DOCUMENT RESUME

ED 420 721 TM 028 443

AUTHOR	Wilson, Linda Dager; Zhang, Liru
TITLE	A Cognitive Analysis of Gender Differences on Constructed-Response and Multiple-Choice Assessments in Mathematics.
PUB DATE	1998-04-00
NOTE	24p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Diego, CA, April 13-17, 1998).
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Age Differences; *Constructed Response; Elementary School Students; Elementary Secondary Education; Mathematical Concepts; *Mathematics Tests; *Multiple Choice Tests; Problem Solving; Secondary School Students; *Sex Differences; State Programs; Tables (Data); Test Construction; *Test Items; Testing Programs
IDENTIFIERS	*Delaware; Iowa Tests of Basic Skills

ABSTRACT

This study is based on data from a state-wide assessment that included both multiple-choice and constructed-response items. The intent of the study was to see whether item types make a difference in gender results. The items on both tests were categorized according to whether they assessed procedural knowledge, concepts, problem solving, or mathematical communication. Data came from Delaware, which in 1995 administered two tests in mathematics to all public school students in grades 3, 5, 8, and 10. A total of 29,809 students were tested on the Delaware "Interim Assessment," a constructed-response test, and the Iowa Tests of Basic Skills (ITBS) (or the ITBS Tests of Achievement and Proficiency Survey Battery in grade 10). On the Interim Assessment in grade 3, the female mean raw score was higher, but at all other grades, males outperformed females. On the ITBS, males scored higher in grades 3 and 8, and there was no difference at grade 5 or 10, although the male mean score was higher at every grade. Differences generated by the broad and narrow interpretations of problem solving show some gender differences that might not have been apparent otherwise. Overall, males outperformed females on problem solving at grades 8 and 10. Results in some ways contradict the more hopeful conclusions of other studies that have shown the gender gap to be narrowing, though they do affirm some results of a study that showed males stronger in problem solving in the high school years. Results suggest that, while the gap is narrowing on traditional multiple choice tests, it is still present on more complex items that require students to construct their own responses and communicate their thinking. It is especially disturbing to see that the gap increases with grade level, which is in keeping with other studies showing females falling behind in adolescence. An appendix contains the item characterization protocol of the 1995 Delaware Mathematics Assessment. (Contains 6 tables and 16 references.) (SLD)

A Cognitive Analysis of Gender Differences on Constructed-Response and Multiple-Choice Assessments in Mathematics

Linda Dager Wilson
University of Delaware

Liru Zhang
Delaware Department of Education

Interest in gender differences in mathematics achievement remains high, after more than thirty years of research. There seems to be some evidence that the wide gender gap that favored males has been narrowing (see following discussion). At the same time, more varied forms of large-scale assessment, in conjunction with more sophisticated psychometrics, give us new evidence that gender differences in mathematics achievement may be more complex than the results from multiple-choice tests indicate. The introduction of different item types, such as student constructed-response items, makes it possible to ask new questions about gender differences, and to probe more deeply into differences that may or may not exist within different cognitive constructs in mathematics.

This study is based on data from a state-wide assessment that included both multiple-choice and constructed-response items. The results from this assessment, given in grades 3, 5, 8 and 10, gave us the opportunity to look at gender differences within two different item types and to compare the results. The intent of the study was see whether new item types make a difference in gender results, and also to use both item types to analyze gender differences on several constructs that were assessed with both types of items. We examined results from the multiple-choice test and also from the constructed-response test, then compared those results. We categorized the items on both tests according to whether they assessed procedural knowledge, concepts, problem solving, or mathematical communication, and analyzed the results on each of the two tests.

