Journal of International Women's Studies

Volume 5
Issue 4 *Women in Science*

Article 2

May-2004

Gender, Context, and Physics Assessment

Laura McCullough

Gender, Context, and Physics Assessment

By Laura McCullough

Abstract

A persistent gender gap exists on one of the most commonly-used physics conceptual tests, the Force Concept Inventory. The test includes many stereotypically male contexts such as hockey, rockets, and cannonballs. A revised version of the test was created using stereotypically female contexts and both versions were randomly administered to 300 college students. While the total correct score did not change for men and women, significant results were discovered when test questions were examined individually. Results suggest that context can affect performance on a physics assessment for both men and women. One implication for instructors is that they should be aware of how their examples and problems can elicit different performance among women and men.

Key Words: science education, assessment, context

Introduction

Issues of gender inequality and science have been under discussion for many years. Despite ongoing concern on the subject, the physical sciences remain heavily male-dominated, with physics demonstrating one of the most severe under-representation of women (NSF 2002). In an increasingly scientific and technological world, society needs to encourage all people to learn and study science and technology. To make educated choices about public officials and politics, and to make the best choices about their health and well-being, people need to understand the basics of science. In general, we are not doing a good job with our women and girls. Women fear or are feared by the culture of science and in consequence are not getting an adequate scientific education, which in turn means that they are not always in a position to make the most informed choices for themselves and their families. It follows that actions to specifically enhance learning for women need to be taken, as well as actions to encourage women to seek careers in these fields.

In the United States women do well in education at a general level. They earn 57% of all bachelor's degrees and 44% of doctorates. Yet in the sciences those numbers drop dramatically, particularly in the physical sciences and engineering.

In physics, about 50% of high school physics students are young women (Ivie & Stowe, 2000). This is encouraging, although the advanced placement courses are still more heavily populated by men. But at the college level only 22% of physics bachelors degrees are earned by women. That number drops further to 14% at the doctoral level. The statistics for participation in physics and other fields in the U.S. are available from the National Science Foundation (2002).

Teachers of physics also illustrate this gender discrepancy; at the high school level, only 29% of physics teachers are women (Neuschatz & McFarling, 2003). In college, women make up only 11% of assistant professors, 10% of associate professors, and 5% of full professors of physics (Nelson & Rogers, 2004). This means that young women have few role models and female mentors available in physics.

Across the world, the percentage of women in physics is not much better (Ivie, Czujko, & Stowe, 2001). France has one of the highest levels, with 27% of physics PhDs going to women. Many Asian nations are lower in ranking, with China at 13% and Japan at 8% of PhDs in physics going to women. This is a worldwide problem, and all countries need to focus on promoting the participation of women in science.

Why is there such a strong gender disparity in physics? Part of the answer lies with physics education. Poor pedagogy is a large factor in students' decisions to leave science (Seymour & Hewitt, 1997). Physics education should be examined closely for biases that may exclude particular learners. Classroom education of any sort consists of three general parts: curriculum, instruction, and assessment. Science curriculum and instruction have been closely scrutinized for gender bias and many positive changes have been made. Textbooks now include pictures of female scientists (Bazler & Simonis, 1990; Potter & Rosser, 1992), and teachers are more aware of the challenges and problems facing girls in education (AAUW, 1995 and 1999). Other aspects of science teaching receive less attention, particularly physics tests and assessments that could be contributing to the unwelcoming atmosphere of the physics classroom.

The question this research project addressed was whether gender-biased contexts in a particular physics assessment could contribute to gender gaps in performance.

Background

The issues surrounding women and science have been much discussed over the last few decades. These discussions have taken many forms. One branch has been the dialogue about the masculine nature of science and how that has affected women's participation in science and the growth of science itself. Londa Schiebinger (1999) gathers much of this debate together in her book "Has Feminism Changed Science?" Science, particularly the "hard" sciences such as chemistry and physics, are typically thought of as being objective, unbiased. Historically, it has not been considered that who does the science might affect the science itself. Yet Schiebinger, Evelyn Fox Keller, and others suggest that this is not the case. Throughout history women have been excluded from science via many different means. The lack of women in science has led to masculine theories and interpretations. One example would be the use of gender to define botanical groupings and the use of sexual metaphors in botanical reproductions (Schiebinger, 1991). Why would plant reproduction be anything like human reproduction? The male researchers created a gendered situation where a nongendered explanation might have sufficed; and their gendered taxonomy ranked "male" parts of a flower higher than "female" parts. This is just one example of how gender has unnecessarily infiltrated aspects of theoretically objective science.

