College Board Report No. 95-2
ETS RR No. 95-14

Performance by Gender on an Unconventional Verbal Reasoning Task: Answering Reading Comprehension Questions Without the Passages

DONALD E. POWERS

College Entrance Examination Board, New York, 1995

Acknowledgments

The author extends his thanks to Patricia O'Reilly, who prepared the data; Ruth Yoder, who processed the text and tables; and Brent Bridgeman and Ann Gallagher, who provided helpful reviews of an earlier draft. Thanks are also extended to the Joint Staff Research and Development Committee of the College Board and Educational Testing Service for its support.

Donald E. Powers is a principal research scientist at ETS.

Researchers are encouraged to freely express their professional judgment. Therefore, points of view or opinions stated in College Board Reports do not necessarily represent official College Board position or policy.

Additional copies of this report may be obtained from College Board Publications, Box 886, New York, New York 10101-0886. The price is $15.

Printed in the United States of America.

Contents

Tables

Abstract

Data were reanalyzed from a previously reported study of the passage dependence of reading comprehension questions being developed for the revised SAT. The objective was to uncover any gender differences in approaches to and performance on a task requiring examinees to answer reading comprehension questions without reading the passages on which the questions were based. Verbally able males and females were compared with respect to their reported use of a variety of test-taking strategies involving reasoning, personal knowledge, and guessing. A few relatively small (and often inconsistent) differences were detected between male and female test takers. However, far more similarities than differences were noted with respect to both test performance and test-taking behavior. If these results are generalizable to more typical tests of verbal ability, they would seem to suggest that males and females employ quite similar approaches to standardized test taking.

Introduction

Gender differences in performance on measures of academic ability and achievement have generated much speculation and considerable research over the years, especially for tests on which the stakes are high. To account for the disparities, a variety of social, biological, educational, and psychological explanations have been offered (see Wilder & Powell, 1989, for example), and many specific factors have been posited as contributing to test score differentials. Among the possibilities that have been considered are genetic makeup (e.g., Benbow & Stanley, 1981), educational background/experience (e.g., Pallas & Alexander, 1983), personality traits and cognitive styles (e.g., Hassmen & Hunt, 1994), method of measurement (e.g., Bolger & Kellaghan, 1990), characteristics of test questions (e.g., Becker, 1990), and approaches to both problem solving and test taking (e.g., Ben-Shakhar & Sinai, 1991; Gallagher, 1992; Johnson, 1984). Test-taking approaches, for instance, may vary with respect to the propensity to guess in the face of uncertainty, the extent to which a slow and deliberate style is used, and the extent to which such traits as test anxiety and testwiseness influence performance.

Undoubtedly, more attention has been devoted to gender differences in mathematical ability/achievement than in other domains, such as verbal reasoning, where the differentials are much smaller and sometimes nonexistent. For instance, on the basis of a comprehensive synthesis of studies of performance on various kinds of verbal ability measures, Hyde and Linn (1988) were willing to assert that "there are no gender differences in verbal ability, at least at this time, in American culture, in the standard ways that verbal ability has been measured," and Feingold (1988) maintained, in both the text and title of his article, that "Cognitive gender differences are disappearing." Apparently, however, not everyone agrees. Halpern (1989), for instance, responded to Feingold's claim by noting that whether "we find gender differences depends on what, who, and when (in the life span) we test" (p. 1157).

Significance of Research on Gender Differences

Assuming that gender differences on measures of academic ability and achievement are indeed small, we might appropriately ask "Why study the possible causes of such differences at all?" The answer, according to Wilder and Powell (1989), is that "There are real, quantifiable educational and social consequences of test performance. Even small differences can add up to major effects in the aggregate. Slight shifts in the ratio of male to female superiority in a domain can alter the nature of the population that qualifies for special awards, scholarships, programs, and educational opportunities" (p. 31).

The relevance of this argument is apparent when examining trends in the SAT verbal scores of male and female test takers. The mean verbal score of females had been higher than that of males until the mid-1980s, when the average score of males first surpassed that of females (Burton, Lewis, & Robertson, 1988). Halpern (1989) noted the impact of this reversal—a drop in the number of female students receiving prestigious scholarships on the basis of high SAT scores. As Stanley, Benbow, Brody, Dauber, and Lupkowski (1992) have observed, even seemingly negligible mean differences can be associated with "sizable [differences in] odds" at the upper end of the score distribution (p. 47).