What do we know about gender-related differences on large-scale assessments in mathematics? Three different meta-analyses show the progression in results over approximately thirty years. Maccoby & Jacklin (1974), after analyzing results from studies in the 1960s through the early 1970s, announced that "boys excel in mathematics ability" (p. 352). Yet, twenty-five years later, Friedman (1989) found the average gender difference to be very small, and concluded that differences in performance were decreasing over the years. Hyde, Fennema & Lamon (1990) also concluded that gender differences are small. They did find that girls showed slight superiority in computation in elementary and middle school, but boys outperformed girls in problem solving during the high school years. These two meta-analyses, analyzing a total of 174 studies, would seem to show that whatever gender-related differences were apparent in 1974 had all but disappeared by 1990.

These findings have been verified by more recent studies. Fan et al (1997) found no gender difference in total group means on the data from the National Education Longitudinal Study of 1988. However, at the high end of the distribution, they found differences favoring males, especially between grades 8 and 12. Tate (1997), analyzing the results of a variety of studies, concluded that there were no significant gender differences on items measuring basic skills. The exception was on the trend data from the National Assessment of Educational Progress, on which 17-year-old males scored higher. When differences do exist, they seem to emerge in secondary school and favor males.

Though all of these studies seem to concur that females have "caught up with" males in mathematics achievement in elementary and middle grades, and that males are still outperforming females at higher grades, we need to take another look at those results. All of the studies mentioned above were based on data from some form of standardized achievement tests, and all the item types were multiple-choice in nature. Many large-scale assessments in mathematics have expanded the types of items they include. For example, the National Assessment of Educational Progress (1995) now includes in its design mathematics items that require students to construct their own responses, in both a short form, such as a number, drawing, or short explanation or a long form, such as a lengthy explanation. As Archbald & Newmann (1988) and others (e.g., Romberg et al, 1990) have argued, multiple-choice items are limited in the kinds of cognitive levels they can adequately assess. Other forms of student-constructed response items, such as open-ended or performance items, are better suited to the assessment of higher-order thinking, such as problem solving. With the advent of these alternative types of items on large-scale mathematics assessment, more research is needed into gender-related differences.

As mentioned above, the 1992 National Assessment of Educational Progress mathematics assessment (grades 4, 8 and 10) contained three types of items: multiple-choice, short constructed-response, and extended constructed-response. In an analysis of the results, Silver et al (1997) concluded that "performance differences between males and females are disappearing" (p. 57) and that there was "little or no difference between males and females on any item type at any grade level, except for a slight advantage for females on extended constructed-response tasks at grade 8" (p. 45). In overall performance, males performed slightly better at grades 4 and 12, and females at grade 8. Also, the percent of males and females classified at or above the Proficient achievement level was similar for all groups. Dossey et al (1993) also analyzed the results of the 1992 NAEP with respect to item types in mathematics. Looking at the percentages of those students nationally who scored at the satisfactory level or better (for the constructed-response items) and at the average percent correct (for multiple-choice items), they concluded that males performed slightly higher at grade 12 on multiple-choice items and also at grade 4 on short constructed-response items, while females performed slightly higher for grade 8 on the extended constructed-response items. All other results were not significantly different. While the differences were small, there did seem to be a difference in results from one item type to another.

A study of gender differences as it relates to item types in science assessment (Klein et al, 1997) compared results from performance assessments with traditional multiple-choice tests. Item type seemed to have little effect on gender differences in scores. They did find, however, that while girls had higher overall means on the performance measures, boys tended to score higher than girls on certain types of questions within a performance task. Specifically, girls tended to do better on questions that required making the correct interpretation of the observed results of the experiment, whereas boys did better on questions that involved making predictions.

Under the assumption that results from more than one item type can provide more robust evidence for the existence or non-existence of gender differences within different mathematical processes, this study combines results from both multiple-choice and constructed-response items. We chose to focus on the three mathematical processes, or ways of thinking, that are described in the NAEP framework: procedural skills, conceptual understanding, and problem solving. In addition, we looked at results for those items that required that students communicate their mathematical ideas.