Physical science can be shown to also have inherent masculine or sexual characteristics. The building of atomic and hydrogen bombs during the Second World War shows just how gendered science can be (Cohn, 1996). The creators of the bombs spoke of giving "birth" to the bombs, and the babies were of course male: Fat Man and Little Boy. The language surrounding the creation of the bombs was strikingly sexual: "…lectures were filled with discussions of vertical erector launchers, thrust-to-weight ratios, soft lay downs, deep penetration…" (Cohn, p. 189). Neutral objective science, indeed.

Other critics of masculinist science have included Donna Haraway, Sandra Harding, and Ruth Bleier. Bleier edited a thoughtful collection of articles in "Feminist Approaches to Science," printed in 1986. Among these are included several critiques of science ranging from epistemological critiques to classroom and educational critiques and assessments of sex differences research. The spectrum of discussions and critiques of science has been wide-ranging over several decades of research and thought on science and gender. Taking a different point of view, Wertheim (1995) posits that the connections between religion and science have also served to keep women out of science. In general, it is now understood that science is not as objective as it was once perceived. Who the scientists are affects how the science grows and what the science is.

Taking feminist and masculinist science ideas into the classroom, Sue Rosser argues that the science classroom is also heavily gendered and the masculine nature of science classes contributes to the lack of women in science (Rosser, 1986). The standard model of teacher as source of wisdom and student as recipient of wisdom is very traditional and tends towards the masculine, particularly in light of the dearth of female science teachers. Rosser contends that using women's studies methods, theories, and pedagogies in the science classroom might serve as a way to attract and retain female students (Rosser, 1990). Cathy Middlecamp (1999, 2000) applies feminist pedagogies in the sciences by using interactive activities and allowing students to determine both questions and answers, in an effort at making science more friendly to women and others.

Gender gaps in assessment have been studied for some time. Many gender researchers are familiar with claims of gender bias against the Scholastic Aptitude Test (Wilder & Powell, 1989 and Navarro, 1989). The Educational Testing Service conducted a study in which they argue that national tests are not biased yet admits that there are persistent gender differences (Cole, 1997). Others have argued that gender gaps on tests are not due to bias but due to innate gender differences (e.g., Benbow & Stanley, 1980).

Eighth graders in the US show no gender gap in mathematics nor in science, according to the Trends in International Math & Science Study (US DOE, 2000). Other studies show that the gender gap tends to show up at higher grade levels (Cole, 1997; AAUW, 1999). Connected with the gender gaps in assessment are gaps due to other factors such as race and socioeconomic status. Of further interest is the TIMSS data showing that among US eighth graders, larger family wealth correlates with a better science score. Also noted in the TIMSS study is that white students outscored Asians, Hispanics, and African-Americans. Race remains an issue in other standardized testing as well ("This Wasn't Supposed to Happen" 1999; Hall & Davis, 1999) and arguments about assessments actually testing for culture, not content, persist (Gleaves, 1994).

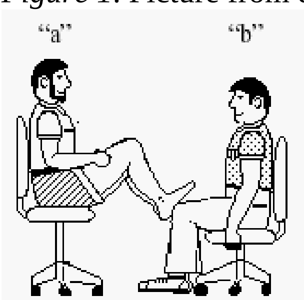
Deeper research into gender differences on tests has produced quite interesting results. In a recent report, Buck et. al. (2002) found that male and female students taking Advanced Placement tests show a significant tendency to do better on questions in content areas favoring their gender. Men did better on questions relating to war, politics and history, among others, while women did better on arts and literature topics, religion, and women's issues.

The issue of context of science questions has been less well studied. Rennie and Parker (1993) studied context in physics problems and found that teachers can create good gender-neutral or gender-inclusive assessment tools by looking at language, portrayal of stereotypes, and particular contexts. The students' reactions suggest that appropriate contexts make problems easier to visualize and more interesting (Rennie & Parker, 1996 & 1998). Concrete problems were preferred over abstract problems. An example of an abstract problem statement follows: "An object is propelled vertically into the air. The object has its maximum potential energy…" whereas a contextualized version starts: "There is a big fireworks display over the Swan River near the city on Australia Day. One rocket is launched into the air…" (Rennie & Parker, 1998, p. 121)

In a South African study of context, Enderstein, et. al. (1998) found that in South African pupils, changing the context of a physical science question substantially affected the responses they received from the pupils. They had changed questions to better target the experiences of urban and rural pupils. Problems with humans or human perspectives elicited different responses than problems without human figures present.