Previous Research and Further Hypotheses

Recently, Powers and Wilson (1993, in press) administered a highly unconventional verbal reasoning test—answering SAT reading comprehension questions without the availability of the passages on which the questions were based—to a sample of verbally able high school juniors. The study was designed to evaluate the reading comprehension questions being developed for the revised SAT, specifically, to evaluate the degree to which performance depends on having read the passages with which the questions are associated. A second objective was to determine the construct relevance of any strategies that students might use when they resort to answering questions without actually reading the passages. The study was partly in response to a recent revival of criticism that, because test takers can, to some degree, answer questions correctly without consulting the reading passages, the verbal portions of such tests as the SAT are not valid measures of reading comprehension (Katz, Lautenschlager, Blackburn, & Harris, 1990).

Although unique, the task had some similarities to the kind of ill-structured, unfamiliar reasoning problems that are encountered in real life (N. W. Burton, personal communication, January 4, 1993). The task also had another interesting feature. Withholding the reading passages was, in a sense, equivalent to stripping away all relevant context, i.e., creating an exercise that was virtually totally decontextualized (and perhaps as *un*authentic as we could possibly imagine). As Goldstein (1994) has pointed out, there is considerable evidence that the context in which a test question is embedded can appreciably alter its difficulty. Linn (1992) suggested that *differential* performance by males and females on various college admission tests was also related to the context in which reasoning was tested. Others have attempted to explain the relevance of context to gender differences by suggesting that females are more attuned to contextual clues than males are. For example, Goldberger (1992) asserted that "Far more than men, women and girls emphasize the importance of care, connection, and context as central to thought and problem solving. Men, more often than women, emphasize abstract principles and universal solutions in decision making" (p. 140).

A reasonable prediction stemming from this statement is that males should outperform females on a reading comprehension task for which the text has been removed. Based on a very small (N=2) informal survey, however, it was clear that alternative forecasts can be advanced with just as much conviction. For example, Graduate Record Examinations Executive Program Director Charlotte Kuh predicted that because of their superior inferencing skills, females would perform better than males. Because they read more literally than females, males would be at a greater disadvantage when not having the passages to reference (personal communication, October 18, 1994). Educational Testing Service (ETS) Research Scientist Ann Gallagher, on the other hand, speculated that males would perform better because of their greater confidence and their willingness to "wing it" in the face of uncertainty (personal communication, November 3, 1994). In any event, as Messick (1994) has noted, the usefulness of context depends on *which* features examinees respond to, and, in particular, the relevance in the present study of characteristics such as degree of concreteness or ambiguity to the construct being measured.

Objectives

The objective of this follow-up study was to compare males and females with respect to their test performances when the reading passages were not accessible, and also to determine any differential tendencies to use particular test-taking strategies or to consider particular test characteristics in this unusual test-like task.

Among the questions of interest were the following:

1. Are males and females equally likely to use a "process-of-elimination" strategy? Do they rule out the same number of alternatives per item?

2. Do males and females attend to the same features of answer choices? Do they use these features in the same way (i.e., to select versus to rule out choices)?

3. Do males and females differ in the kinds of guessing (random, informed, patterned, or intuitive) in which they engage?

4. Are males and females equally inclined to invoke personal knowledge or experience? Does this inclination differ according to the content of reading passages?

5. Do males and females resort to various reasoning strategies with the same degree of frequency?

These types of comparisons have the potential to shed additional light on how test-taking styles may contribute to gender differences in performance on traditional multiple-choice tests.

Method

Data collected in a previously reported study were reanalyzed. The specific procedures used in the study are described in some detail elsewhere (Powers & Wilson, 1993, in press). It will, however, be useful here to summarize some of the study's salient features.

Six reading passages and the questions associated with them were selected from among a larger pool that had been pretested for the revised SAT. The nature of these passages and the number of questions associated with each one were as follows:

1. A passage of approximately 900 words on language, in which the author, a Japanese American, recounts an experience he had just after the United States entered the Second World War (12 questions).
2. A 500-word passage adapted from an excerpt of a memoir written by Elizabeth Bishop about the poet Marianne Moore (6 questions).
3. An 800-word passage about Clarence Darrow and the Communist trial of 1920 (9 questions).
4. A 600-word passage that presents a theory about the nature of the object that exploded above Tunguska in 1908 (9 questions).
5. Two passages totalling about 800 words that present two views of the architectural design of cities. One discusses planned, medium-sized cities; the other offers a critique of modern cities (13 questions).
6. A 500-word passage excerpted from a book of literary criticism analyzing the work of Richard Wright (1908–1960) (5 questions).

