Gender Differences in Cognition and Educational Performance

Introductory article

Marcia C Linn, University of California, Berkeley, California, USA
Cathy Kessel, University of California, Berkeley, California, USA

CONTENTS

Gender differences in cognitive and educational performance refer to the apparent differences in the cognitive abilities of men and women, especially with respect to spatial and verbal tasks. They have been ascribed to both cultural and biological influences.

HISTORICAL AND CULTURAL INFLUENCES

Meaningful interpretations of gender differences in cognitive and educational performance require a knowledge of the relevant cultural, societal, and historical contexts. In the United States, as in other countries, women's participation in education has indeed been a function of culture and history, as well as economic need and social justice. Women began to gain access to selected institutions of higher education by the late 1800s, although many still believed that study might overtax females by drawing blood away from their ovaries to their brains. Research discredited this belief, but it resonated with popular opinion, patterns of participation in intellectual pursuits, and the images of intellectual women as spinsters. When women started to enroll as undergraduate and graduate students, they tended to earn higher grades than men – a trend that has continued to the present.

Almost a century later, when Title IX came into effect in the United States, most colleges lifted their quotas for women. Now almost every institution of higher education admits both men and women; few professional programs retain gender quotas. As a result, over half of undergraduates are women, and graduation rates of women from high school and college have soared and surpassed those of men. Females comprise an increasing proportion of students in professional schools including medicine, law, and business. National Science Foundation statistics show that almost half of all PhDs are earned by women, and over a third of the doctorates in science and engineering go to women.

The picture is less rosy with regard to employment. Cultural expectations concerning who can succeed, especially in positions of leadership, impact on selection decisions. However, exceptions are widespread. Men have excelled in fields traditionally viewed as feminine, including nursing and office administration, and women have succeeded in endeavors traditionally viewed as male, including conducting orchestras, leading technology companies, and winning intellectual prizes. Nevertheless, progress in these areas has been slow for nontraditional participants, and, in some cases such as computer science, initial gains have subsided.

The subtle but powerful pressures of cultural expectations, giving the work of females less value, are apparent, for example, even at the prestigious Massachusetts Institute of Technology. Female faculty jointly assessed the space they were given, the treatment of their requests for research funds, the assignment of courses, and the assignment of committee responsibilities. Comparing practices across departments revealed systematic and substantial differences in treatment for female compared with male faculty. MIT publicized the discrepancies and remedies in an effort to increase equitable opportunities for women at all institutions. Despite this publicity most other institutions have resisted similar analyses and possible remedies.

Individuals in power may be slow to welcome those who do not resemble themselves. As described, for example by Virginia Valian, individuals who are different from their peers frequently experience difficulties in succeeding and often are more closely scrutinized than their more traditional peers. These conditions are exacerbated by expectations that tend to place greater responsibility for family and children on females than on males, reducing available female human capital.

Cultural expectations may affect performance as well as selection. The psychological research of Claude Steele and his colleagues has demonstrated how context may make these expectations salient for particular groups. When students of a subgroup that traditionally performs poorly in a given domain are reminded of this fact, they tend to perform poorly on difficult tests; without this reminder, they perform similarly to their peers. These and other findings suggest that cultural expectations permeate all social interactions, including those in research laboratory settings, and may influence outcomes of experimental and classroom studies in subtle and unexpected ways.

Cultural expectations may affect educational participation and performance as well. For example, in 1980 Geoffrey Driver compared the academic performances of students who were white and those of West Indian origin. He found that although they experienced the same general academic program, when tested at age 16, West Indians tended to outperform their English counterparts, English boys tended to outperform English girls – but West Indian girls tended to outperform West Indian boys. A possible explanation is that the West Indian girls assumed a family structure more prevalent in the West Indies in which women are responsible for family subsistence – and were preparing for their futures as primary wage-earners.