Another good piece of evidence for the contextual dependency of performance lies in a story related by a colleague teaching in Thailand (personal correspondence). He gave the test under consideration to his students, but was unprepared for their response to the question referring to the situation of a person putting his bare feet on another person's knees (see Figure 1). This act would be extremely rude in Thailand, where the feet are considered an unclean part of the body. The students had severe trouble answering this question because the situation was so unbelievable to them. The *context blocked the physics*.

Figure 1. Picture from question on conceptual test



The present study looks at a particular physics assessment test, the Force Concept Inventory (Hestenes, Wells, & Swackhamer, 1992). The test is a thirty-question, multiple-choice conceptual test, which covers the topics usually found in the first semester of introductory physics courses. There is minimal math required to answer the questions and the non-correct answers (distracters) were written based on physics education research on common student misconceptions. The FCI is used across the United States in high schools, colleges and universities. Some instructors give course credit for taking the test; others make the test un-graded and/or voluntary. The most common use for the test is to assess the instructional effectiveness of the course. The FCI has been shown to correlate strongly with pedagogy (Hake, 1998).

The FCI has a significant gender gap favoring male students (McCullough, 1996 and McCullough & Crouch, 2002). This gender gap is not explained by physics background; when broken out by previous physics coursework men receive higher scores than women at each educational level. Nor does a course in physics ameliorate the problem; the gender gap is maintained from pre-instruction to post-instruction.

If women are receiving artificially lower scores on this test because of contextual bias, then not only is the test doing them a disservice, but it is also an inappropriate measure of instructional effectiveness for instructors and educational researchers who use the test. The goal of this research is to determine if context affects performance, and if so, to develop a version of the test in which context does not contribute to the gender gap on the test.

Methods

In order to determine if the contexts of the questions were affecting performance, the context had to be separated from the physics. A direct way to do this was to create a new version of the test in which the physics was kept identical to the original but the context of each question was changed. The original test used mostly male persons and either male-oriented or school- or lab-oriented contexts. Questions about rockets, cannonballs, hockey and male figures were included alongside questions about steel balls rolling off a horizontal table. The revised version of the test, called the Revised FCI or RFCI, used *stereotypically* female contexts such as shopping, cooking, jewelry and stuffed animals. Every figure and person mentioned is female. The contexts were pushed as far female-centric as manageable. This was done so that if performance exhibited context-dependence it would be as visible as possible. If only small changes in context had been made, corresponding small changes in performance might not have been detectable within the small populations under study. If large changes were made and large changes in performance were seen, then there would be more reason to believe that context did have an effect on performance.

The other significant change made to the test was to make the abstract, school-oriented contexts more concrete and closer to daily-life situations. As mentioned above, research (Rennie & Parker 1998) has found that female students prefer more concrete problems. Context-less problems such as a nondescript ball being thrown were changed to be more specific.

The RFCI was pilot-tested in several populations and revisions were made based on problems found in the test, such as a figure which was too big to be usable and a context which was not quite identical to the original in a physics sense (McCullough 2001). The version of the RFCI used in this study was the third version of the test. The physics of each question is the same as the physics in the original to the extent that this assessment is designed to test. For example, minor issues such as air resistance are ignored by most people who use this test.

The population chosen for the study was a group of non-physics students in general education classes such as English at a mid-sized Midwestern state university. Because this study was concerned about context, the context of the class was also a concern. Physics classes are male-dominated and may be a contributor to female under-performance. Claude Steele (1997) has found that cuing people in on gender in gender-discriminatory situations such as math and science testing can cause the minority population to do more poorly. This "stereotype threat" was a significant concern in this study. In order to reduce the possible gender-discriminatory cuing, non-physics classes were chosen and the test was introduced in a way that didn't mention gender at all.

With IRB approval, eight classes were chosen based on the instructor's willingness to give up class time to this project. Five were English or Literature classes, two were

Sociology classes, and one was a math class. In all but one class the researcher herself explained and administered the test using prepared notes (not a formal script). For the other class, the instructor introduced the study using a more formal script covering the same points. Taking the test was voluntary, but less than 5 students chose not to take the test.