Three different test forms, each consisting of questions concerning two different passages, were assembled and administered to volunteering students at eight secondary schools in seven states. Classroom teachers distributed the tests in a spiral fashion so that approximately one-third of all students (*N*=350) took each form. After attempting the task, examinees were asked to reveal how they had approached it—for example, the extent to which they employed different kinds of reasoning strategies, the degree to which they invoked personal knowledge or experience, the extent (and kind) of guessing they used, and the particular features of answer choices (e.g., specificity, length, abstractness, etc.) that were considered to either select or rule out options. Information on the extent to which examinees eliminated answer alternatives was also collected during the test.

The following comparisons are based on 335 study participants whose sex was either known from SAT files (*N*=268) or inferred from first name (*N*=67). The sex of 15 students could not be determined because no SAT records could be located and because first name (e.g., Robin, Lee, BJ) did not allow a trustworthy inference. These 15 subjects were therefore not included in the analyses.

TABLE 1

SAT Scores of Study Sample by Gender

	Male		*Female*		
	M	S.D.	*M*	S.D.	*d*
SAT-V	552	86	534	89	.21
Reading	55.0	8.8	53.4	9.4	.18
Vocabulary	54.1	9.1	52.7	9.2	.15
SAT-M	640	93	578	100	.63*
TSWE	51.7	5.8	52.3	7.2	-.08

Note: *N* = 122 males and 146 females for whom SAT records were available. *d* = effect size, i.e., difference between means divided by the pooled within-groups estimate of the standard deviation.
*p<.001

Results

Description of Sample

SAT records were available for 80 percent of the sample. Table 1 compares performances on the SAT and on the Test of Standard Written English (TSWE) for males and females in the study sample. Table 2

TABLE 2

Characteristics of Study Sample by Gender

	Male *(N = 122)*	*Female* *(N = 146)*
Honors English (%)	55	52
High School Rank (%)		
Top tenth	34	27
Second tenth	20	15
Second fifth	10	12
Third fifth	5	7
Fourth fifth	0	0
Fifth fifth	0	0
Unavailable	32	39
High School GPA (%)	27	28
A+	18	14
A	20	14
A–	11	18
B	24	26
C	<1	0
D, E, or F	0	0
Unavailable	27	28

shows comparisons with respect to high school achievement (grade-point average and rank in class) and course participation (honors English course). As the tables show, students in the sample were extremely capable, on average, in terms of each of these indicators. Except for the significantly better average performance of males on the SAT mathematical section (SAT-M), there were no large differences between males and females in the study on any of several indicators of academic ability.

Test Performance Without Reading Passages

As Table 3 shows, females performed significantly better (p<.05) than males on the passage about Richard Wright. This particular question set differed from the other five included in the study mainly in the extent to which personal knowledge was invoked when answering the questions, as is shown later. The effect size (of nearly .5) is considered "medium" according to one widely used standard (Cohen, 1977). There were no significant gender differences on any of the other five question sets, nor was there any tendency for males or females to perform better overall. One of our reviewers suggested that the failure to detect gender differences in performance on the task may have been due at least in part to the overall difficulty of the task, resulting in a relatively restricted range of performance by both males and females.

At the level of individual questions, significant (p<.05) between-group differences in question difficulty were detected for 4 of the 54 questions on all forms—about what would be expected by chance. All were in favor of males, and 3 were based on the Tunguska passage. The correlation of question difficulty for males with that for females (with respect to a simple rank ordering of difficulty) was .79, .74, and .93 for test forms A, B, and C, respectively.

