

## WOMEN, MOTIVATION, AND ACHIEVEMENT

Janet Shibley Hyde  
*University of Wisconsin*

Kristen C. Kling  
*St. Cloud State University*

Women's educational and occupational achievements are crucial to the economic productivity and prosperity of the nation, as well as to the mental health of women and their families. In this article we review psychological research on motivation and on educational achievement, focusing on gender and the contributions that have been made by feminist researchers. Feminist psychologists noted the sex bias and methodological flaws in traditional research on achievement motivation and proposed vastly improved models, such as Eccles's expectancy x value model of achievement behavior. Contrary to stereotypes, gender similarities are typically found in areas such as mathematics performance. Policymakers should be concerned about gender bias in the SAT and about the Female Underprediction Effect. Additional threats to girls' and women's achievements include stereotype threat and peer sexual harassment in the schools.

Women's achievements are crucial to the economic productivity and prosperity of the nation, as well as to the well-being of women and their families. From 1964 to 1999, approximately 71 million jobs were added to the U.S. economy, permitting remarkable economic expansion. Of those 71 million jobs, 43 million were occupied by women and 28 million by men (U.S. Department of Labor, 2000a). Thus women's work achievements have been crucial to the expansion. Women's educational attainments, in turn, have been essential to their participation and success in the labor force; the higher the level of education, the more likely a woman is to be in the labor force (U.S. Department of Labor, 2000b). In 1999, women with less than a high school diploma had a labor force participation rate of 31.9% and an unemployment rate of 8.2%, whereas women who were college graduates had a 74.8% participation rate and a 1.8% unemployment rate (U.S. Department of Labor, 2000b). This pattern is consistent across ethnic groups in the United States. For example, Black women with less than a high school diploma have a labor force participation rate of 30.2% and an unemployment rate of 7.4%, whereas Black women who are college graduates have a labor force participation rate of 81.2% and an unemployment rate of 2.6% (U.S. Department of Labor, 1997). Education and the good jobs that it brings can make the difference between poverty and prosperity for women and their families. Because another article in this special issue is devoted to work (Gutek, 2001), we focus here on educational achievements and motivation.

Address correspondence and reprint requests to: Janet Hyde, Department of Psychology, University of Wisconsin, 1202 W. Johnson St., Madison, WI 53706. E-mail: jshyde@facstaff.wisc.edu

Policies designed to foster girls' and women's educational achievements will necessarily help our nation achieve the Decade of Behavior's priority of a better-educated citizenry.

Education and other achievements by women are important not only for the productivity of the nation, but also for women's mental health. Research shows that higher levels of education are associated with and predictive of mental health (Kessler & Cleary, 1980; Miech, Caspi, Moffitt, Wright, & Silva, 1999).

The Decade of Behavior provides an opportunity to step back and assess the current state of scientific knowledge about the psychology of motivation and achievement and to craft an agenda for the next decade of research. Feminist psychology is particularly helpful in these efforts because it has provided cogent critiques noting the flaws in much of the traditional research on these questions, while also articulating new theoretical models, supported by research, that provide better explanations of motivation and achievement for both men and women. To facilitate the goals of the Decade of Behavior, we will review traditional research on gender, motivation, and achievement, as well as the major new contributions to this field that have been made by feminist psychology.

### MOTIVATION

#### Achievement Motivation

Achievement motivation is the desire to accomplish something of value or importance through one's own efforts and to meet standards of excellence in what one does. The traditional method of measuring achievement motivation, developed in the 1950s, uses a projective technique in

which people's stories in response to an ambiguous cue are scored for achievement imagery (McClelland, Atkinson, Clark, & Lowell, 1953). The person being tested is shown a series of pictures and is asked to write a story about each, after being told that this is a test of creative imagination. For example, one of the pictures shows a young man standing on a sidewalk with a broom in his hand, looking off into the distance. If the person wrote a story about how the young man is dreaming about going to college if he can earn enough money at this job, that would indicate high achievement motivation. If the person wrote a story about how the young man is thinking of sneaking out of his job so that he can go to the beach with his friends, that would indicate low achievement motivation.

McClelland and Atkinson's classic theory of achievement motivation has been the object of feminist criticism (Spence & Helmreich, 1983; Stewart & Chester, 1982). For example, Spence and Helmreich (1983), like many others, raised questions about the validity of projective tests that McClelland and Atkinson used to measure achievement motivation. Stewart and Chester (1982) provided a critique of the experimental methods used by McClelland and Atkinson to arouse achievement motivation and noted substantial flaws. In the original research, experimental conditions were manipulated in ways that, theoretically, should increase achievement motivation—for example, by telling participants that the projective test measured not only intelligence but also capacity to act as a leader. Under these conditions, men's achievement motivation increased sharply but women's did not. McClelland and Atkinson then excluded women from their studies because they did not perform as the theory said they should. McClelland went so far as to say "Clearly we need a differential psychology of motivation for men and women" (1966, p. 481). When women's behavior was discrepant from predictions, McClelland and Atkinson, rather than questioning their theory, excluded women from study and viewed women's behavior as problematic.

In summary, traditional research on achievement motivation was androcentric and seriously flawed methodologically. Feminist researchers have emerged from this tangled thicket to suggest new methods of measurement and new theoretical models.

Feminist psychologist Janet Spence developed a non-projective, objective, self-report measure of motivation that, additionally, expanded on the classic unidimensional view of achievement motivation to recognize multiple domains of motivation (Spence & Helmreich, 1983). Spence and Helmreich's research uncovered three dimensions of achievement motivation: work ("I like to work hard"), mastery ("Once I undertake a task, I persist"), and competitiveness ("I feel that winning is important in both work and games"). This research is the precursor to contemporary research on achievement goals, which is reviewed later in this article.

Research in the 1950s and 1960s indicated that females had a lower level of achievement motivation than males

(Hoffman, 1972; Tyler, 1965), consistent with society's emphasis at the time that women should remain in the home instead of venturing out into the achievement-oriented world of work. The evidence suggests that women's achievement motivation has increased over time. Veroff, Depner, Kukla, and Douvan (1980) found that achievement motivation increased among American women from 1957 to 1976, and Jenkins (1987) found increases from 1967 to 1981. More recent studies show no gender differences in achievement motivation (Mednick & Thomas, 1993).

It also seems likely that the opening of educational opportunities and career options for women over the last several decades has increased achievement motivation for women as they experience these employment situations and for girls as they anticipate jobs with exciting possibilities for achievement. Jenkins (1987), for example, found that achievement motivation in female students who were college seniors in 1967 predicted their employment in achievement-oriented occupations 14 years later. Even more intriguing is the finding that women employed as college professors or as business entrepreneurs showed significant increases in their achievement motivation compared with their scores in college, whereas those in other occupations showed no change in achievement motivation (Jenkins, 1987).

Feminist psychologists emphasize attention not only to gender, but to race/ethnicity and class in psychological research. The traditional research on achievement motivation is silent on issues of race and ethnicity. The limited research available, conducted in the 1970s, suggested that middle-class, Black women were more achievement-oriented and displayed less fear of success compared with White women (reviewed by Mednick & Thomas, 1993).

### The Rise and Fall of the Motive to Avoid Success

Seeking alternatives to traditional models of achievement motivation, Horner (1969) conducted research on the motive to avoid success or a fear of success among bright, high-achieving women. In attempting to understand the gender differences in achievement that were present in the 1960s, Horner observed that achievement situations were more anxiety-provoking for females than for males. To measure this phenomenon, Horner devised a projective test in which respondents completed a story that began "After first-term finals, Anne (John) finds herself (himself) at the top of her (his) medical-school class." Women wrote about Anne, men about John.

Men's stories generally indicated happiness and feelings of satisfaction over achievement. Women's responses, in contrast, were far more negative, indicating fears of social rejection, worries about maintaining womanhood, and denial of the reality of success. In Horner's sample from the University of Michigan, 65% of the women showed such negative responses, in contrast to 10% of the men.

Horner collected her original data in 1965 for her doctoral dissertation. The publication of the findings (1969) attracted widespread attention from the popular media, and it was required reading for students in many courses. The research was appealing because it appeared at a time of the emergence of the women's movement and concern over women's equal opportunity. The research seemed to offer a sensible explanation for why more women had not succeeded in high-status occupations—they simply feared success.

In the cold light of day more than 30 years later, the research does not seem nearly as appealing. It has been criticized on a number of grounds (Mednick, 1989; Shaver, 1976; Tresemer, 1977; Zuckerman & Wheeler, 1975): (1) Other studies using Horner's techniques often found men displaying as much motive to avoid success as women. Therefore, there is no reason to believe that it is found only in women or even that it is more frequent in women. If that is the case, it cannot be used to explain women's lesser occupational achievements. (2) Anne's success was in a field that, at the time, was stereotyped as male, namely medical school. Therefore, the research might not indicate a generalized fear of success so much as a fear of being successful in a way that violates stereotypes. Indeed, when Anne was presented as successful in nursing school, women did not show anxiety about her success (Cherry & Deaux, 1978). (3) The research method confounded gender of stimulus person with gender of respondent. That is, women wrote about Anne and men wrote about John. Perhaps women are not anxious about their own success, but rather Anne's success stimulates anxiety, whether a woman or man writes about her and, in fact, one study showed exactly that (Monahan, Kuhn, & Shaver, 1974). The technique may simply have measured cultural stereotypes about women rather than deep unconscious conflicts.