Both versions of the test were used and were handed out in alternating order (A-B-A-B) so that students sitting next to one another had different tests. That there were two versions of the test was not mentioned. Students answered the test questions on a separate answer sheet which included nine demographic questions on the back. Students were asked about 15 minutes into the test to answer the demographic questions after they finished the test questions. This was done to again minimize the stereotype threat. The demographic questions included: previous physics and math courses taken, major, year in school, gender, years of high school math and science, and semester courses of college math and science. A total of 312 students participated in the study.

Results

An initial look at the overall scores of the four groups shows that changing the version did not significantly change the scores of the women, though there was a measurable difference in the male students' scores. Table 1 lists the average percent correct for the four groups under consideration.

Table 1. Average percent correct by test version and gender (Number of students and standard deviation)

	Original FCI	Revised FCI
Men	33.7 (N=56, sd=14.4)	28.5 (N=71, sd=11.7)
Women	21.7 (N=106, sd=9.7)	22.3 (N=79, sd=8.1)

The difference between the women's scores is not significant ($p>.05$). The difference between the men's scores is significant ($p=.005$) in favor of the original version. Changing the context of the questions towards more female-oriented contexts negatively affected the overall male score while having no effect on the overall female score. An initial interpretation of this result could be that women are less affected than men by changes in context. However, the literature does not support this interpretation, suggesting a need for closer examination of the data.

These somewhat unexpected results suggested that a closer look was needed. The next step of analysis was to look at each question and analyze student performance by men and women on both versions. The item analysis leads to a different interpretation of the effect of the version change.

The first question asked for the question-by-question analysis was this: on how many questions did women improve on the revised gender version? Looking at the number of women who got each question correct on each version, women did better on the gender version on 13 out of the 30 total questions. In contrast, men did better on the gender version on only 5 questions. There was no change in the women's or men's scores on 2 questions. Another way to view this is to say that the gender version *decreased* or depressed women's scores on 15 questions and men's scores on 23 questions. This is in accordance with the overall results which suggest that the gender version hurts men's performance but neither helps nor hinders women's overall performance.

Why might this be happening? Perhaps women are so acclimated to bias in their science texts and tests, that they do not even consciously see the bias, and so it matters little to them. Or men may be so accustomed to male contexts that it shakes them out of their usual habits and thinking patterns to see the bias reversed. Alternatively, men may be more context-sensitive in general than women. It also should be noted that women's scores were close to what would be achieved by random guessing; since the scores are so low, it might be that we are seeing different patterns of random guessing among the women. Men's slightly higher score may be reflecting more thought going into the questions, which might increase the effect of context. However, since the FCI was designed to draw out students' right and wrong ideas, and the incorrect answers (distracters) were specifically chosen to be common misconceptions, it is unlikely that random guessing can explain this effect.

The next layer of analysis was to determine if there were particular questions with notable performance patterns. Were there questions where women improved but men did not? Vice-versa? The answer is clearly a resounding yes. Every combination of relationships imaginable showed up in the test. Particular questions serving as examples are shown below. Because the authors of the FCI are very concerned about the test becoming easily available to students, which they believe would reduce the test's usefulness I will not include the actual questions on the test. Educators can download the test by visiting the FCI website (see references) and requesting a password.

Cannon/bowl: The original question asks about the path of a cannonball fired off the top of a cliff. The picture shows a figure, more male- than female-looking, firing the cannon. The revised question asks about the path of a bowl shoved by a (female) baby off her high chair. The physics is identical between the questions at the level of analysis asked of the students. The only difference is the *context*. The results in Table 2 suggest that changing the context helped the women without hurting the men.

Table 2. Percentage of students choosing correct answer on cannon/bowl question by test version and gender.

	Original	Revised
Women	34	51
Men	66	66

Balls/oranges: Another question asks students about two steel balls rolling off a table and how far from the table they hit the floor, given that one ball weighs twice as much as the other. The revised question simply changes the balls to oranges, and inserts the modifier "kitchen" before "table". This change was one of the "daily-life experience" changes, to make the question less formal and school-oriented and more informal and familiar. Here, the revision actually increased the gender gap on the question, as seen in Table 3. The women performed more poorly on the revised question, while the men performed better on the more female-oriented question. This makes drawing conclusions more complicated.

Table 3. Percentage of students choosing correct answer on ball/orange question by test version and gender.