Methods

This study analyzes data from the state of Delaware, which in 1995 administered two tests in mathematics to all public school students in grades 3, 5, 8, and 10. A total of 29,809 students were tested. The first test, termed the "Interim Assessment," was comprised entirely of student constructed-response items. Each grade level test had between 10 and 15 items, all situated within a single "real-world" context, and concentrating on one major mathematical domain (such as number or measurement). These tests were scored by trained scorers. Students in grades 3, 5, and 8 also took the Iowa Test of Basic Skills (ITBS) Survey Battery (Hoover et al, 1993), while students in grade 10 took the ITBS Tests of Achievement and Proficiency Survey Battery (Scannell et al, 1993). Both were norm-referenced, machine-scored, and composed entirely of multiple-choice items. These tests were used for Title I assessment and linking purposes only and were administered on a different day than the Interim Assessment. Scores for the Interim Assessment were reported on the individual, school, district, and state levels.

For this study each item from both sets of tests was categorized into one of three categories: Procedural, Conceptual, and Problem Solving. (See Appendix A for complete item categorization protocol.) In brief, procedural items were defined to be those that demand routine computation or the application of a routine procedure; they may have multiple steps.

Conceptual items were seen as those whose primary focus is on understanding a concept; they may require explaining a concept, or they could require the application of a concept in a limited way. Problem Solving items were those that demand that a student combine concepts or apply them in a new way to a novel situation; they require that students devise a plan or strategy and carry it out to reach a solution.

Each constructed-response item was categorized a second time, according to whether or not the item demands communication (through analysis of the multiple-choice items we determined that none assessed communication). In these items, the focus is on clarity, completeness, and mathematical accuracy of an explanation, argument, conjecture, etc; the scoring rubric takes into account the communication of mathematical ideas, whether it be through explaining an answer, showing work, drawing a graph, making a table, drawing a picture, writing an equation, or making an argument.

The intent was to identify categories that could be used across all four grade levels, and would be inclusive of all pertinent mathematical domains. Thus we chose the process, or mathematical thinking categories, rather than content categories, such as whole number operations, as such a scheme would not incorporate all grade levels.

Characterizing the items according to whether or not they assess communication was done for two reasons. First, constructed-response items are more likely to assess communication than more traditional types of items. Examining results from items that assess communication allows an opportunity to analyze whether there are gender differences for this critical aspect of mathematics (NCTM, 1989). In addition, language arts-related skills have often been seen as a particularly female strength, so items that assess these skills would be of particular interest in a study of gender differences.

As with all categorization schemes, it is not always possible to reach agreement on what each of these categories mean, nor to fit each item neatly into only one category (Silver, Kenney, & Salmon-Cox, 1992). The attempt was to identify the primary category that seemed to be elicited by each item. The items were categorized by three independent raters. On the second categorization, for those items that assess communication, there was unanimous agreement. However, when the raters categorized the items according to the other cognitive areas of procedures, concepts, and problem solving, it became clear that there were disagreements as to what constitutes problem solving. Two of the raters tended to take a narrow view of problem solving, reserving that category for those items that were deemed to be novel or non-routine in nature for that grade level. The third rater took a broader view of the kinds of situations that might be considered novel or non-routine. An example from a (hypothetical) third grade item illustrates this:

Joe has 82 red marbles and 69 blue marbles. About how many more red marbles does Joe have than blue marbles?

A. *10*

B. *20*

C. *25*

D. *30*

Under the narrow interpretation of problem solving, this item was seen as primarily procedural. The judgment was that most third graders would solve this problem by using a standard rounding procedure, either rounding 82 and 69 to 80 and 70, and then subtracting, or subtracting 69 from 82 and then rounding the result. Under a broader interpretation, the item was seen as requiring several steps and demanding that students make sense of a contextual situation.

Since there seemed to be no resolution to these different interpretations, we decided to analyze the results both ways. Thus, items classified under a narrower interpretation of problem solving were analyzed for gender differences, and then those same items were analyzed again under a broader classification. By including both results, we reasoned, we can learn more about any gender differences that occur, and what those might indicate for the assessment of problem solving.