TABLE 3

Test Performance (Without Passages) by Gender

Test Form/Passage		*Males*	*Females*	*d*
A Language	*M*	3.13	3.18	-.03
(N = 12 items)	*S.D.*	1.93	1.65	
A Marianne Moore	*M*	1.66	1.89	-.18
(N = 6 items)	*S.D.*	1.30	1.25	
B Clarence Darrow	*M*	3.37	3.54	-.10
(N = 9 items)	*S.D.*	1.78	1.67	
B Tunguska	*M*	3.58	3.17	.25
(N = 9 items)	*S.D.*	1.73	1.57	
C Architecture	*M*	3.96	3.52	.29
(N = 13 items)	*S.D.*	1.78	1.39	
C Richard Wright	*M*	2.63	3.19	-.48*
(N = 5 items)	*S.D.*	1.27	1.05	

Note: N = 53 males, 71 females for Form A; 52 males, 52 females for Form B; and 49 males, 58 females for Form C.
d = effect size, i.e., difference between means divided by the pooled within-groups estimate of the standard deviation.
*p<.05

Use of Various Test-Taking Strategies

Tables 4, 5, and 6 show the frequencies, by test form, with which male and female test takers used each of a variety of strategies related to reasoning (or inference), personal knowledge, and guessing. Overall, there were few statistically significant differences between the genders, and no consistent differences across test forms for any of the strategies. For test form C, females were less likely than males to report that, for a large proportion of questions, they had either "Chose[n] an answer because it seemed to be consistent with something stated in the other questions" or "Guessed a particular choice." Females were also more likely than males to state that they recognized the Marianne Moore passage or knew its source. Each of these differences can best be described as approaching "medium" in effect size.

Use of Features of Answer Choices

Table 7 shows that, with respect to the use of 26 possible answer-choice features to *select* answers, male and female test takers differed significantly on only two. They differed with regard to the use of three features for *eliminating* answer choices (Table 8). The only consistency was that males had a greater tendency, compared to females, to select choices they regarded as "outdated or old-fashioned," and to eliminate choices that were seen as "in tune with current thinking." Table 9, which displays the mean number of answer choices eliminated by males and females, does not suggest any differential tendency for males or females to rule out alternatives as implausible.

Consistency in Use of Strategies by Gender

It is evident from the various tables discussed previously that there were few detectable differences between male

TABLE 4

Use of Reasoning Strategies by Gender

Strategy	*Test Form*	*Males*	*Females*	*d*
Chose an answer because it seemed to be consistent with something stated in the other questions	A	2.75	2.68	.08
	B	2.79	2.79	.00
	C	2.67	2.24	.42*
Tried to determine the meaning of a word, or phrase, or the way in which it was used, from the other questions in the set	A	2.70	2.55	.14
	B	2.58	2.56	.02
	C	2.39	2.69	−.31
Ruled out an answer because it seemed to contradict something in the other questions	A	2.42	2.28	.12
	B	2.66	2.50	.14
	C	2.35	2.47	−.12
Assumed, guessed, or knew the answers to some questions and then, on the basis of these answers, reasoned what the answer to a later question would have to be (or what it could not be)	A	2.42	2.31	.11
	B	2.46	2.33	.12
	C	2.24	2.24	.00
Chose an answer because it resembled something in the question: I associated a word, phrase, or idea in the question with something in the answer I chose	A	2.10	2.03	.06
	B	2.18	2.32	−.13
	C	2.09	1.88	−.17
Tried to reconstruct the theme or main idea of the missing passage by reading all the questions and answers	A	62%	50%	.23
	B	57%	52%	.10
	C	64%	57%	.14

Note: Entries for the first five strategies are means on a 0 to 4 scale, with points indicating use of the strategy for few or no questions (0), about 25% of the questions (1), about half of the questions (2), about 75% of the questions (3), or all or nearly all questions (4). Entries for the sixth strategy are percentages of "yes" responses.
d = effect size, i.e., mean for males minus mean for females divided by the pooled within-groups standard deviation.
*p<.05

TABLE 5

Use of Personal Knowledge by Gender

Kind of Personal Knowledge Used	*Test Form/Passage*	*Males*	*Females*	*d*
Knowledge about the topic learned in or outside of school	A Language	1.53	1.52	.01
	A Marianne Moore	0.77	0.99	−.21
	B Clarence Darrow	1.60	1.42	.14
	B Tunguska	1.33	1.12	.16
	C Architecture	1.02	1.07	−.04
	C Richard Wright	1.84	2.12	−.20
Recognized the passage or knew its source	A Language	4%	4%	−.02
	A Marianne Moore	0%	8%	−.40*
	B Clarence Darrow	10%	4%	.23
	B Tunguska	16%	8%	.25
	C Architecture	2%	0%	.01
	C Richard Wright	22%	36%	−.30
Recognized the author and was familiar with his/her opinions	A Language	2%	1%	.04
	A Marianne Moore	4%	4%	−.03
	B Clarence Darrow	8%	2%	.28
	B Tunguska	4%	2%	.10
	C Architecture	2%	0%	.21
	C Richard Wright	46%	31%	.31