Today, research on motive to avoid success has virtually disappeared. Nonetheless, it provides an important object lesson on the popular appeal of attributing women's lesser achievements to internalized, intrapsychic factors and how, ultimately, such factors were unsuccessful in accounting for the striking gender differences in occupational achievement that characterized the 1950s and 1960s. As we chart the course of the Decade of Behavior, models that assume widespread intrapsychic deficits in women are unlikely to be productive. The models reviewed below show far more promise.

## Current Motivation Research

### *Achievement Goal Theory*

Contemporary research on achievement motivation is framed by achievement goal theory (e.g., Barron & Harackiewicz, 2001; Pintrich & Schunk, 1996). Generally two types of goals are recognized: mastery goals (also called intrinsic goals) and performance goals (or extrinsic goals). When an individual adopts a mastery goal in a par-

ticular activity, her or his purpose is to develop personal competence by acquiring knowledge and skills (e.g., "My goal is to learn as much as I can in this class;" Barron & Harackiewicz, 2001; Midgley, Kaplan, & Middleton, 2001). If the individual pursues performance goals, in contrast, the objective is to demonstrate personal competence and outperform others (e.g., "My goal is to do better than other students in this class"). Performance goals, therefore, tend to be competitive.

Research on achievement goals is generally silent on the question of gender (Midgley, Kaplan, & Middleton, 2001). Two possible reasons for this omission suggest themselves. First, achievement-goals researchers may be ignoring gender. Second, these researchers may be routinely checking for gender differences and finding no significant differences; finding the gender similarity uninteresting, they fail to report it. Supporting the second interpretation, Barron and Harackiewicz (2001) found a non-significant correlation between gender and performance goals ( $r = .10$ ) and a significant but small correlation between gender and mastery goals ( $r = .16$ ), with women being more likely to adopt mastery goals. It seems likely that, for achievement goals, gender similarities are the rule and, when gender differences appear, they are small. This conclusion must remain tentative, though, because contemporary motivation researchers have paid so little attention to gender. The recognition of gender similarities in achievement goals is important as a counter to a possible assertion that boys and girls, and men and women, have entirely different achievement goals and therefore should receive different treatment in the classroom or the workplace.

An alternative to examining gender differences in achievement goals is to ask whether situations that encourage performance goals or those that encourage mastery goals are more beneficial—in terms of fostering achievement and interest—for males or females. For example, are performance goals, perhaps set by a teacher, more facilitative for boys than for girls? Research with French-speaking, Canadian college students indicates that both mastery and performance goals are significantly correlated with academic performance for both women and men (Bouffard, Boisvert, Vezeau, & Larouche, 1995). The research base is limited, but suggests that gender similarities may be found in the role that achievement goals play in performance.

Feminist researchers would question the use of gender as the prime category of analysis. Why ask whether performance goals are more beneficial to boys (on average) or to girls (on average)? Researchers might more profitably search for factors that are more important than gender in predicting who thrives under performance goals and who thrives under mastery goals. Feminist perspectives, then, can do much to enrich research on achievement goals.

*The role of ethnicity and culture.* It is entirely possible that current theory and research on achievement goals

characterize White Americans and do not extend to other ethnic groups within the United States or to other cultures around the world (e.g., Graham, 1994). Most of the research has been conducted with college students, who are a predominantly White population. The emphasis on performance goals as one of the two fundamental motivations may be rooted in an individualistic society.

Theorists distinguish between individualistic and collectivistic cultures (e.g., Triandis, Bontempo, Villareal, Asai, & Lucca, 1988). In individualistic cultures, the emphasis is on individual freedom and the right of the individual to do what she or he wishes. In collectivistic cultures, individuals may be expected to subordinate their personal goals to the goals of some collective, which is typically a stable in-group (e.g., family, tribe, clan). The dominant culture of the United States is highly individualistic, whereas Asian cultures are collectivistic (Triandis et al., 1988).

McClelland and Atkinson's (1953) classic research on achievement motivation and current research on achievement goals clearly focus on individual achievement and individualistic goals. Feminist and cross-cultural researchers urge major expansions of the theories that would include factors such as group goals and motivation (in addition to individual goals); interpersonal, affiliative, and family goals; and cooperation (in addition to competition or performance goals) as a stimulus to achievement.

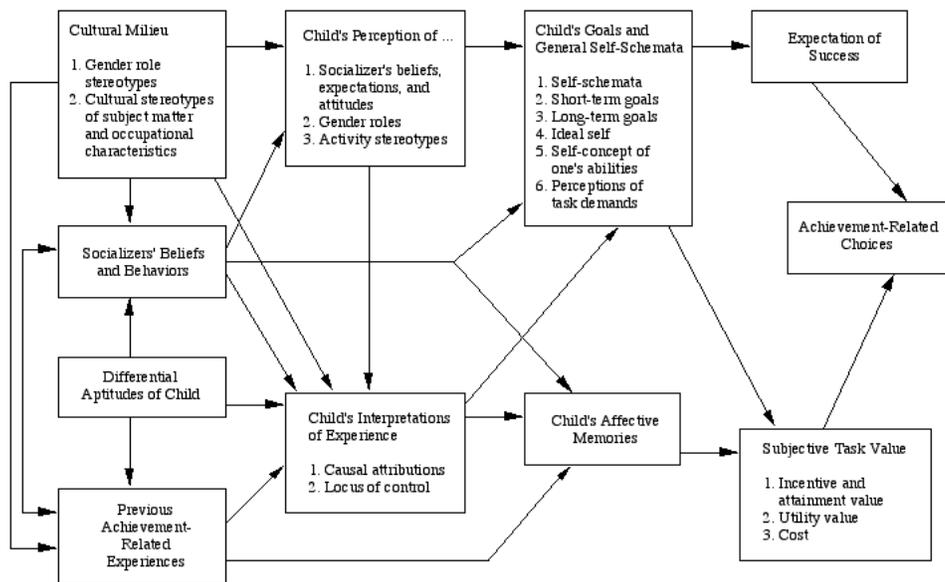
Dabul and Russo (1996; Dabul, 1995), working from a feminist perspective to address these issues, developed an expanded model of achievement motivation, goals, and attributions. In this model, mastery and performance goals are understood to be rooted in an independent, individualistic goal structure. In a collectivistic society, in contrast, fulfilling social obligations of duty and responsibility are the achievement goals, and group goals become the goals

of the individual. Biracial or multiracial individuals may have multiple self-construals that may be engaged depending on the context. This model is too new to have been tested empirically, but it provides exciting possible conceptualizations for future research.

*Eccles: Expectancy x Value Theory*

The massive theoretical and empirical contributions of feminist psychologist Jacquelynne Eccles are crucial to contemporary motivation research and in particular, to understanding gender, motivation, and achievement (for excellent comprehensive summaries, see Eccles, 1987, 1994).

Eccles (1987, 1994) uses an expectancy x value model of achievement motivation and has framed her theory specifically to address gender differences in educational or occupational choices—for example, a high-school girl deciding whether to take calculus during her senior year, or a college woman deciding whether to go to medical school. The model (see Figure 1) specifies that the choice to take on an achievement-related task is the result of two sets of beliefs: the individual's expectations for success at the task (expectancy, top portion of Figure 1) and the extent to which the individual values the task (value, bottom portion of Figure 1). These beliefs in turn have been influenced by social and cultural factors, including socializers (especially parents and teachers), gender-role stereotypes, the individual's self perceptions, and perceptions of the task. All of these influence the expectations for success that the individual holds for various achievement-related options and the subjective value that the individual attaches to the options. The theory addresses the question: Why do girls and boys, and men and women, make the educational and occupational choices they do?



**Fig. 1.** Eccles's expectancy x value model of achievement behaviors. (Source: Eccles, J. S. [1994]. Understanding women's educational and occupational choices: Applying the Eccles et al. model of achievement-related choices. *Psychology of Women Quarterly*, 18, 585–610).

*Expectations for success.* Expectations for success at a task are closely related to task self-confidence. In studies of the general population, Eccles's research consistently shows gender differences in expectations for success in mathematics, athletics, and English, especially among junior and senior high-school students (e.g., Eccles, 1994; Eccles, Wigfield, Harold, & Blumenfeld, 1993). Global self-esteem is not the issue here. Rather, it is task-specific self-confidence. In studies of the general population, consistent with stereotypes, boys are more confident about mathematics (Hyde, Fennema, Ryan, Frost, & Hopp, 1990) and girls are more confident about English. In contrast, in studies of academically talented or gifted youth, the typical result is no gender difference in self-confidence in mathematical ability (Benbow & Stanley, 1982; Eccles, 1994; Eccles [Parsons], Adler, & Meece, 1984).