Means and standard deviations were compared for male and female students by grade, test (or item type), and cognitive category of test items. A two-tailed t-test was used to determine the significant mean differences between gender groups on different item types (e.g., multiple-choice from the ITBS and constructed-response from the Interim Assessment) and within different cognitive categories in mathematics.

Results

The results are given here with reference to the research questions.

1. Are there overall gender differences on ITBS or the Interim Assessment at each grade level?

Insert Tables 1 and 2 about here

. On the Interim Assessment (all student constructed-response items) for grade 3, the female meanraw score was higher (by .09), but the difference was not significant. At all three other grade levels, males outperformed females. The meanraw score difference was .391 at grade 5, 1.277 at grade 8, and .931 at grade 10. Converted to scale scores, the greatest mean difference occurred at grade 8.

On the ITBS (all multiple-choice items), males scored higher in grades 3 and 8, and there was no difference at grade 5 or 10. At every grade, the male mean score was higher. The greatest difference (converted to scale scores) occurred at grade 8.

2. For each test, are there gender differences within different cognitive categories?

Insert Tables 3 and 4 about here

In general, there was little difference in results on the Interim Assessment from the two categorization schemes (the broad and narrow interpretations of problem solving). Results for both interpretations of problem solving were the same, with the exception of grade 8. Here the broad interpretation of problem solving did not have sufficient items in the procedural or conceptual categories to allow for reliable statistical analysis. Males did score significantly higher on the problem solving items at grade 8. The narrow interpretation, which did yield sufficient data points in all categories, showed that males were significantly higher on procedural, conceptual, and problem solving items at grade 8. Females scored significantly higher on procedural constructed-response items at grade 3, but then males scored significantly higher on all process categories in grade 8, and on problem solving items at grade 10.

Insert Tables 5 and 6 about here

On the ITBS at grade 3, when multiple-step, routine problems are considered problem solving (the broad interpretation), males outperformed females at the .01 level. When those items are categorized as either procedural or conceptual (under the narrow view), males are stronger in those categories (.05 level for procedural and .01 level for conceptual). Those types of items seem to be the ones making the difference between male and female results at grades 3. At grade 5, males performed better on the conceptual multiple-choice items, while females performed slightly better in problem solving, but only under a broad view. At grade 8, males scored higher on conceptual and problem solving items, as well as on procedural items when these included routine multiple-step problems. At grade 10, males scored higher on basic procedural items and slightly higher on non-routine problem solving items.

3. Are there gender differences on Interim Assessment items that require mathematical communication?

The male raw mean score was higher at every grade level on the communication items. The difference was not significant at grades 3 or 5, but at grades 8 and 10 the males outperformed the females at the .01 level (see Tables 3 or 4).

Discussion

The differences generated by the two interpretations of problem solving seem to point to some gender differences that might not have been apparent otherwise. Overall, males outperformed females on problem solving at grades 8 and 10. The results from the two different interpretations of problem solving would seem to indicate that, at those grade levels, males also scored higher on problems that might be considered more routine in nature, or on procedural items that are more complex and require multiple steps. At grade 5, however, females had a higher mean score on the Interim Assessment problem solving items (under either interpretation). Generally, the male dominance is most pronounced on nonroutine problem solving at grades 8 and 10.

There is an interesting phenomenon when the procedural items are analyzed across grade levels. At grade 3, females did better, both on the constructed-response items and on the multiple-choice items that assessed procedural knowledge. At grade 5 there were no differences. But by grade 8 males outperform on procedural items, both the constructed-response and the multiple-choice, and that difference continues at grade 10 for the multiple-choice items.

The conceptual items showed males stronger on those that were multiple-choice at grade 3 and 5, and males also higher on both kinds of items at grade 8. At grade 10 there were insufficient items on the Interim Assessment, and no significant difference on the multiple-choice conceptual items.