	Original	Revised
Women	25	18
Men	25	37

Ball in channel/waterslide: A different pattern was seen on a question asking about the path taken by a ball exiting a horizontal circular channel. The original question gives a paragraph of introduction describing the setup (a frictionless horizontal circular channel is secured to a table top; the ball enters the channel at high speed and exits at the end of the channel). The revision involved changing the situation to a girl on a water slide; the last section of the slide is a horizontal circular arc. Again, the physics of the situations is the same. The path of a ball and a girl as they leave the circle are the same. In this case, the revision improved the women's scores while depressing the men's scores.

Table 4. Percentage of students choosing correct answer on channel/waterslide question by test version and gender.

	Original	Revised
Women	46	53
Men	77	67

The remaining questions show all the possible patterns: women up, men same; both same; men up, women down; etc. The single all-encompassing result of this study is that changing the context can and does affect student performance, but in ways that are hard to predict in terms of gender bias.

Discussion

The context in which a question is couched can affect how a student responds to that question. Replacing male-oriented contexts with female-oriented ones did reduce the gender gap on a popular physics conceptual test. This came at the cost of lowering men's performance rather than raising the women's. There are several interpretations of this data. Women may be accustomed to male-oriented examples and are less disturbed by changes to more female-oriented contexts. Perhaps the stereotypical contexts chosen (in 1998) do not match with current young women's worlds, and so there remains a mismatch in contexts for the women, but the contexts are far enough away from men's experiences to make them uncomfortable. What is certain is that context does interact with gender to affect how students perform on test questions.

The current version of the Revised FCI is not yet ready for answering the question of whether or not the context of the questions is lowering women's scores. What has been shown is that context does affect performance. The next steps in this research are to further investigate how context affects performance. Interviews of students answering the same questions with different contexts are being planned. Also planned are shorter versions of the test which include both contexts for some questions.

Since the test is used primarily by physics instructors in physics classes, the Revised FCI must be tested with physics students and this is already underway at six different institutions across the country.

The underlying issues of this study have implications for instructors at many levels. When writing tests, instructors need to be aware of possibly biased contexts in their questions, not only for gender but for culture as well. Asking a question about a pop-fly baseball can alienate not only women (and men) not interested in the sport but students from other cultures who may not even know what a baseball is. In textbooks, back-of-chapter questions should be examined for their gender and cultural context and appropriateness to the students at hand. Questions about such things as driving on ice may be inappropriate for students who come from warmer climates. Even in day-to-day discussions and examples, instructors need to be aware of how the contexts they use may be affecting some students' learning and understanding.

The context of a question can affect how a student interprets, relates to and responds to that question. Further research is needed to examine more thoroughly the connection and relationships between context and response. The current research suggests that teachers need to be aware of the contexts they use in their classrooms, so that they do not inadvertently disadvantage the women they teach. By learning more about context and contextual bias, our classrooms can become more accommodating to women and we can invite broader participation of women in science and in science classrooms.

References

AAUW (1995). *How Schools Shortchange Girls*. New York: Marlowe & Co.

AAUW (1999). *Gender Gaps.* New York: Marlowe & Co.

Bazler, J. and D. Simonis. (1990). Are women out of the picture? Sex discrimination in science texts. *Science Teacher 57*(9): 24-26.

Benbow, C. P. and J. Stanley (1980). Sex differences in mathematical ability: Fact or artifact? *Science 210*: 1262-1264.

Bleier, R. (1986). *Feminist Approaches to Science.* New York: Pergamon Press.

Buck, G., I. Kostin, et al. (2002). *Examining the relationship of content to gender-based performance differences in advanced placement exams.* New York, College Board/ETS. Report #2002-12.

Cohn, C. (1996). Sex and death in the rational world of defense intellectuals. In *Gender and Scientific Authority*, Laslett, B., S. G. Kohlstedt, et. al. (Eds). Chicago, IL: University of Chicago Press, pp. 183-214.

Cole, N. (1997). *The ETS Gender Study*. Princeton, NJ: Educational Testing Service.

Enderstein, L.G. & P.E. Spargo (1998). The effect of context, culture and learning on the selection of alternative options in similar situations by South African pupils. *International Journal of Science Education, 20*(6): 711-736.

Force Concept Inventory available on the Internet at http://modeling.asu.edu/R&E/Research.html

Gleaves, K. (1994). Urban African-American students' perceptions of school and tests: Re-examining the cultural bias theme. *International Journal of Educational Reform*, 3(2), 194-201.

Hake, R. R. (1998). Interactive-engagement vs traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*: 64-74.

Hall, C. & N. Davis. (1999). Gender and racial differences in mathematical performance. Journal of Social Psychology, 139(6), 677-689.