Cultural expectations may explain differences in participation found by a Public Policy of California study. In California, middle school girls' course enrollment was greater than boys' for English, foreign language, mathematics, science (except that required for college), social science, but not for computer science courses. The situation was similar for college preparatory high school courses. Female enrollment was greater than that of males – again, except for computer science. This enrollment pattern reverses the earlier gender gap in mathematics and science courses. Moreover, in each ethnic group in the study categories, girls enrolled in more mathematics and science courses than boys. However, gaps were far larger for Filpinos, Hispanics, and African Americans than for whites and non-Filipino Asians.

Cultural expectations and instructional practices may explain the findings of David Byrnes and his colleagues. In contrast with long-standing experience in the United States, no gender gap was found when the SAT was administered to a sample of Chinese high school students. Instructional practices, but perhaps not cultural expectations, may be responsible for a similar result for high school students enrolled in an innovative mathematics course in the United States.

In the past, cultural interpretations of gender differences often followed a deficit model. From this perspective, male behavior was considered normative and desirable, and deviations from this norm were considered deficits. This model also influenced research. Studies looked for differences between males and females and accumulated less successful performances of females to paint a picture of females as having 'deficits'. This deficit model was sometimes elaborated into a compensation model, in which typical male behavior was the norm in 'male' activities, including academic, intellectual, and athletic performance, and female behavior was the norm in activities typed as female, including child care, family relations, and homemaking. These models drew attention away from the large overlap in the distributions of male and female performance in every area, and neglected the rapid improvements that result when barriers to participation or success disappear.

SPATIAL AND VERBAL ABILITIES

Historically, accounts and definitions of human abilities have varied substantially. In the United States at the beginning of the twentieth century, research views of human ability as unitary were reinforced by measures of general ability that yielded a single score. As researchers performed more detailed analyses, they identified more specific and often contextualized abilities. A substantial commercial enterprise in the design, sale, and scoring of tests for these abilities developed.

Some of this proliferation in abilities stemmed from newly available statistical techniques. As factor analysis developed, individual researchers used this technique to distinguish more and more abilities. At the same time, a research program concerning the nature of ability constructs took shape. The identification, statistical properties, and labeling of human ability constructs has continued as an active research area. Researchers seek to distinguish 'traits' that are innate or unresponsive to instruction from achievements that respond to instruction. Good methods for determining whether an ability is inherent or responsive to instruction have been difficult to develop. Sorting out the effects of cultural expectations on human abilities has proven problematic. When cultural expectations limit opportunity to learn, constructs often appear innate. For example, many assumed there are innate differences in spatial ability only to discover that performance responds to instruction and varies across educational systems. Similar differences in mathematical attainment have narrowed, as Rosenthal observed, 'faster than the gene can travel'. A serious debate about the inherent nature of abilities plays out in each of the areas highlighted in this review.

Spatial Performance

Research on spatial ability has its origins in Francis Galton's work on imagery in the 1880s. Findings concerning gender differences soon followed. Galton found that 'scientific men as a class have feeble powers of visual representation. There is no doubt whatever on the latter point, however it may be accounted for'. G. Stanley Hall reported in his 1891 study, 'The contents of children's minds on entering school', that girls excelled in space concepts and boys in number. Systematic research on this topic began in the 1920s.

Tests of spatial reasoning such as Raven's progressive matrices were viewed as independent of verbal skill. However, using factor analysis to distinguish verbal and nonverbal constructs proved difficult. Raymond Cattell argued that fluid and crystallized ability, although distinct constructs, were correlated factors. Despite these difficulties, by World War II spatial ability measurement played an important role in assessing and assigning army recruits, and measures of spatial ability were used in other career decisions. Eventually, the belief that females had greater difficulty in spatial reasoning was associated with their underrepresentation in mathematics and science.

In 1974, Eleanor Maccoby and Carol Jacklin reviewed all the research on gender differences and concluded that, although performance across tests used to measure spatial ability varied considerably, in general males excelled in spatial tasks. These tests included Raven's progressive matrices, the embedded figures test, and a newly developed measure of speed in mental rotation.