Conclusions

The results in some ways contradict the more hopeful conclusions of other studies that have shown the gender gap to be narrowing, though they do affirm some of the results of the Hyde et al study that showed males stronger in problem solving at the high school years. The results suggest that, while the gap may be narrowing on traditional multiple choice tests, it is still present on more complex items that require students to construct their own responses and to communicate their thinking. It is especially disturbing to see that the gap increases with grade level, which is in keeping with earlier studies showing females falling behind in adolescence.

References

Archbald, D. & Newmann, F. (1988). *Beyond standardized testing: Assessing authentic academic achievement in the secondary school.* Reston, VA: National Association of Secondary School Principals.

Dossey, J., Mullis, I., & Jones, C. (1993). *Can students do mathematical problem solving?* Washington, DC: National Center for Educational Statistics.

Fan X., Chen, M., & Matsumoto, A. (1997). Gender differences in mathematics achievement: findings from the National Education Longitudinal Study of 1988. *The Journal of Experimental Education, 65*(2), 229-242.

Friedman, L. (1989). Mathematics and the gender gap: A meta-analysis of recent studies on sex differences in mathematical tasks. *Review of Educational Research, 59*(2), 185-213.

Hoover, H.D., Hieronymus, A.N., Frisbie, D.A. & Dunbar, S.B. (1993). *Iowa tests of basic skills: survey battery (Form K, Level 9).* Iowa: Riverside.

Hoover, H.D., Hieronymus, A.N., Frisbie, D.A. & Dunbar, S.B. (1993). *Iowa tests of basic skills: survey battery (Form K, Level 11).* Iowa: Riverside.

Hoover, H.D., Hieronymus, A.N., Frisbie, D.A. & Dunbar, S.B. (1993). *Iowa tests of basic skills: survey battery (Form K, Level 14).* Iowa: Riverside.

Hyde, J., Fennema, E., & Lamon, S. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*(2), 139-155.

Klein, S., Jovanovic, J., Stecher, B., McCaffrey, D., Shavelson, R., Haertel, E., Solano-Flores, G., & Comfort, K. (1997). Gender and racial/ethnic differences on performance assessments in science. *Educational evaluation and policy analysis, 19*(2), 83-97.

Maccoby, E. & Jacklin, C. (1974). *The psychology of sex differences.* Stanford, CA: Stanford University Press.

National Assessment Governing Board (1995). *Mathematics framework for the 1996 National Assessment of Educational Progress.* Washington, DC: Author.

Romberg, T., Zarinnia, A., & Collis, K. (1990). A new world view of assessment in mathematics. In G. Kulm (Ed.), *Assessing Higher Order Thinking in Mathematics,* pp.21-38. Washington, DC: American Association for the Advancement of Science.

Scannell, D., Haugh, O., Loyd, B., & Risinger, C. (1993). *Tests of achievement and proficiency: Survey battery (Form K, Level 16).* Iowa: Riverside.

Silver, E., Kenney, P., & Salmon-Cox, L. (1992). The content and curricular validity of the 1990 NAEP mathematics items: A retrospective analysis. In National Academy of Education, *Assessing Student Achievement in the States: Background Studies,* pp, 157-218. Stanford, CA: The Academy.

Silver, E., Strutchens, M., & Zawojewski, J. (1997). NAEP findings regarding race/ethnicity and gender: Affective issues, mathematics performance, and instructional context. In P. Kenney & E.Silver (Eds.) *Results from the Sixth Mathematics Assessment of the National Assessment of Educational Progress*, pp. 33-60. Reston, VA: National Council of Teachers of Mathematics.

Tate, W. (1997). Race-ethnicity, SES, gender, and language proficiency trends in mathematics achievement: an update. *Journal for Research in Mathematics Education, 28*(6), 652-

679.