Note: Entries for "Knowledge about . . ." are means on a 0 to 4 scale, with points indicating use of the strategy for few or no questions (0), about 25% of the questions (1), about half of the questions (2), about 75% of the questions (3), or all or nearly all questions (4).
For "Recognized the passage . . ." and "Recognized the author . . . ," the entries are percentages of "yes" responses.
d = effect size, i.e., mean for males minus mean for females divided by the pooled within-groups standard deviation.
*p<.05

and female test takers with regard to the use of strategies to select answer choices. In fact, the degree of similarity in patterns of strategy use was far more apparent than were any dissimilarities. Table 10 reveals that the similarity between males and females with regard to the frequency of use of various strategies (correlations of the rank order of frequency of use) was generally very strong. For example (although not shown in Table 10), for test forms B and C, both males and females "Used vague hunches or intuition" more frequently than any other guessing strategy. "Guessed among two or more choices that couldn't be eliminated" was the second most frequent, "Guessed randomly among all choices" was the third most frequent, and "Guessed a particular choice" was the least frequent guessing strategy used, resulting in a perfect rank order correlation of 1.00. The greater inconsistency between males and females in the extent to which they used personal knowledge was due mainly to the fact that females more frequently recognized the Marianne Moore passage.

Discussion

A variety of test-taking behaviors was examined in order to determine any gender differences in approaches to and performance on an unusual verbal reasoning task. With regard to these behaviors, we found no consistent differences between males and females in eliminating answer choices because they were implausible. Nor were there any consistent differences between the genders in their reported use of numerous characteristics of response alternatives either to rule out choices or to select them. This finding is significant, because a "process-of-elimination" strategy is one that is often strongly advocated by commercial coaching schools as a way to "beat" standardized, multiple-choice tests.

TABLE 6

Use of Guessing Strategies by Gender

Strategy	*Test Form*	*Males*	*Females*	*d*
Guessed among two or more choices that couldn't be eliminated	A	2.23	2.38	-.14
	B	2.29	2.29	-.01
	C	2.41	2.62	-.20
Used vague hunches or intuition	A	2.55	2.38	.14
	B	2.08	2.13	-.05
	C	2.04	2.14	-.09
Guessed randomly among all choices	A	.73	1.07	-.35
	B	.67	.90	-.27
	C	.88	1.05	-.16
Guessed a particular choice (e.g., A or C)	A	.47	.41	.07
	B	.54	.39	.15
	C	.86	.41	.40*

Note: Entries are means on a 0 to 4 scale, with points indicating use of the strategy for few or no questions (0), about 25% of the questions (1), about half of the questions (2), about 75% of the questions (3), or all or nearly all questions (4).
d = effect size, i.e., mean for males minus mean for females divided by the pooled within-groups standard deviation.
*p<.05

TABLE 7

Frequency of Use of Various Answer-Choice Features to Select Answers by Gender

Feature	*Males* (N = *154)*	*Females* (N = *181)*
More carefully worded/qualified	76%	77%
More concrete	75	65
More definite/absolute	66	63
More specific	64	70
At the center of all the other choices, something in common with all choices	62	62
In tune with current thinking	60	66
More positive/less critical in tone/mood	58	62
More complex	55	52
Most obvious	48	45
More common/normal	48	45
Least ambiguous	44	39
More uncommon/unusual	40	36
Longer	39	32
More general	38	36
Simpler	37	41
Least obvious	34	45*
Less definite/more relative	32	33
Shorter	32	29
More abstract	30	31
Most ambiguous	29	29
More negative/more critical in tone/mood	29	28
Outdated, old-fashioned	27	15**
Too neutral	26	20
Too similar to other choices	25	20
Too extreme	18	18
Less qualified	14	13

*p<.05
**p<.01

(This strategy is also typically mentioned in test familiarization materials provided by test sponsors as an appropriate basis for making informed guesses.) That no notable differences were found is also meaningful in light of speculation (e.g., Linn, 1992) that males and females differ with respect to their willingness to take risks. (Insofar as eliminating the *right* choice is a distinct likelihood, this strategy could be a shaky one.)