In the late 1970s, Marcia Linn and Anne Petersen extended Maccoby and Jacklin's findings on spatial performance. They sought to distinguish the varied constructs captured under the broad category of spatial ability. They used meta-analysis to group tasks in categories with uniform characteristics and identified three constructs.

First, syntheses of performances on tasks such as embedded figures, paper folding, and Raven's progressive matrices that asked students to reason about figural rather than verbal information revealed virtually no gender differences.

Second, tasks measuring what is often referred to as the *water level task* revealed large gender differences. This task, originally used by Jean Piaget, asks students to predict the level of water in a tilted glass when given information about the water level in an upright glass. Many factors can influence performance, including whether the glass is square or curved, and whether the respondent believes that the glass has just been moved to a new location or has been resting in that new location for some time.

Third, tasks requiring students to mentally rotate two- and three-dimensional objects to a new orientation revealed large differences, primarily in speed. Most respondents obtained the right answer; males responded more rapidly. Some respondents took over twice as long as others, and the studies of strategies for mental rotation showed that some participants mentally rotated an entire object and chose an answer, while others rotated features and repeated the process until the answer was chosen. In these studies, women were more likely than men to use a feature matching approach, and therefore to take longer to respond to mental rotation items. Changing the instructions to encourage rapid responding reduces differences, supporting the idea that a strategy preference is a factor in these differences.

Subsequent research has shown that spatial ability, rather than being an enduring trait, appears quite responsive to instruction and, for instance, multiple strategies can be used to solve most spatial reasoning tasks. For example, Sherry Hsi and Alice Agogino studied spatial reasoning performance among first-year engineering students at a competitive university, a group likely to have been successful in high school mathematics classes. They found large initial differences between males and females on measures of spatial reasoning. Two Saturday morning voluntary training sessions reduced or eliminated these gender differences.

Moreover, in-depth interviews of successful engineers revealed that they rarely used the rotation of whole objects in their work. Instead, most reported that they were likely to rely on descriptive geometry, feature matching, and the rotation of a subset of familiar shapes and forms, rather than using their mental capacities to rotate whole objects. This suggests that success in engineering may not require the use of mental rotation.

Another area where spatial reasoning ability is frequently implicated concerns the interpretation of maps. Here again, cultural expectations and anecdotal evidence suggest that males are more successful at way-finding, map reading, and map interpretation. Nevertheless, extensive study of young children's ability to find their way in complex settings reveals a fairly systematic result of males being more successful than females. A study of adults found that males were more successful at way-finding and, as with the engineering students, that females improved with training.

All these results underscore the importance of instruction promoting varied spatial abilities. Studies attempting to link spatial ability to activities requiring spatial reasoning such as sports and hobbies have primarily illustrated the complexity of the distinction between opportunity to learn and performance. Scrutiny of the US curriculum, however, reveals little opportunity to learn spatial reasoning. Until recently, elementary mathematics in the United States was almost synonymous with arithmetic. This is followed by a sudden jump to Euclidean geometry in high school and later (often in college) to the geometry concerned in calculus. Curriculum designers consciously chose to reduce or eliminate emphasis on spatial reasoning to be fair to all. This intriguing policy illustrates the role of culture in decision making. It appears that exactly the opposite decision would have been more sensible. By equalizing opportunity to learn, gender differences might have been reduced. Limited instruction and experience with spatial reasoning in the curriculum may exacerbate gender differences.

In summary, in the area of spatial performance, the definition of the construct 'spatial reasoning' has varied substantially. Research shows that the definition of the construct determines the existence and magnitude of gender differences in performance. Moreover, for tasks involving mental rotation where the largest gender differences are found, opportunity to learn as well as instructions to respond rapidly are both important factors in reducing or even eliminating gender differences in performance. Although the connection of the ability to rotate a whole object mentally with performance in mathematics, science, or engineering remains unclear, mental rotation continues to be an object of study.