According to the model, choices are shaped not simply by task self-confidence, but by the individual's hierarchy of self-confidence in different domains. For example, a girl may feel self-confident about both math and English, but if she believes she is more talented in English, that will pull her toward choosing English courses and a related occupation. Eccles's studies of gifted girls show that, although their confidence in their math ability equals that of the gifted boys, the girls have more confidence in their reading ability than their math ability (Eccles & Harold, 1992). This same pattern also has been found in studies of the general population (Eccles, 1994).

The other major component of the model is *subjective task value* (bottom portion of Figure 1). Even if individuals anticipate success at a task, they will not undertake it unless they value it. Value involves components such as interest and usefulness. In longitudinal research on college-bound students, Eccles has found that gender differences in choices to enroll in advanced mathematics courses were mediated primarily by gender differences in the value attached to mathematics (Eccles [Parsons], Adler, & Meece, 1984). Compared with boys, girls felt that math was less important, less useful, and less enjoyable. It is therefore not surprising that they are less likely to choose to take optional mathematics courses. Research with high school seniors on their occupational aspirations showed that, for every occupational category, sense of personal efficacy in that domain was highly predictive of plans to enter that occupation, and the values attached to aspects of a job were predictors of occupational aspirations (Josefowicz et al., 1993, cited in Eccles, 1994). For example, valuing creativity predicted women's plans to become artists or writers, and valuing helping others predicted *not* aspiring to a career in the physical sciences.

Task self-confidence and subjective task value, of course, are powerfully influenced by gender socialization processes, including messages from parents, teachers, textbooks, and the mass media. For example, consistent with traditional gender roles, in a study of high-school seniors, girls placed more value than boys on having a job that allows one to help others, whereas boys placed more value

on becoming famous and making lots of money (Josefowicz et al., 1993, cited in Eccles, 1994). Parents who endorse traditional gender-role stereotypes underestimate their daughters' talent in a male-typed activity like physics and overestimate their sons' talent (Jacobs, 1991). These parental views can, in turn, undermine girls' self-confidence in their ability in these areas (Jacobs, 1991). Indeed, parents' perceptions of their children's talent in math and English are stronger predictors of the children's self-confidence than the children's grades in the relevant courses (Frome & Eccles, 1998).

In regard to ethnicity, research conducted in the 1960s showed Black children have lower perceptions of their academic abilities (Caplan, 1969; Wylie, 1963). More recent studies, however, find no race differences or find Blacks displaying higher perceptions of their academic abilities than Whites (Graham, 1994). These studies, however, typically do not examine gender and race simultaneously. Perhaps the more important question is whether Eccles's theoretical model works as well for ethnic minorities as it does for Whites (Eccles's samples, drawn from the public schools, have, in fact, had excellent representation of ethnic minority children). Eccles has tested links in her model (predicted correlations between variables) with African American samples and finds that the links are the same as for Whites (Eccles, personal communication, 2001; Eccles, Roeser, Wigfield, & Freedman-Doan, 1999). The model has yet to be tested with other ethnic groups. The differences between groups will probably lie not in the processes specified by the model, but in average levels on some constructs, such as math self-confidence or value attached to taking an advanced math course.

### Conclusions: Motivation

We have reviewed here a half-century of research on motivation, the last 30 years of it more informed by feminist approaches and by women researchers. Feminist psychology has made several crucial contributions to this research. First, it unmasked the males-only theorizing on achievement motivation by McClelland and Atkinson (Spence, 1983; Stewart & Chester, 1982). Second, feminist psychologist Matina Horner proposed the Motive to Avoid Success as a psychological explanation for women's lesser achievements (1969), but then other feminist psychologists pointed out the flaws in the theory and the research methods; the theory was subsequently dropped. Thus feminist psychologists, like all good scientists, can self-correct and can falsify poorly formulated theories.

Feminist psychologists emphasize the importance of the situation or context and note gender similarities as well as gender differences and, in particular, note similarities in women's and men's achievement goals. Feminist researchers argue that not only gender, but race/ethnicity, social class, and disability are crucial to understanding motivation and that motivation research—ranging from McClelland's achievement motivation to

contemporary research on achievement goals—has been based almost exclusively on White, middle-class samples. Finally, and perhaps most importantly, feminist psychologist Jacquelynne Eccles has formulated an alternative theory, which is supported by massive amounts of data, that examines achievement-related choices in terms of expectations for success and task value, both of which are powerfully influenced by the forces of gender-role socialization. This theory provides many of the answers we need in understanding the underrepresentation of women in math- and science-related occupations.

### Achievements

In this section we consider research on gender and achievements in several areas: achievement tests, grades in school, and earning advanced degrees.

An important task of feminist psychology is to challenge stereotypic ideas about gender and test the stereotypes against data. Over the years, many achievement-related stereotypes have been challenged by feminist research. For example, long-standing stereotypes hold that boys and men are better at math, whereas girls and women are better at verbal tasks (Hyde et al., 1990; Swim, 1994). Feminist researchers and others concerned about equity in education have spent a great deal of energy questioning whether these stereotypes reflect reality. Some of these researchers have examined well-sampled, national data sets and others have used the technique of meta-analysis.

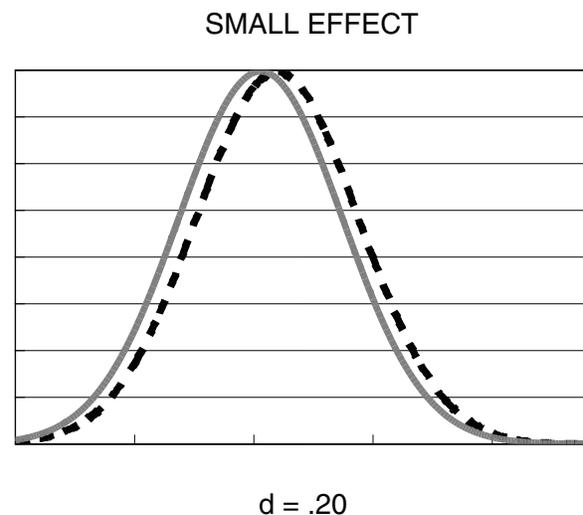
To perform a meta-analysis, the researchers collect all of the studies that they can find on a particular topic, for example, gender differences in mathematics performance. The statistics from each study are extracted and, for each study, an effect size,  $d$ , is computed, which reflects the magnitude of the gender difference. An effect size is the difference between the mean score for males and the mean score for females, divided by the average within-sex standard deviation. The effect sizes from all studies are then averaged to generate an effect size across all studies. In interpreting the magnitude of effect sizes, the following convention is often used: a  $d$  value of .20 is small, .50 is moderate, and .80 is large (Cohen, 1969). When an effect size for a gender difference is close to zero, the researcher may conclude that there is no gender difference, countering erroneous gender stereotypes.

### Gender and Mathematics Test Performance

The question of gender differences in math performance has been studied for decades. Using the technique of meta-analysis, feminist researcher Janet Hyde, with her colleagues Elizabeth Fennema and Susan Lamon (1990) located more than 100 studies representing the testing of over three million test-takers. The overall gender difference in math performance was .20, a small difference favoring males. The effect size varied considerably, however, depending on age and the cognitive level of the test. In elementary school, for tests measuring simple computa-

tion,  $d = -.20$ , indicating a small difference favoring girls. In understanding of mathematical concepts in the high school years,  $d = 0.07$ , indicating no gender difference. And for complex problem solving in the high school years,  $d = 0.29$ , indicating a small-to-moderate difference favoring males.

The pattern of results revealed by this meta-analysis challenges the stereotypes that males are better than females at math in a number of ways. First, the variety of skills that were assessed in the studies that were reviewed demonstrate that math performance is not a unitary construct and that gender differences vary according to the type of skill assessed. Global statements about male superiority at math can and should be rejected. Second, when only studies of the general population were analyzed, excluding selective samples, the gender difference actually favors females, although only by a slight amount,  $d = -.05$ , again challenging the stereotype. Third, the magnitude of the effect size indicates that there is substantial overlap between the male and female distributions (Figure 2). In other words, the similarities between males and females far outweigh the differences, and within-gender variations are far greater than between-gender differences. Fourth, the age trend showed that girls perform better than boys in elementary school. Not only does this finding provide a counterexample to the stereotype of male superiority at math, it suggests that developmental experiences contribute to patterns of gender differences in math performance. For example, as students grow older, they are given more freedom to choose their classes.



**Fig. 2.** Two normal distributions that are 0.20 standard deviations apart (i.e.,  $d = 0.20$ ). This is the approximate magnitude of the gender in mathematics performance. (Source: Hyde, J. S., Fennema, E., & Lamon, S. J. [1990]. Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, 107, 139–155. Figure 1, p. 149).

Investigations into gendered patterns of course-taking in high school suggest that the tendency for girls to take fewer math classes is partially responsible for their lower scores on standardized math tests (Kimball, 1989). Although historically males have taken more math classes than females (Armstrong, 1981; Elmore & Vasu, 1986; Fennema & Sherman, 1977), data from the 1999 SAT cohort shows that females were more likely than males to have taken four years of math classes (56% vs. 44%, College Board Online, 1999), at least for the academically talented group that takes the SAT. In sum, this meta-analysis challenges the stereotypes that men are better at math. However, many of the findings require a nuanced understanding of research and of the nature of mathematics performance that may be lacking in those who do not have advanced training in psychology or education.

