the grade but feeling distressed: Gender differences in academic performance and internal distress. *Journal of Educational Psychology, 94,* 396–404.

Santapau, S. L. (2001). *The nation's report card: Mathematics highlights 2000*. Washington, DC: National Center for Education Statistics.

Schank, R. C. (1990). Case-based teaching: Four experiences in educational software design. *Interactive Learning Environments, 1(4),* 31–53.

Schoenfeld, A. H. (1988). Problem solving in context(s). In R. I. Charles & E. A. Silver (Eds.), *The teaching and assessing of mathematical problem solving* (pp. 82–92). Reston, VA: National Council of Teachers of Mathematics.

Song, L. & Hill, J. R. (2007). A conceptual model for understanding self-directed learning in online environments. *Journal of Interactive Online Learning, 6*(1). Retrieved March 12, 2008, from http://ncolr.org/jiol/issues/viewarticle.cfm?volID=6&IssueID=19&ArticleID=98.

Stevens, R. H. (1991). Search path mapping: a versatile approach for visualizing problem-solving behavior. *Academic Medicine, 66* (9), S72–S75.

Stevens, R. H., & Soller, A. (2005). Machine learning models of problem space navigation: The influence of gender. *ComSIS, 2*(2), 83–98.

Stevens, R., Soller, A., Cooper, M., & Sprang, M. (2004). Modeling the development of problem solving skills in chemistry with a web-based tutor. In J. C. Lester, R. M. Vicari, & F. Paraguaca (Eds.), *Intelligent Tutoring Systems*. Springer-Verlag Berlin Heidelberg, Germany. 7th International Conference Proceedings, pp. 580–591.

Tatsuoka, K. K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold, & M. G. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Tatsuoka, K. K. (1993). Item constructive and psychometric models appropriate for constructed responses. In R .E. Bennett & W. C. Ward (Eds.), *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing and portfolio assessment.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Tatsuoka, K. K. (1995). Architecture of knowledge structures and cognitive diagnosis: A statistical pattern recognition and classification approach. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively diagnostic assignment.* Hillsdale, NJ : Lawrence Erlbaum Associates.

Wainer, H. (1983). *An exploratory analysis of performance on the SAT* (RR-83-36). Princeton, NJ: Educational Testing Service.