Verbal Performance

Measures of ability, starting with the Binet measure of intelligence, have generally relied on verbal skills. The Binet test includes vocabulary items as well as comprehension and interpretation items. Tasks such as verbal analogies, opposites, and sentence completion and paragraph comprehension characterize a broad range of items and predominate in widely used tests, including the SAT verbal measure, the Graduate Record Examination verbal measure, and a multitude of achievement tests administered in schools.

Cultural expectations typically include the view that women excel in 'verbal ability', variously defined. This perception is consistent with established findings that on average girls learn to speak and to read earlier than boys. In addition, serious problems with reading are more common among boys than girls; however, individual reports of these difficulties may be affected by cultural expectations. Maccoby and Jacklin reported that differences in verbal ability were well established in their 1974 summary of gender differences. Subsequent meta-analyses conducted by Janet Hyde and Marcia Linn suggest that these differences are very small and vary by the selectivity of the sample as well as the discipline represented in the item.

Differential opportunities to develop specific forms of verbal ability influence performance on verbal measures. Vocabulary items that require the interpretation of word meanings without supplying a context often require students to rely far more on broad verbal experience and test-taking expertise than do more contextualized items such as sentence completion or verbal comprehension tasks. Furthermore, performance on verbal comprehension tasks that strongly rely on information from a specific discipline often display patterns of gender differences compatible with participation in courses from these fields. Because more males than females participate in engineering and physical science courses, passages with physics or engineering content tend to favor males. Similarly, because more females participate in humanities and social sciences, passages from these domains tend to favor females.

Another important consideration in analyzing gender differences in verbal ability concerns the selectivity of the sample. In general, differences in performance between males and females are relatively small when the general population is used but are amplified when selective populations are studied. A very large group of college-bound students take the SAT but this is sample is selective. An interesting pattern of results has emerged over the years in SAT performance. Early measures of verbal ability for this test tended to favor females. However, in the late 1980s, this measure started to favor males. This change in gender differences coincided with an increased reliance on science passages in the comprehension section. Hyde and Linn demonstrated that when the selective SAT sample was considered, gender differences favoring males were significant, but when a representative group of students were used, these differences disappeared. One might hypothesize that among the selective sample the advantage of specific knowledge about the topics might have a bigger effect on overall differences between males and females. This effect, while still operative, would be very small in a population of individuals when most lacked specific knowledge of the topics.

Recent research calls for viewing verbal ability, verbal performance, and verbal communication skill as nested in the discipline and context where it is used. Ability to write news articles, poetry, short stories, novels, technical manuals, scientific papers, or law briefs depends far more on opportunity to learn, diligence, and motivation than on gender. Viewed through this lens, gender differences play a minor role in verbal performance. By far, the larger differences occur in the age of first speech.

BIOLOGICAL BASIS FOR ABILITY

Belief in a biological basis for gender differences in intellectual performance has motivated numerous research programs. Many studies have examined differences in distributions of ability for males and females. The notion that males are more variable dates back at least to Charles Darwin, who made this generalization from observations and studies of physical characteristics. Scientists sometimes assumed that mental traits were inherited, and would therefore show distributions similar to those of physical traits. Edward Thorndike used this supposition to argue that because 'men differ in intelligence and energy by wider extremes than do women', women should be not educated for professions that required giftedness rather than average ability.

Leta Hollingworth noted that psychologists had variable meanings for 'variability', including: a wider range in distribution of a trait, or the same range with greater frequency in the extremes. Moreover, her large-scale study of neonate measurements did not support any interpretation of 'greater male variability'. Subsequent large-scale studies have reinforced Hollingworth's view that variability between the genders reflects cultural rather than biological factors, and the hypothesis of greater variability has lost currency as an explanation of why 'leadership in the world's affairs … will inevitably belong oftener to men'.