Other researchers have examined data from well-sampled, national studies (e.g., Hedges & Nowell, 1995; Willingham & Cole, 1997). For the High School and Beyond sample, the magnitude of the gender differences was 0.22, a small difference (Hedges & Nowell, 1995). Looking across multiple such data sets, Willingham and Cole (1997) showed that among American 17-year-olds, girls perform better on tests assessing computational skills, whereas boys perform better on math tests that assess understanding of concepts.

Hedges and Nowell supplemented their analysis with an examination of gender differences in variability. For the most part, male scores were more variable than female scores. They demonstrated that the larger male variance, when combined with a small difference in average scores favoring males, results in lopsided gender ratios in the tails of the distributions. For example, the math test administered to the High School and Beyond sample yielded an effect size of .22 favoring males. In this sample, the male distribution was significantly more variable than the female distribution. If excellence on this test was defined as scoring in the top 5% of the distribution, then twice as many boys as girls would meet this criteria. Because test scores are often used as criterion, the demonstration of the gender inequity is extremely important. No one, to our knowledge, has successfully explained why males are more variable.

The standardized math test that receives the most attention, in terms of both press coverage and advance preparation on the part of students, is the Scholastic Assessment Test, or SAT (formerly known as the Scholastic Aptitude Test). In the year 1999 alone, over 1.2 million students took the SAT as part of their preparation to apply to college. Part of the appeal of using the SAT to make decisions about students is that the test provides a way to compare students who come from vastly different schools. Because results from this test are used in both admission decisions and for scholarship allocation, the test has been described as a "high-stakes" test-taking situation. Males score higher on the math section of the SAT, and the size of the gender gap has remained generally consistent since 1972 (College Board Online, 1999). In 1999, the

average scores on the math SAT were 531 and 495 for males and females, respectively. The standard deviation was 114, yielding an effect size of .32 (College Board Online, 1999). This effect is noticeably larger than those obtained through meta-analysis and national databases. A discussion of potential reasons for this discrepancy is presented in a later section that considers reasons for gender differences on the SAT as a whole.

#### *Gender and Verbal Skills*

Another achievement-related stereotype is that women have better verbal skills (Swim, 1994). As with math skills, both meta-analysis and national databases have been used to examine whether this stereotype is true. A meta-analysis by feminist psychologists Janet Hyde and Marcia Linn (1988) located 165 relevant studies. Averaged over all studies,  $d = -.11$ , a difference favoring females that is so small that it could be considered no difference. The magnitude and even the direction of the gender difference depended on the kind of verbal skill that was assessed. For example, the gender difference favoring females was more pronounced for speech production ( $d = -.33$ ) and anagrams ( $d = -.22$ ). However, males performed better on tests of analogies ( $d = 0.16$ ).

Data from the National Assessment of Educational Progress (NAEP), a well-sampled, national data set, indicate that the largest female advantage appears for tests of writing,  $d = 0.57$  (Willingham & Cole, 1997). Females also scored better on tests of language use (e.g., grammar and spelling) with an effect size around 0.40 across numerous national studies (see Hedges & Nowell, 1995, for roughly similar results). In summary, the national data from American students consistently show that females have better verbal skills, although the differences are often small and the magnitude of the difference varies according to the type of skill that is assessed.

These findings suggest that women should score higher on the verbal section of the SAT and, until the 1970s, they did. When the verbal section of the test was revised in 1972, changes were implemented in the content that eliminated the female advantage (Linn, 1992; Petersen & Dubas, 1992). In fact, since 1972, males have consistently scored higher, on average, than females on the verbal section of the SAT, although the difference is small. For example, in 1999, the average scores on the verbal section of the SAT were 509 and 502 for males and females, respectively; with a standard deviation of 111, the effect size was 0.06 (College Board Online, 1999). This effect size is very close to zero and could be considered to be no difference. The SAT verbal score, however, is typically not considered in isolation. Instead, the SAT verbal and math scores are summed to generate an overall SAT score, which favors males because of the larger difference on the math portion.

What is notable about the effect size for the SAT-Verbal is that it is inconsistent with effects favoring females documented by meta-analysis and in well-sampled national data sets.

### *Potential Explanations for Gender Differences on the SAT*

Many possible factors have been suggested as potential explanations for gender differences on the SAT. First, those who take the SAT are a selective sample, and samples selected for higher ability tend to display a larger gender difference favoring males in math (Hyde et al., 1990). In 1999, 44% of the students who took the SAT reported that they ranked in the top 20% of their high school class (College Board Online, 1999). Such selectivity, combined with greater male variability, results in higher average scores for men (Hedges & Nowell, 1995). Second, women are more likely to take the SAT than men (54% vs. 46% in 1999). Thus, the female pool of test-takers essentially goes farther down into the distribution of female talent than the male pool does. Supporting this line of reasoning, data from the 1999 SAT sample indicate that females were more likely to be disadvantaged in terms of parental income and education. Women represented 62% of the test-takers who reported family incomes less than \$20,000 and 60% of the test-takers whose parents had not earned a high school diploma. Although some suggest that these factors are sufficient to explain away the gender difference in SAT scores and that there is nothing wrong with the test itself (College Board, February 1998; June 1998), others believe that the test is inherently biased and should be either substantially revised or abandoned entirely (e.g., Lavergne, 2001; Mandula, 1990; Mann, 1997; Rebhorn & Miles, 1999; Rothstein, 2001; Schevitz, 2001). Given the ongoing controversy and the importance of SAT scores in the lives of students, feminist psychologists should continue to monitor research on the SAT and pressure the College Board to ensure a gender-fair test.

### *Gender, Grades, and Other School Outcomes*

Success in school can be measured in many ways. Using virtually any measure of such success, females consistently perform better as students than males. The most obvious measure of achievement in school is grades, and female superiority in grades, at all grade levels from elementary school to high school, has been documented repeatedly (e.g., Kimball, 1989; Willingham & Cole, 1997). For example, in the 1999 cohort of students to take the SAT, the average reported GPA for males was 3.16 compared to 3.31 for females (College Board Online, 1999) despite the fact that the females are a less selective sample than the males. Estimates of the effect size range from approximately .33 in nationally representative studies (e.g., High School and Beyond and the National Education Longitudinal Study of 1988), to approximately .17 in more selective samples (e.g., students who took the ACT and SAT in 1992; Willingham & Cole, 1997). The female advantage in college grades holds even if differences in high school grades are controlled (Astin, 1993). The gender difference in grades consistently favors females in all ethnic groups; however, data from the 1992 administration of the ACT indicate that the magnitude of the effect varies,

with the largest effect emerging in samples of African Americans students, and the smallest effect in samples of Asian Americans (Willingham & Cole, 1997). Although these data on ethnic differences are limited in their generalizability, because they are derived just from students who took the ACT in 1992, they do suggest that it is important to understand the role of both gender and ethnicity in predicting academic achievement (e.g., Mickelson, 1989).

Other indicators, aside from grades, suggest that females are better students. For example, more boys are required to repeat a year of schooling, and boys represent over two-thirds of the students in special education classes (U.S. Department of Education, 2000). In elementary school, boys are more likely to have their parents contacted about their behavior or their schoolwork (U.S. Department of Education, 2000). In high school, according to data from the 1999 SAT test-takers (College Board Online, 1999), girls are more likely to report having taken honors classes (e.g., 55% of math honors classes, 62% of English honors classes, 56% of natural sciences honors classes). Data from the 1992 assessment wave of the National Education Longitudinal Study shows that high-school girls were more likely to be on the honor roll and to be elected as class officers (Willingham & Cole, 1997). Finally, female students are more likely to belong to academic clubs and are more likely to be engaged in community service (U.S. Department of Education, 2000). In summary, female students are better students on a wide variety of indicators.

### *Women's Educational Achievements at the Postsecondary, Graduate, and Professional Levels*

Women have made remarkable advances in higher education over the last three decades. In 1970, women earned 43% of the bachelor's degrees conferred; by 1996, the figure had risen to 55% (U.S. Department of Education, 2000). In addition to attending college at higher rates, women have made large inroads into areas that were previously male-dominated. For example, in 1970, women earned 9% of the undergraduate business degrees, but earned 48% of them in 1996. Regarding bachelor's degrees in biological or life sciences, women earned about one-third of such degrees in 1970 (30%); the figures for 1996, however, show that women earned the majority of such degrees (53%, U.S. Department of Education, 2000).

Women have also made remarkable progress in earning graduate and professional degrees in some areas. Women represented 30% of full-time graduate students in 1970; in 1996, the percentage had risen to 51% (U.S. Department of Education, 2000). Regarding professional degrees, in 1970 women earned fewer than 1% of the degrees in dentistry, compared with 36% in 1996. During the same time period, medical doctor degrees awarded to women rose from 8% to 41%, and the corresponding percentages for law degrees showed an increase from 5% to 44% (U.S. Department of Education, 2000).

Nonetheless, some educational specialties remain heavily male-dominated. For example, in the field of

engineering, women earned 16% of the undergraduate degrees and 13% of the doctoral degrees in 1996. Computer and information sciences also remains male-dominated, with women earning 28% of the undergraduate degrees and 15% of the doctoral degrees in 1996 (U.S. Department of Education, 2000).