Hestenes, D., M. Wells & G. Swackhamer (1992). A Force Concept Inventory. *The Physics Teacher, 30*: 141-151.

Ivie, R., R. Czujko & K. Stowe (2001). *Women Physicists Speak: The 2001 International Study of Women in Physics.* AIP Report. http://www.aip.org/statistics

Ivie, R. & K. Stowe (2000). *Women in Physics, 2000.* AIP report R-430. College Park, MD: American Institute of Physics.

McCullough, L. (1996). *Gender differences in multiple-choice conceptual tests.* Paper presented at the national meeting of the American Association of Physics Teachers. College Park, MD. Available at http://physics.uwstout.edu/staff/mccullough/physicseduc.htm.

McCullough, L. (2001). *A gender context for the Force Concept Inventory*. Paper presented at the national meeting of the American Association of Physics Teachers. San Diego, CA. Available at http://physics.uwstout.edu/staff/mccullough/physicseduc.htm.

McCullough, L. & C.H. Crouch (2002). *Gender, Educational Reform, and Instructional Assessment, Part 1.* Paper presented at the national meeting of the American Association of Physics Teachers. Philadelphia, PA. Available at http://physics.uwstout.edu/staff/mccullough/physicseduc.htm.

Middlecamp, C. and B. Subramaniam (1999). What is feminist pedagogy? Useful ideas for teaching chemistry. *J. of Chem. Ed.* **76**(4): 520-524.

Middlecamp, C. and A.-M. Nickel (2000). Doing science and asking questions: An interactive exercise. *J. of Chem. Ed.* **77**(1): 50-52.

National Science Foundation (2002). Division of Science Resources Statistics; *Science and Engineering Degrees: 1966-2000.* NSF report 02-327. Arlington, VA: NSF. Author: Susan T. Hill. http://www.nsf.gov/sbe/srs/nsf02327/start.htm

Navarro, C. (1989). *Why do women have lower average SAT-Math scores than men?* Paper presented at American Educational Research Association meeting, San Francisco, CA.

Nelson, D. and D. Rogers. (2004). *A National Analysis of Diversity in Science and Engineering Faculties at Research Universities.* Research Report, University of Oklahoma. Available at http://www.now.org/issues/diverse/diversity_report.pdf

Neuschatz, M. and M. McFarling. (2003). *Broadening the Base: High School Physics Education at the Turn of a New Century*. AIP Report R-439. Available at http://www.aip.org/statistics/trends/reports/hsreport2003.pdf

Potter, E. and S. Rosser (1992). Factors in life science textbooks that may deter girls' interest in science. *Journal of Research in Science Teaching 29*(7): 669-686.

Rennie, L. & L. Parker (1993). Assessment in physics: Further exploration of the implications of item context. *Australian Science Teachers Journal, 39*(4): 28-32.

Rennie, L. & L. Parker (1996). Placing physics problems in real-life context: Students' reactions and performance. *Australian Science Teachers Journal, 42*(1): 55-59.
Rennie, L. and L. Parker (1998). Equitable measurement of achievement in physics: High school students' responses to assessment tasks in different formats and contexts. *Journal of Women and Minorities in Science and Engineering* **4**(2): 113-127.
Rosser, S. (1986). *Teaching Science and Health from a Feminist Perspective*. New York: Pergamon Press.
Rosser, S. (1990). *Female Friendly Science: Applying Women's Studies Methods and Theories to Attract Students.* New York: Pergamon Press.
Seymour, E. and N. Hewitt (1997). *Talking about leaving: Why undergraduates leave the sciences.* Boulder, CO: Westview Press.
Schiebinger, L. (1999). *Has Feminism Changed Science?* Cambridge, MA: Harvard University Press.
Steele, C. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52*(6): 613-629.
This wasn't supposed to happen: The Black-White SAT gap is actually growing larger. 1999. *Journal of Blacks in Higher Education*, 25, p. 95-100.
U.S. Department of Education. National Center for Education Statistics.
Mathematics and Science in the Eighth Grade: Findings from the Third International Mathematics and Science Study, NCES Report 2000-014, Washington, DC, 2000.
Wertheim, M. (1995). *Pythagoras' Trousers: God, Physics and the Gender Wars.* New York: W. W. Norton.
Wilder, G. Z. and K. Powell (1989). *Sex Differences in Test Performance: A Survey of the Literature.* New York: College Entrance Examination Board.