Researchers now use a variety of new, and rapidly advancing methods, for studying the brains of males and females. These include non-invasive measurement of neurotransmitter levels, synaptic strength, number of synapses, and receptor levels as well as postmortem studies comparing the size of various brain areas. Studies often reveal gender differences but no consistent pattern has emerged and new methods often yield findings that contradict prior work. Both small sample sizes and difficulty associating brain functioning with complex performance limit the generalizability of current research. Recent findings offer some insight into the diagnosis and treatment of problems related to brain functioning. For example, when reading performance is measured, men show a higher rate of dyslexia. However, Sally Shaywitz and her colleagues have found that the incidence of dyslexia in brain function appears to be equally distributed by gender. To explain this apparent contradiction, Shaywitz reports evidence that women and men process language in different parts of their brains when reading (men use the right inferior frontal gyrus, women use the left gyrus as well).

ACADEMIC ACHIEVEMENT

As the discussions of spatial and verbal abilities suggest, the distinction between educational attainments in mathematics, science, language and other domains, and spatial and verbal ability has blurred over the past decade. However, it is still a long way from the laboratory to the classroom. Experimental research suggests constraints and draws attention to certain capabilities but does not often yield direct instructional implications. Educational accomplishments depend on opportunity to learn; cultural expectations have a profound impact on who participates and who persists in a variety of fields. Initially, access to education varied by gender, and today gender patterns in enrollment and persistence in courses remain.

The most profound differences in gender, access, and participation are in mathematics and science, where men predominate. In the past 20 years, fields like computer science have gained prominence and importance both in the economy and in academia. After an initial surge, to approximately one-third of the student body, the proportion of women has declined, and today women receive about 12 percent of the computer science PhDs in the United States.

Participation in the natural sciences varies both by field and by nationality. For example, in the United States, equal numbers of women and men enroll in calculus courses and select mathematics as a college major. By graduate school, according to counts in 1999, women earned over one-third of the mathematics PhDs granted to US citizens by US universities. Since the United States educates a large percentage of graduate students from other countries, when all nationalities are considered only 28 percent of those receiving mathematics PhDs from US universities were women. In contrast to the situation for computer science, both the number and proportion of women receiving PhDs in mathematics has increased. Proportions differ considerably in other fields: women currently earn about 20 percent of PhDs in physics but 40 percent of those granted in biology.

Opportunity to learn effects vary by institution of higher education as well. For example, Marcia Linn and Cathy Kessel found that proportions of bachelor's degrees granted to women from the 'top ten' mathematics departments ranged from 9 to 47 percent. This variability is consistent with research done by Elaine Seymour and Nancy Hewitt. Their analysis of a national sample of 800 000 undergraduates showed that about two-thirds of students who enter college intending to study mathematics or statistics switch to another field. Moreover, 72 percent of the females and 60 percent of the males switched to other fields, but half of the males and two-thirds of the females switched to a field outside of mathematics, science, and technology.

Seymour and Hewitt's more in-depth study of 335 undergraduates in science and engineering at seven institutions suggests an explanation for why men and women switched fields – and moreover, that explanations differed for men and women. Although these undergraduates were considered 'highly qualified' (their SAT-M scores were at least 650), the primary concern of both switchers and persisters was poor teaching: courses were fast-paced, instructors were often unclear, unavailable, and uninspiring, even lecturing by reading from the textbook. Women more than men, however, were repelled by poor instruction. This study suggests that context of instruction rather than 'lack of ability' is a better explanation for students' lack of persistence in science and engineering fields. Possibly reflecting their status as nontraditional students, other motivations for participation – lack of acceptance by male peers, hence lack of opportunity to learn from peers outside of class – contributed to switching.

Other evidence suggests that women may be more often affected by poor instruction, but that improvements advantage all. For example, Marcia Linn and Michael Clancy demonstrated that new computer science courses tend to favor males, contrary to the usual experience of higher female grades. When courses are iteratively refined and improved, they tend to be equally successful in improving the learning and understanding of males and females. In general, when instructors use information from past courses to redesign their instruction to be more responsive to their students, individuals who have been at risk for failure, and groups that have traditionally performed less well, gain more than individuals who are traditional course participants. This finding resonates with the comments of switchers in the Seymour and Hewitt study and that students tend to persist more in nontraditional fields with effective instruction. The pattern of success and persistence also coincides with accounts of success of women who attended all-female colleges in the past.