#### *The Female Underprediction Effect*

As advertised, the SAT does predict grades in the first year of college. The most recent data from the College Board web site indicates that the validity coefficient calculated from a sample of over 48,000 students in the freshman class of 1995 was .52 (College Board Online, 1999). Gender effects in prediction, however, require special attention. Consider simultaneously the gender effect favoring males on the SAT and the fact that women earn higher grades in college than men (e.g., Kimball, 1989), and a problem with the SAT becomes strikingly apparent. The SAT underpredicts the grades of women, an effect that has been documented for over 25 years (Hewitt & Goldman, 1975; Linn, 1973; Stricker, Rock, & Burton, 1993; Wainer & Steinberg, 1992; Willingham & Cole, 1997). The "female underprediction effect" (FUE) means that women earn higher grades in college than their SAT scores would suggest.

Many potential factors have been offered as explanations for the FUE. Some argue that women are more likely to take classes in the humanities and that grading is easier in these classes (Hewitt & Goldman, 1975). Indeed, in two different samples of students, researchers were able to completely account for the FUE using variations in department grading standards (Elliott & Strenta, 1988; Young, 1991). However, other accounts of the FUE suggest that women are better students than men and are able to compensate for the negative bias in standardized tests through good scholastic work (Stricker, Rock, & Burton, 1993).

The practical impact of the FUE was examined in a comprehensive study of undergraduate admissions at the University of California, Berkeley. Leonard and Jiang (1999) examined the academic performance of approximately 10,000 students who were admitted between 1986 and 1988. Berkeley uses a linear combination of SAT scores and high school grade point average, called the Academic Index Score (AIS), for admissions purposes (for details on the construction of the AIS, see Leonard & Jiang, 1999, p. 381). When examining the ability of the AIS to predict grades, Leonard and Jiang controlled for discipline-related variation in grading standards by testing for the FUE separately for the largest majors on campus. Even after controlling for differences in courses taken, women earned higher grades than their AIS scores would predict in 15 of the 22 majors. As in prior studies, the magnitude of the FUE was small (.06). In the context of admission procedures at a large, highly competitive university, however, a small effect can have a detrimental impact on substantial numbers of women. According to their estimates of how well the AIS predicts grades at

Berkeley, women's scores on the AIS in their sample should have been 150 points higher than they were (AIS scores range from 2000 to 8000). Although this difference is small, the authors estimated that between 200 and 300 nonminority women were denied admission each year at that institution alone because of gender bias in the SAT.

Leonard and Jiang (1999) reviewed some of the proposed explanations for the FUE. Having ruled out differential course selection as an explanation, they noted that researchers have also suggested that female students have better study skills than their male counterparts (Stricker, Rock, & Burton, 1993). Leonard and Jiang (1999) pointed out that suggesting that women earn higher grades than their SAT scores would predict because they study harder after being admitted does not undo the discrimination that occurred at the point of admission. The logic of this argument, in fact, suggests that women who were denied admission would have earned grades that were similar to, if not better than, the men who were admitted.

Leonard and Jiang (1999) also addressed the possible ways that the system could be fixed to restore equity. They noted that the fixes that have been recommended to compensate for the gender bias in the SAT are not logically consistent. For example, in a court case against the State of New York (*Sharif v. New York State Education Department*, 1989), the state was directed to weigh high school GPA more heavily in awarding the Empire State Scholarship. Presumably, because girls earn higher grades in high school, this correction would fix the problem. However, in order to undo gender bias in the SAT in the decision process, it would be necessary to incorporate a scale that is *biased towards females* not just one on which females score better. In other words, high school grades could only be used as a corrective factor if they overpredicted college grades for women. In the Berkeley study (Leonard & Jiang, 1999), however, high school grades were unbiased in their prediction of college grades. Based on this evidence, it is "impossible to correct the gender bias in the SATs with a compensating weight for high school grade point average" (Leonard & Jiang, p. 394). On the other hand, it would be possible to correct for gender bias in the SAT by using gender-specific criteria or equations in the admissions process. For example, a female student's math performance could be scored relative to other women who took the test, rather than to the entire sample of test-takers. Unfortunately, such actions are untenable in the current political climate. Any action taken by a college or university that treats women differently, even if the spirit of the action is to restore gender equity, is likely to be challenged in court by men who were denied admission. It is at this point, argue Leonard and Jiang, that the makers of the test have an obligation to step in and take the lead:

Only if the College Board and ETS were to publicly label their tests as underpredicting women's grades, and to warn colleges that they should be used only with appropriate correctives, would admissions offi-

cers in competitive state universities be able to treat men and women applicants differently in the present legal and political climate (Leonard & Jiang, 1999, p. 401).

Leonard and Jiang raised many important issues about gender equity that should be monitored by feminist psychologists and policymakers. The empirical findings of Leonard and Jiang need to be replicated at other universities, of course. Nonetheless, the work of Leonard and Jiang (1999) raises important questions about testing and women's equal access to education.

#### *Additional Threats to Female Academic Achievement*

Aside from gender bias in the SAT, a number of other threats to female academic achievement have been identified and studied. Broadly speaking, these threats can be classified as either existing in the educational environment itself or existing internally, in the minds of female students, albeit created by social forces. Turning first to the environmental threats to female achievement, the list of potential factors is lengthy, ranging from overt sexual harassment to acts of subtle discrimination that may go unnoticed at first. The term "chilly classroom" was coined to describe a pattern of gender bias in the classroom (Sandler & Hall, 1986; Hall & Sandler, 1982). Some specific behaviors that contribute to a chilly classroom include: male students interrupting female professors more frequently than male professors (Brooks, 1982), teachers providing more specific and helpful feedback to boys (Eccles & Blumefeld, 1985; Golombok & Fivush, 1994; Sadker & Sadker, 1994), attributing boys' academic failures to motivational problems and girls' academic failures to lack of ability (Dweck, Davidson, Nelson, & Enna, 1978), and using gender-stereotypic examples such as male doctors and female secretaries. These patterns may, over time, cause female students to feel unwelcome in the classroom and to become doubtful of their own ability. Any one of these behaviors may not, in and of itself, have important negative consequences, but the small "microinequities" (Rowe, 1990) may compound over time into a profoundly different educational experience for boys and girls.

Recent feminist research into the nature of modern sexism provides a framework for understanding the wide range of behaviors that contribute to a "chilly classroom." Specifically, Glick and Fiske (1996; 1997) have researched a concept they call "Ambivalent Sexism," which consists of both "hostile" and "benevolent" sexism. Hostile sexism most closely matches what people typically think about when hearing the term sexism. That is, hostile sexism is based on the stereotypic idea that women should be subservient to men and contains an element of resentment toward the results of the women's movement. Items that are used to assess hostile sexism include, "Most women fail to appreciate fully all that men do for them," and "Women are too easily offended" (Glick & Fiske, 1996). Benevolent sexism, in contrast, is more affectively positive in nature,

yet it simultaneously consists of ideas and behaviors that are restrictive to women. An individual who is high in benevolent sexism would agree with statements such as, "Women should be cherished and protected by men," and "Women, compared to men, tend to have a superior moral sensibility" (Glick & Fiske, 1996). The distinction between benevolent and hostile sexism has the potential to be very useful in providing a more nuanced understanding of how gender is experienced in today's cultural climate.

Hostile sexism in the schools takes many forms and is experienced by both students and teachers alike. On the part of female teachers, hostile sexism can be experienced as discrimination in the hiring process and lower pay relative to men. For example, in 1993, women represented 84% of elementary school teachers and 53% of secondary school teachers (U.S. Department of Education, 2000). Although this pattern may be partially explained by gender differences in interests, it is also possible that the differences could be attributed to discrimination in the hiring process. Pay discrimination in academia has been documented; among full-time faculty at institutes of higher education, women earn 81% of what men earn (U.S. Department of Education, 2000). On the part of students, hostile sexism can take the form of sexual harassment (Murnen & Smolak, 2000; Pellegrini, 2000; Stein, 1995). The passage of Title IX in 1972 prohibited all forms of sex discrimination in the schools, including sexual harassment. In 1999, the Supreme Court heard a case involving a female student who was repeatedly victimized, both physically and verbally, while in the classroom (*Davis v. Monroe County Board of Education*, 1999). Despite repeated complaints to the school district, the administration did not take any actions to protect the girl. In their decision against the school system, the Supreme Court ruled that schools could be held liable for damages for failing to stop documented cases of sexual harassment. The decision places a greater responsibility on school systems for eliminating sexual harassment. As school systems respond to this decision, it is likely that the prevalence of sexual harassment in the schools will decrease.

Other types of behavior that have been identified in the classroom can be classified as benevolent sexism. For example, the tendency for teachers to attribute the failure of girls to lack of ability (Dweck et al., 1978) may be motivated by a desire to protect "fragile" girls from humiliation and embarrassment. The desire to protect girls might also lead teachers to call on boys more frequently when asking difficult questions in class (Sadker & Sadker, 1994). While the short-term consequences of such behavior may be protective of the feelings of girls, they also deprive girls of many of the experiences that are necessary for a good education.