Some researchers have suggested that, when forced to guess, females may favor certain kinds of options over others, e.g., familiar, nontechnical answer choices over unfamiliar, technical ones (Strang, 1977). We found little evidence, however, of any differences in preference according to gender, aside from one curious (and inexplicable) finding—that relative to females, males may be more inclined to select answer choices they consider to be "outdated or old-fashioned" and to eliminate choices they regard as "in tune with current thinking." If more than a chance finding, this result could be due to different beliefs about how to outwit the test maker. Before much credence is placed on this isolated outcome, however, follow-up is required.

The extent to which personal knowledge was invoked to answer questions seemed to be more a function of the content of the passages/questions than of gender. Although there was some hint that, with respect to this type of strategy, males and females may differ according to the content of the reading passages, there was little consistency in the findings. Nor did we detect any noteworthy differences between males' and females' use of several strategies that we regarded as illustrating reasoning.

To summarize, in a highly able sample of secondary school juniors we found remarkably few differences between males and females with respect to either performance on, or strategies used for, an unconventional reading task. The occasional differences that were noted were not consistent across the three test forms, suggesting that the few differences that were identified may have been more a function of the particular content of the questions (and the missing reading passages) than of the task itself. This inference is consistent with the fact that the passages (and questions) are based on a variety of subject matter that may have been differentially familiar or interesting to male and female test takers.

TABLE 8

Frequency of Use of Various Answer-Choice Features to Eliminate Choices by Gender

Feature	*Males* (N = *154)*	*Females* (N = *181)*
Less qualified	75%	69%
Too extreme	70	66
More abstract	68	56*
Too similar to other choices	64	66
Less definite/more relative	64	54
More general	62	59
Too neutral	60	60
More negative/more critical in tone/mood	56	56
Outdated, old-fashioned	55	55
Most ambiguous	53	45
Simpler	52	49
Least obvious	52	45
Most obvious	47	49
More uncommon/unusual	47	46
More complex	36	43
More common/normal	36	33
More specific	34	33
Least ambiguous	34	32
Shorter	33	29
More positive/less critical in tone/mood	31	22
At the center of all the other choices, something in common with all choices	30	18*
More definite/absolute	27	29
Longer	27	27
More concrete	22	22
In tune with current thinking	22	11**
More carefully worded, qualified	21	14

*p<.05
**p<.01

TABLE 9

Use of an Answer Elimination Strategy by Gender

	Mean Number of Answer Choices Eliminated		
Test Form/Passage	*Male*	*Female*	*d*
A Language	12.7	14.3	–.16
A Marianne Moore	5.0	5.7	–.12
B Clarence Darrow	10.4	13.5	–.38
B Tunguska	9.6	11.0	–.15
C Architecture	18.1	15.1	.27
C Richard Wright	7.2	6.0	.24

Note: d = effect size, i.e., mean for males minus mean for females divided by the pooled within-groups standard deviation.

Even more noteworthy than these infrequent differences, however, were the considerable similarities found between the genders in their approach to and performance on the exercise. We found little consistent evidence to suggest that performance on the task differs by gender. Apparent differences in the ways in which males and females confronted the assignment were detected only about as frequently as might be expected by chance alone. In fact, the frequencies with which males and females employed various strategies were extraordinarily similar. If any conclusion is to be drawn from these data then,' it is that males and females are far more alike than they are different with respect to both performance on and strategies used for the task studied here.

A possible limitation of the study, which may moderate the findings, is the low-stakes nature of the task that was presented. That is, test performance did not "count" in any meaningful way for the subjects participating in the study. Under operational conditions, in which test takers can be expected to be more motivated, more anxious, etc., the test-taking strategies studied here might be used differently by males and females. Still other strategies might also come into play. This speculation is perhaps moot, however, as this sort of task is unlikely to be encountered outside the context of a research study.

There were several other limitations as well. One reviewer was skeptical about the accuracy of examinee reports on strategy use, and would have preferred an alternative means of gathering these data. The rationale for our choice (and a discussion of the trade-offs involved) has been presented elsewhere (Powers and Wilson, in press). Despite these limitations, the study results are informative, we think, as further documentation of how males and females approach standardized, multiple-choice tests.

TABLE 10

Correlations Between Males and Females of Rank Ordering of Frequency of Use of Various Strategies

		Test Form		
Category of Strategy		*A*	*B*	C
Reasoning		.97	.90	.56
Guessing		.95	1.00	1.00
Personal knowledge:				
About the topic	.94[a]			
Recognized the passage or its source	.53[a]			
Recognized the author or his/her opinions	.85[a]			
Use of answer-choice features to:				
Select answers	.96[b]			
Eliminate choices	.97[b]			

[a] Computed over all six passages.
[b] Computed over all subjects, regardless of test form.