CONCLUSION

The history of views of gender differences in intellectual ability has responded to changes in research methods. As methods for identifying and assessing gender differences have developed, new interpretations have emerged. Both more powerful tests and new methods for representing results, such as path analysis and meta-analysis, have influenced thinking. Analysis of aspects often categorized as 'noise' in early studies has strengthened the view that opportunity and inclination to learn play a powerful role in measured differences.

The nature of research findings depends on the methods used. The majority of psychological investigations of gender differences tests whether score differences for males and females significantly differ from chance. As the sample size increases and the power of the test increases, the likelihood of concluding that there are statistically significant differences also increases. Measures that tap aspects of broad constructs, such as verbal ability, spatial ability, and school achievement, often yield contradictory results. Synthesis techniques such as meta-analysis make the often unrealistic assumption that the studies replicate each other. In contrast, neurobiological studies with small numbers of women and men offer hints of results to come but are limited by sample size and by complex relationships between indicators and the potential constructs they measure.

In conclusion, the investigation of differences between males and females over the past decade reflects advances in methodological, conceptual, and cultural views of the field. Advances in the understanding of complex behavior have undermined the once prevalent deficit model and strengthened the belief in opportunity to learn. Changing cultural views of who can succeed in intellectual endeavors have enhanced opportunity and increased the choices available to all students. Most importantly, research on gender differences in intellectual attainment reinforces the importance of effective educational programs and establishes the value of research designed to enhance the learning of students with disparate prior experiences and opportunities.

Acknowledgments

This material is based upon research supported by the National Science Foundation under grants REC 98-73160 and REC 98-05420. Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors are grateful for input from members of the Web-based Integrated Science Environment project, as well as Susan Klein.

Further Reading

Caplan P, Crawford M, Hyde JS and Richardson JTE (eds) (1997) *Gender Differences in Human Cognition*. New York, NY: Oxford University Press.

Gould SJ (1981) *The Mismeasure of Man*. New York, NY: W. W. Norton.

Linn MC and Petersen A (1985) Emergence and characterization of sex differences in spatial ability: a meta-analysis. *Child Development* **56**: 1479–1498.

Linn MC and Kessel C (1996) Success in mathematics: increasing talent and gender diversity. In: Schoenfeld A, Dubinsky E and Kaput J (eds) *Research in Collegiate Mathematics Education II*, pp. 101–144. Providence, RI: American Mathematical Society.

Maccoby E and Jacklin C (1974) *The Psychology of Sex Differences*. Stanford, CA: Stanford University Press.

National Center for Education Statistics (2000) *Trends in Educational Equity of Girls and Women* (NCES 2000–030). Washington, DC: US Department of Education, Office of Educational Research and Improvement. Available at: http://nces.ed.gov/spider/webspider/2000030.shtml

Rossiter M (1982; 1995) *Women Scientists in America*. Baltimore, MD: Johns Hopkins University Press.

Science (1994) **263**: 1345–1532. [Special issue: Women in science: Comparisons across cultures.]

Seymour E and Hewitt N (1997) *Talking About Leaving: Why Undergraduates Leave the Sciences*. Boulder, CO: Westview Press.

Shaywitz S (1996) Dyslexia. *Scientific American*. Available at http://www.sciam.com/1196issue/1196shaywitz.html

Shields SA (1982) The variability hypothesis: the history of a biological model of sex differences in intelligence. *Signs* **7**(4): 769–797.

Steele C (1997) A threat in the air: how stereotypes shape intellectual identity and performance. *American Psychologist* **52**(6): 613–629.

Sternberg R (2000) The Holey Grail of General Intelligence. *Science* **289**: 399, 401.

Tavris C (1992) *Mismeasure of Woman*. New York, NY: Simon & Schuster.

Valian V (1998) *Why So Slow? The Advancement of Women*. Cambridge, MA: MIT Press.