Researchers who study sexism have noted that as American society has moved toward more egalitarian attitudes towards women (Spence & Hahn, 1997; Twenge, 1997) and as norms have shifted to universal condemnation of sexism, the nature of sexism has changed. In her

introduction to the report on the status of female faculty at MIT, Lotte Bailyn noted that

The key conclusion that one gets from the report is that gender discrimination in the 1990s is subtle but pervasive, and stems largely from unconscious ways of thinking that have been socialized into all of us, men and women alike. This makes the situation better than in previous decades where blatant inequities and sexual assault and intimidation were endured but not spoken of. . . . But the consequences of these more subtle forms of discrimination are equally real and equally demoralizing (MIT, 1999, p. 6).

In order for women to make progress in terms of their educational achievements, our society as a whole needs to become more aware of how nonconscious processes may contribute to sexism and how benevolent sexism can be cloaked in the form of courtesy.

Limitations on women's academic achievement come not only from the external environment, but from internalized ideas and beliefs. For example, when compared to men, women have less confidence in their own abilities, especially in domains that are stereotyped as male (Beyer & Bowden, 1997; Eccles, Wigfield, & Schiefele, 1998; U.S. Department of Education, 2000). Regarding beliefs about intelligence, when Furnham and Rawles (1995) asked a sample of undergraduates to estimate their own IQ, the estimates of females were significantly lower than the estimates provided by males (112 vs. 118). Students in this study were also asked to estimate the IQ of their parents, and a similar gender effect emerged such that the average estimate of mothers' intelligence was lower than that for fathers (108 vs. 115). Finally, Brenda Major's work on entitlement shows that when undergraduates are trained on the same task in the laboratory, women will pay themselves less for the job than men (Bylsma & Major, 1992; Major, 1989). Recent work has documented that the gender difference in entitlement to pay has persisted over the years (Desmarais & Curtis, 1997). Until women feel that they are of equal worth, it is likely that their education will be impaired. Girls, of course, are not born with these ideas. They acquire them through the processes of gender socialization. Special attention to the messages that girls receive from parents and teachers about their worth relative to boys is needed to remove this internalized barrier to women's achievement.

A new body of work on "stereotype threat" offers another example of how a combination of situational factors and internalized beliefs may influence academic achievement (e.g., Brown & Josephs, 1999; Steele 1997; Walsh, Hickey, & Duffy, 1999). As described by Claude Steele (1997), members of stereotyped groups such as African Americans are especially wary of situations in which their behavior can confirm the negative reputation that their group lacks a valued ability. The extra pressure caused by the fear of reinforcing the negative stereotype interferes with performance, resulting in lower scores.

Thus, the stereotype threat process describes how knowledge of a negative stereotype about one's own social group can lead to behavioral outcomes that reinforce the stereotype. The focus on social factors embeds this work firmly in feminist principles.

To document the influence of stereotype threat in women, Spencer, Steele, and Quinn (1999) conducted a series of studies in which they administered a difficult set of GRE math problems to undergraduates. The researchers manipulated the instructions given to the students prior to taking the test. Some students were informed that the test had, in the past, yielded significant gender effects. The researchers believed that this information would be interpreted to mean that males usually score higher on the test. In another condition, participants were informed that men and women typically perform equally well on the test. When the instructions included the information that gender differences were expected, the males in the sample subsequently scored higher than the females. However, when the instructions indicated that the test had not been known to yield gender differences, men and women performed equally well. Importantly, the GRE items were the same across the groups—only the instructions were manipulated. The authors argued that the negative influence of stereotype threat on female performance was lifted when the women were told that men and women perform equally well on the test. Presumably, the removal of stereotype threat decreased anxiety levels in the women, thereby allowing them to earn a higher score that more accurately reflected their ability.

In an important follow-up (Spencer, Steele, & Quinn, 1999), a control group was given the same difficult math test with no mention of expected gender differences. In this condition, men scored higher than women, mirroring the findings from the condition in which the participants were told that the test generally yields gender differences. Thus, in the absence of information about gender differences, undergraduates behave as though they expect men to earn higher scores than women on a difficult math test. Importantly, the control condition of the follow-up study is similar to the testing situation that students face when taking the SAT. Thus, it is likely that when taking the SAT math section, male and female students alike believe that men will score higher because the gender difference on the SAT is published so extensively. If this is true, then stereotype threat may diminish the performance of women.

As researchers have come to appreciate the power of stereotype threat, an important topic that has been studied is what contextual factors increase or decrease the likelihood that stereotype threat will operate. In one recent study, researchers demonstrated that the presence of men during a math test is sufficient to create a "threatening intellectual environment" for women (Inzlicht & Ben-Zeev, 2000, p. 365). In a series of studies, the researchers administered math tests to groups of undergraduates. In the majority condition, three female participants complet-

ed the math test at the same time. In the minority condition, one female participant completed the math test along with two male confederates. As anticipated, women in this study performed worse on the math test when men were present during the testing session. One implication of these results is that performance by women during math tests can be reduced depending on the gender composition of the testing group. Situations in which women are a minority, particularly a minority of one, may be detrimental to women's performance.

Because the process of stereotype threat can be applied to any stereotyped social group, research on the topic embodies the feminist principle of inclusion and attention to ethnicity. For example, Steele and Aronson (1995) documented the effect of stereotype threat on academically talented African Americans. In this series of studies, stereotype threat was manipulated by describing different uses for the test scores. In the diagnostic condition, the participants completed a test that they believed would yield an accurate assessment of their skills. Participants in the nondiagnostic condition were told that their test performance would yield insight into problem-solving processes. Participants in this second group were explicitly told that their ability would not be evaluated. The researchers believed that the explanation of the test as nondiagnostic would remove stereotype threat. In a series of four studies, the researchers found that the performance of African American undergraduates on difficult verbal tests was worse in the diagnostic condition and better in the non-diagnostic condition. Thus, Steele and Aronson (1995) have provided a convincing argument that the process of stereotype threat interferes with performance of African American students. More recent work has extended the stereotype threat model to Latinos (Aronson & Salinas, 1997) and individuals from low socioeconomic backgrounds (Croizet & Claire, 1998).

Feminist scholars have pointed out the importance of understanding the influence of multiple social roles (e.g., Barnett & Hyde, *in press*; Espin, 1987). A recent study on stereotype threat in Asian women (Shih, Pittinsky, & Ambady, 1999) provides an elegant example of how to study and understand the psychological complexities that arise for individuals belonging to multiple social groups. The researchers chose Asian women as participants because they have two identities, each with opposite linkages with performance in mathematics. Specifically, the researchers hypothesized that priming the participants' female identity would lower their math scores, whereas priming the participants' Asian identity would increase their math scores. Priming was manipulated using two different questionnaires that were completed prior to the administration of a difficult math test. One questionnaire highlighted the participant's gender by asking questions about the gender composition of her living arrangements (e.g., whether they preferred to live in coed or mixed sex floors). The other questionnaire primed the participants' Asian identity by focusing on their families, including questions about the languages that their family members

spoke and how many generations of their family have lived in America. In addition to these two groups, a control group was included, who answered a series of generic questions about campus life prior to taking the test. Significant group differences on the math test matched the researchers' expectations. Specifically, women in the Asian-salient group earned the highest scores, followed by women in the control group, and women in the gender-salient condition earned the lowest scores. In documenting the effects of both positive and negative stereotypes within the same study, Shih, Pittinsky, and Ambady (1999) have provided an excellent demonstration of the flexibility and power of the stereotype threat model and have demonstrated how multiple identities can influence behavior in different contexts.

The studies reviewed above show that the concept of stereotype threat has strong empirical support. The theoretical implications of stereotype threat, however, also deserve some attention. First, individuals do not have to believe that a given negative stereotype is true for it to influence their behavior (Steele, 1997). For example, if a particular woman believes that women can do as well as men in math, she may still worry that other people who hold the traditional math stereotype may interpret any failure on her part as supportive of the stereotype. In considering the feminist goal of challenging erroneous stereotypes, this reasoning suggests that it is not sufficient to arm individuals with data that counter popular negative stereotypes. In order to combat the negative impact of stereotypes effectively, stereotypes about stigmatized groups must be entirely eliminated from the public consciousness. Second, stereotype threat sets up a mutually reinforcing system that will be difficult to disrupt. That is, the fear of confirming a negative stereotype leads to behavior that confirms the stereotype. Paradoxically, then, the people most concerned about changing stereotypes of their social groups may be most likely to experience a decrement in their performance due to stereotype threat (Steele, 1997). Third, the current conditions in many academic settings are likely to foster stereotype threat. For example, being the only woman in the group is sufficient to induce stereotype threat in females while taking a math test (Inzlicht & Ben-Zeev, 2000). In addition, Inzlicht and Ben-Zeev (2000) noted that as women who are interested in math pursue their education, their gender becomes increasingly salient as the percentage of females in the classroom drops. This combination of factors is likely to inhibit the performance of the women who are in the best position to serve as models of women who do well in math. Finally, the documented gender difference in the math SAT, which is likely to be due, at least in part, to stereotype threat, has many negative consequences for women's academic achievement. To the extent that SAT scores are used as a selection criterion, women may be unjustly denied academic scholarships and admission to college. In summary, stereotype threat represents a pernicious threat to the achievement of women, as well as any other stereotyped group.