References

Becker, B. J. (1990). Item characteristics and gender differences on the SAT-M for mathematically able youths. *American Educational Research Journal, 27,* 65–87.

Benbow, C. P., & Stanley, J. C. (1981). Mathematical ability: Is sex a factor? *Science, 212,* 118–119.

Ben-Shakhar, G., & Sinai, Y. (1991). Gender differences in multiple-choice tests: The role of differential guessing tendencies. *Journal of Educational Measurement, 28,* 23–36.

Bolger, N., & Kellaghan, T. (1990). Method of measurement and gender differences in scholastic achievement. *Journal of Educational Measurement, 27*(2), 165–174.

Burton, N. W., Lewis, C., & Robertson, N. (1988). *Sex differences in SAT scores* (College Board Rep. No. 88-9; ETS Research Rep. No. 88-58). New York: College Entrance Examination Board.

Cohen, J. (1977). *Statistical power analyses for the behavioral sciences* (Rev. ed.). New York: Academic Press.

Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist, 43,* 95–103.

Gallagher, A. M. (1992). *Sex differences in problem-solving strategies used by high-scoring examinees on the SAT-M* (College Board Rep. No. 92-2; ETS Research Rep. No. 92-33). New York: College Entrance Examination Board.

Goldberger, N. (1992). Discussion. In Educational Testing Service (Ed.), *Sex equity in educational opportunity, achievement, and testing: Proceedings of the 1991 ETS Invitational Conference* (pp. 137–144). Princeton, NJ: Educational Testing Service.

Goldstein, H. (1994). Recontextualizing mental measurement. *Educational Measurement: Issues and Practice, 13*(1), 16–19, 43.

Halpern, D. F. (1989). Comment: The disappearance of cognitive gender differences: What you see depends on where you look. *American Psychologist, 44*(8), 1156–1158.

Hassmen, P., & Hunt, D. P. (1994). Human self-assessment in multiple-choice testing. *Journal of Educational Measurement, 31,* 149–160.

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104,* 53–69.

Johnson, E. S. (1984). Sex differences in problem solving. *Journal of Educational Psychology, 76*(6), 1359–1371.

Katz, S., Lautenschlager, G. J., Blackburn, A. B., & Harris, F. H. (1990). Answering reading comprehension items without passages on the SAT. *Psychological Science, 1,* 122–127.

Linn, M. C. (1992). Gender differences in educational achievement. In Educational Testing Service (Ed.), *Sex equity in educational opportunity, achievement, and testing: Proceedings of the 1991 ETS Invitational Conference* (pp. 11–59). Princeton, NJ: Educational Testing Service.

Messick, S. (1994). *Alternative modes of assessment, uniform standards of validity.* Paper presented at a conference on Evaluating Alternatives to Traditional Testing for Selection. Bowling Green State University (Oct. 25–26, 1994).

Pallas, A. M., & Alexander, K. L. (1983). Sex differences in quantitative SAT performance: New evidence on the differential coursework hypothesis. *American Educational Research Journal, 20*(2), 165–182.

Powers, D. E., & Wilson, S. T. (1993). *Passage dependence of the new SAT reading comprehension questions* (College Board Rep. No. 93-3; ETS Research Rep. No. 93-60). New York: College Entrance Examination Board.

Powers, D. E., & Wilson, S. T. (in press). Passage dependence of reading comprehension questions on the new SAT. *Journal of Educational Measurement.*

Stanley, J. C., Benbow, C. P., Brody, L. E., Dauber, S., & Lupkowski, A. (1992). Gender differences on eighty-six nationally standardized aptitude and achievement tests. In N. Colangelo, S. G. Assouline, & D. L. Ambroson (Eds.), *Talent development: Proceedings from the 1991 Henry B. and Jocelyn Wallace National Research Symposium on Talent Development* (pp. 42–65). Unionville, NY: Trillium.

Strang, H. R. (1977). The effects of technical and unfamiliar options upon guessing on multiple-choice test items. *Journal of Educational Measurement, 14,* 253–260.

Wilder, G. Z., & Powell, K. (1989). *Sex differences in test performance: A survey of the literature* (College Board Rep. No. 89-3; ETS Research Rep. No. 89-4). New York: College Entrance Examination Board.