## SUMMARY AND CONCLUSIONS

We have reviewed research and theory on gender, motivation, and educational achievement. Increasing the educational achievement of Americans is one of the top priorities of the Decade of Behavior and of the nation. Educated workers are crucial to the economic growth of the nation, particularly in the post-industrial era. Education equips individuals to lead happier, more satisfying lives. Feminist psychologists have done much to provide a more accurate understanding of motivational patterns for girls and boys, men and women. They have also done much to highlight gender similarities and dispel erroneous stereotypes about mathematics performance. Feminist psychologists have contributed extensively to our understanding of barriers to women's educational achievement and have, at the same time, suggested ways to remove these barriers. It is crucial to maintain and expand this feminist perspective as we move into a new century and work toward improving educational opportunities for all.

*Initial submission: June 16, 2001*

*Initial acceptance: July 20, 2001*

*Final acceptance: July 24, 2001*

## REFERENCES

- Armstrong, J. M. (1981). Achievement and participation of women in mathematics: Results of two national surveys. *Journal for Research in Mathematics Education, 12*, 356–372.
- Aronson, J., & Salinas, M. F. (1997). *Stereotype threat, attributional ambiguity, and Latino performance*. Unpublished manuscript, University of Texas, Austin.
- Astin, A. W. (1993). *What matters in college? Four critical years revisited*. San Francisco: Jossey-Bass.
- Barnett, R. C., & Hyde, J. S. (in press). Women, men, work, and family: An expansionist theory. *American Psychologist, 56*.
- Barron, K. E., & Harackiewicz, J. M. (2001). Achievement goals and optimal motivation: Testing multiple goal models. *Journal of Personality and Social Psychology, 80*, 706–722.
- Benbow, C. P., & Stanley, J. C. (1982). Consequences in high school and college of sex differences in mathematical reasoning ability: A longitudinal perspective. *American Educational Research Journal, 19*, 598–622.
- Beyer, S., & Bowden, E. M. (1997). Gender differences in self-perceptions: Convergent evidence from three measures of accuracy and bias. *Personality and Social Psychology Bulletin, 23*, 157–172.
- Bouffard, T., Boisvert, J., Vezeau, C., & Larouche, C. (1995). The impact of goal orientation on self-regulation and performance among college students. *British Journal of Educational Psychology, 65*, 317–329.
- Brooks, V. R. (1982). Sex differences in student dominance behavior in female and male professors' classrooms. *Sex Roles, 8*, 683–690.
- Brown, R. P., & Josephs, R. A. (1999). A burden of proof: Stereotype relevance and gender differences in math performance. *Journal of Personality & Social Psychology, 76*, 246–257.
- Bylsma, W., & Major, B. (1992). Two routes to eliminating gender differences in personal entitlement: Social comparisons and performance evaluations. *Psychology of Women Quarterly, 16*, 193–200.
- Caplan, M. (1969). The relationship between self-concept and academic achievement. *Journal of Experimental Education, 37*, 13–16.
- Cherry, F., & Deaux, K. (1978). Fear of success versus fear of gender-inappropriate behavior. *Sex Roles, 4*, 97–102.
- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York, NY: Academic Press.
- College Board. (February, 1998). SAT and gender differences. *Research Summary, RS-04*.
- College Board. (June, 1998). Common sense about SAT score differences and test validity. *Research Notes, RN-01*.
- College Board Online. (1999). 1999 College-Bound Seniors, National Report. Retrieved from <http://www.collegeboard.org/sat/cbsenior/yr1999/NAT/cbs1999.html>.
- Croizet, J.-C., & Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin, 24*, 588–594.
- Dabul, A. J. (1995). *Understanding achievement motivation and behavior as a function of cultural values and the social context*. Unpublished doctoral dissertation, Arizona State University.
- Dabul, A. J., & Russo, N. F. (1996). Rethinking psychological theory to encompass issues of gender and ethnicity: Focus on achievement. In K. F. Wyche & F. J. Crosby (Eds.), *Women's ethnicities: Journeys through psychology* (pp. 183–199). Boulder, CO: Westview.
- Davis v. Monroe County Board of Education*, No. 97-843, 1999 U.S.
- Desmarais, S., & Curtis, J. (1997). Gender and perceived pay entitlement: Testing for effects of experience with income. *Journal of Personality and Social Psychology, 72*, 141–150.
- Dweck, C., Davidson, W., Nelson, S., & Enna, B. (1978). Sex differences in learned helplessness: II. The contingencies of evaluation feedback in the classroom. III. An experimental analysis. *Developmental Psychology, 14*, 268–276.
- Eccles, J. S. (1987). Gender roles and women's achievement-related decisions. *Psychology of Women Quarterly, 11*, 135–172.
- Eccles, J. S. (1994). Understanding women's educational and occupational choices: Applying the Eccles et al. model of achievement-related choices. *Psychology of Women Quarterly, 18*, 585–610.
- Eccles (Parsons), J. S., Adler, T., & Meece, J. L. (1984). Sex differences in achievement: A test of alternate theories. *Journal of Personality and Social Psychology, 46*, 26–43.
- Eccles, J. S., & Blumefeld, P. (1985). Classroom experiences and student gender: Are there differences and do they matter? In L. C. Wilkinson & C. B. Marrett (Eds.), *Gender influences in the classroom* (pp. 79–113). New York: Academic Press.
- Eccles, J. S., & Harold, R. D. (1992). Gender differences in educational and occupational patterns among the gifted. In N. Colangelo, S. G. Assouline, & D. L. Ambrosio (Eds.), *Talent development: Proceedings from the 1991 Henry B. and Jocelyn Wallace national research symposium on talent development* (pp. 3–29). Unionville, NY: Trillium Press.
- Eccles, J. S., Roeser, R., Wigfield, A., & Freedman-Doan, C. (1999). Academic and motivational pathways through mid-

- dle childhood. In L. Balter & C. Tamis-LeMonda (Eds.), *Child psychology: A handbook of contemporary issues* (pp. 287–317). Philadelphia: Psychology Press/Taylor & Francis.
- Eccles, J. S., Wigfield, A., Harold, R. D., & Blumenfeld, P. (1993). Ontogeny of children's self-perceptions and subjective task values across activity domains during the early elementary school years. *Child Development, 64*, 830–847.
- Eccles, J. S., Wigfield, A., & Schiefele, U. (1998). Motivation to succeed. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology. Vol. 3, Social, emotional, and personality development*. 5th ed., NY: Wiley.
- Elliott, R., & Strenta, A. C. (1988). Effects of improving the reliability of the GPA on prediction generally and on comparative predictions for gender and race particularly. *Journal of Educational Measurement, 25*, 333–347.
- Elmore, P. B., & Vasu, E. S. (1986). A model of statistics achievement using spatial ability, feminist attitudes and mathematics-related variables as predictors. *Educational and Psychological Measurement, 46*, 215–222.
- Espin, O. M. (1987). Issues of identity in the psychology of Latina lesbians. In Boston Lesbian Psychologies Collective, *Lesbian psychologies*. Urbana, IL: University of Illinois Press.
- Fennema, E., & Sherman, J. (1977). Sex related differences in mathematics achievement, spatial visualization and affective factors. *American Educational Research Journal, 14*, 51–71.
- Fineran, S., & Bennett, L. (1999). Gender and power issues of peer sexual harassment among teenagers. *Journal of Interpersonal Violence, 14*, 626–641.
- Furnham, A., & Rawles, R. (1995). Sex differences in the estimation of intelligence. *Journal of Social Behavior and Personality, 10*, 741–748.
- Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology, 70*, 491–512.
- Glick, P., & Fiske, S. T. (1997). Hostile and benevolent sexism: Measuring ambivalent sexist attitudes towards women. *Psychology of Women Quarterly, 21*, 119–135.
- Golombok, S., & Fivush, R. (1994). *Gender development*. New York: Cambridge University Press.
- Graham, S. (1994). Motivation in African Americans. *Review of Educational Research, 64*, 55–117.
- Gutek, B. A. (2001). Women and paid work. *Psychology of Women Quarterly, 25*, 379–393.
- Hall, R. M., & Sandler, B. R. (1982). *The classroom climate: A chilly one for women?* Washington, DC: Association of American Colleges, Project on the Status and Education of Women.
- Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high scoring individuals. *Science, 269*, 41–45.
- Hewitt, B. N., & Goldman, R. D. (1975). Occam's razor slices through the myth that college women overachieve. *Journal of Educational Psychology, 67*, 325–330.
- Hoffman, L. W. (1972). Early childhood experiences and women's achievement motives. *Journal of Social Issues, 28*(2), 129–155.
- Horner, M. S. (1969). Fail: Bright women. *Psychology Today, 3*(6), 36.
- Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*, 139–155.
- Hyde, J. S., Fennema, E., Ryan, M., Frost, L. A., & Hopp, C. (1990). Gender comparisons of mathematics attitudes and affect: A meta-analysis. *Psychology of Women Quarterly, 14*, 299–324.
- Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104*, 53–69.
- Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science, 11*, 365–371.
- Jacobs, J. E. (1991). Influence of gender stereotypes on parent and child ability beliefs in three domains. *Journal of Personality and Social Psychology, 63*, 932–944.
- Jenkins, S. R. (1987). Need for achievement and women's careers over 14 years: Evidence for occupational structure effects. *Journal of Personality and Social Psychology, 53*, 922–932.
- Kessler, R. C., & Cleary, P. D. (1980). Social class and psychological distress. *American Sociological Review, 45*, 463–478.
- Kimball, M. M. (1989). A new perspective on women's math achievement. *Psychological Bulletin, 105*, 198–214.
- Lavergne, G. M. (2001, March 4). Is this the end for the SAT? [Letter to the editor]. *The New York Times*, Section 4, p. 15.
- Leonard, D. K., & Jiang, J. (1999). Gender bias and the college prediction of the SATs: A cry of despair. *Research in Higher Education, 40*, 375–407.
- Linn, M. (1992). Gender differences in educational achievement. In *Sex equity in educational opportunity, achievement and testing: Proceedings of the 1991 ETS invitational conference* (pp. 11–50). Princeton, NJ: Educational Testing Service.
- Linn, R. L. (1973). Fair test use in selection. *Review of Educational Research, 43*, 139–161.
- Major, B. (1989). Gender differences in comparisons and entitlement: Implications for comparable worth. *Journal of Social Issues, 45*(4), 99–115.
- Mandula, B. (1990). Is the SAT unfair to women? *AWIS Newsletter, 19*(1), 8–10.
- Mann, J. (March 26, 1997). SAT fails the bias test again. *The Washington Post*, p. E13.
- Massachusetts Institute of Technology. (1999). *A study on the status of women faculty in science at MIT*. Cambridge, MA: Author. Retrieved from <http://web.mit.edu/fnl/women/women.html>, June 23, 2001.
- McClelland, D. (1966). Longitudinal trends in the relation of thought to action. *Journal of Consulting Psychology, 30*, 479–483.
- McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, F. L. (1953). *The achievement motive*. New York: Appleton-Century-Crofts.
- Mednick, M. T. (1989). On the politics of psychological constructs: Stop the bandwagon, I want to get off. *American Psychologist, 44*, 1118–1123.
- Mednick, M. T., & Thomas, V. G. (1993). Women and the psychology of achievement: A view from the eighties. In F. L. Denmark & M. A. Paludi (Eds.), *Psychology of women: A handbook of issues and theories* (pp. 585–626). Westport, CT: Greenwood.
- Mickelson, R. A. (1989). Why does Jane read and write so well? The anomaly of women's achievement. *Sociology of Education, 62*, 47–63.

- Midgley, C., Kaplan, A., & Middleton, M. (2001). Performance-approach goals: Good for what, for whom, under what circumstances, and at what cost? *Journal of Educational Psychology, 93*, 77–86.
- Miech, R. A., Caspi, A., Moffitt, T. E., Wright, B., & Silva, P. A. (1999). Low socioeconomic status and mental disorders: A longitudinal study of selection and causation during young adulthood. *American Journal of Sociology, 104*, 1096–1131.
- Monahan, L., Kuhn, D., & Shaver, P. (1974). Intrapsychic versus cultural explanations of the “fear of success” motive. *Journal of Personality and Social Psychology, 29*, 60–64.
- Murnen, S. K., & Smolak, L. (2000). The experience of sexual harassment among grade-school students: Early socialization of female subordination? *Sex Roles, 43*, 1–17.
- Pellegrini, A. D. (2001). A longitudinal study of heterosexual relationships, aggression, and sexual harassment during the transition from primary school through middle school. *Journal of Applied Developmental Psychology, 22*, 119–133.
- Petersen, A. C., & Dubas, J. (1992). Strategies for achieving sex equity in postsecondary education. In *Sex equity in educational opportunity, achievement and testing: Proceedings of the 1991 ETS invitational conference* (pp. 119–135). Princeton, NJ: Educational Testing Service.
- Pintrich, P. R., & Schunk, D. H. (1996). *Motivation in education: Theory, research and applications*. Englewood Cliffs, NJ: Merrill/Prentice Hall.
- Rebhorn, L. S., & Miles, D. D. (1999). High-stakes testing: Barrier to gifted girls in mathematics and science. *School Science and Mathematics, 99*, 313–319.
- Rothstein, R. (March 28, 2001). The SAT debate ought to be broader. *The New York Times*, p. B8.
- Rowe, M. (1990). Barriers to equality: The power of subtle discrimination to maintain unequal opportunity. *Employee Responsibilities and Rights Journal, 3*, 153–163.
- Sadker, M., & Sadker, D. (1994). *Failing at fairness: How our schools cheat girls*. New York, NY: Touchstone.
- Sandler, B. R., & Hall, R. (1986). *The campus climate revisited: Chilly for women faculty, administrators and graduate students*. Washington, DC: Association of American Colleges and Universities.
- Schevitz, T. (February 17, 2001). UC Chief wants to drop SAT for admissions. *The San Francisco Chronicle*, p. A1.
- Sharif v. New York State Education Department*, 709 F.Supp 345 (SDNY, 1989).
- Shaver, P. (1976). Questions concerning fear of success and its conceptual relatives. *Sex Roles, 2*, 305–320.
- Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10*, 80–83.
- Spence, J. T., & Hahn, E. D. (1997). The Attitudes Toward Women Scale and attitude change in college students. *Psychology of Women Quarterly, 21*, 17–34.
- Spence, J. T., & Helmreich, R. L. (1983). Achievement-related motives and behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives: Psychological and sociological approaches* (pp. 7–74). San Francisco: Freeman.
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women’s math performance. *Journal of Experimental Social Psychology, 35*, 4–28.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52*, 613–629.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*, 797–811.
- Stein, N. (1995). Sexual harassment in school: The public performance of gendered violence. *Harvard Educational Review, 65*(2), 145–162.
- Stewart, A. J., & Chester, N. L. (1982). The exploration of sex differences in human social motives: Achievement, affiliation, and power. In A. J. Stewart (Ed.), *Motivation and society* (pp. 172–218). San Francisco: Jossey-Bass.
- Stricker, L. J., Rock, D., & Burton, N. W. (1993). Sex differences in predictions of college grades from scholastic aptitude test scores. *Journal of Educational Psychology, 85*, 710–718.
- Swim, J. K. (1994). Perceived versus meta-analytic effect sizes: An assessment of the accuracy of gender stereotypes. *Journal of Personality & Social Psychology, 66*(1), 21–36.
- Tresemmer, D. (1977). *Fear of success*. New York: Plenum.
- Triandis, H. C., Bontempo, R., Villareal, M. J., Asai, M., & Lucca, N. (1988). Individualism and collectivism: Cross-cultural perspectives on self-ingroup relationships. *Journal of Personality and Social Psychology, 54*, 323–338.
- Twenge, J. M. (1997). Attitudes toward women, 1970–1995. *Psychology of Women Quarterly, 21*, 35–51.
- Tyler, L. E. (1965). *The psychology of human differences*. New York: Appleton-Century-Crofts.
- U.S. Department of Education/National Center for Education Statistics. (2000). *Trends in educational equity for girls & women*, NCES 2000-030, by Y. Bae, S. Choy, C. Geddes, J. Sable, & T. Snyder. Washington, DC: Author.
- U.S. Department of Labor. (1997). *Black women in the labor force*. Washington, DC: U.S. Department of Labor, Women’s Bureau. Retrieved from <http://www.dol.gov/dol/wb>, June 15, 2001.
- U.S. Department of Labor. (2000a). *Women’s jobs: 1964–1999*. Washington, DC: U.S. Department of Labor, Women’s Bureau. Retrieved from <http://www.dol.gov/dol/wb>, June 15, 2001.
- U.S. Department of Labor. (2000b). *20 facts on women workers*. Washington, DC: U.S. Department of Labor, Women’s Bureau. Retrieved from <http://www.dol.gov/dol/wb>, June 15, 2001.
- Veroff, J., Depner, C., Kukla, R., & Douvan, E. (1980). Comparison of American motives: 1957 versus 1976. *Journal of Personality and Social Psychology, 39*, 1004–1013.
- Wainer, H., & Steinberg, L. S. (1992). Sex differences in performance on the mathematics section of the Scholastic Aptitude Test: A bidirectional validity study. *Harvard Educational Review, 62*, 323–336.
- Walsh, M., Hickey, C., & Duffy, J. (1999). Influence of item content and stereotype situation on gender differences in mathematical problem solving. *Sex Roles, 41*, 219–240.
- Willingham, W., & Cole, N. (1997). *Gender and fair assessment*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Wylie, R. (1963). Children’s estimates of their schoolwork ability, as a function of sex, race, and socioeconomic level. *Journal of Personality, 31*, 203–224.
- Young, J. (1991). Gender bias in predicting college academic performance: A new approach using item response theory. *Journal of Educational Measurement, 28*, 37–47.
- Zuckerman, M., & Wheeler, L. (1975). To dispel fantasies about the fantasy-based measure of fear of success. *Psychological Bulletin, 82*, 932–946