Table 1

Results from Interim Assessment (constructed-response items)

Grade	3	5	8	10
n	7906	7713	7884	6306
Number of items	14	11	15	12
Total points possible	28	30	29	27
Mean raw score (SD)	16.193 (6.235)	10.269 (6.674)	13.215 (7.577)	11.332 (6.704)
Male n	4035	3889	3974	3095
Mean raw score (SD)	16.148 (6.308)	10.463** (6.819)	13.848** (7.713)	11.806** (6.980)
Female n	3871	3824	3910	3211
Mean raw score (SD)	16.239 (6.158)	10.072 (6.518)	12.571 (7.382)	10.875 (6.395)

*p<.05

**p<.01

__ insufficient data

Table 2

Results from ITBS (Multiple-Choice Items)

Grade	3	5	8	10
n	7913	7734	7684	6204
Number of items	30	35	45	36
Total possible points	30	35	45	36
Mean raw score (SD)	19.38 (6.01)	20.90 (6.64)	21.95 (7.27)	17.16 (6.85)
Male n	4059	3945	3877	3075
Mean raw score (SD)	19.55** (6.17)	20.97 (6.79)	22.27** (7.52)	17.33* (7.29)
Female n	3854	3789	3807	3129
Mean raw score (SD)	19.21 (5.84)	20.83 (6.48)	21.62 (6.99)	16.99 (6.38)

*p<.05
**p<.01

Table 3
Constructed-Response Results for Cognitive Categories:
Broad Interpretation of Problem Solving

Grade	3		5		8		10	
	Mean (SD)	# items	Mean (SD)	# items	Mean (SD)	# items	Mean (SD)	# items
Total Procedural	5.255 (1.303)	3	5.268 (2.761)	4	__	1	__	0
Conceptual	8.288 (4.845)	9	__	1	__	1	__	0
Problem Solving	__	2	.437 (.590)	6	11.150 (7.208)	13	11.332 (6.704)	12
Communication	6.746 (4.005)	6	5.845 (4.319)	7	7.180 (4.784)	7	8.258 (5.570)	8
Male Procedural	5.216 (1.351)		5.289 (2.793)		__		__	
Conceptual	8.245 (4.874)		__		__		__	
Problem Solving	__		.405 (.572)		11.804** (7.312)		11.806** (6.98)	
Communication	6.694 (4.011)		5.927 (4.442)		7.513** (4.914)		8.669** (5.799)	
Female Procedural	5.296** (1.250)		5.248 (2.729)		__		__	
Conceptual	8.331 (4.815)		__		__		__	
Problem Solving	__		.470 (.606)		10.485 (7.039)		10.875 (6.395)	
Communication	6.799 (3.999)		5.763 (4.191)		6.841 (4.623)		7.861 (5.310)	

**p<.01

__ insufficient data

Table 4
Constructed-Response Results for Cognitive Categories:
Narrow Interpretation of Problem Solving

Grade[1]	8		10	
	Mean (SD)	# items	Mean (SD)	# items
Total Procedural	4.605 (2.769)	6	__	1
Conceptual	3.014 (2.038)	3	__	0
Problem Solving	5.595 (3.637)	6	10.878 (6.304)	11
Communication	7.180 (4.784)	7	8.258 (5.570)	8
Male Procedural	4.844** (2.783)		__	
Conceptual	3.133** (2.099)		__	
Problem Solving	5.871** (3.679)		11.346** (6.658)	
Communication	7.513** (4.914)		8.689** (5.799)	
Female Procedural	4.362 (2.734)		__	
Conceptual	2.893 (1.968)		__	
Problem Solving	5.315 (3.572)		10.427 (6.098)	
Communication	6.841 (4.623)		7.861 (5.310)	

[1] Grades 3 and 5 are identical to Table 6; **p<.01; __insufficient data

Table 5

Multiple-Choice Results for Cognitive Categories:
Broad Interpretation of Problem Solving

Grade	3		5		8		10	
	mean (SD)	# items	mean (SD)	# items	mean (SD)	# items	mean (SD)	# items
Total Procedural	4.12 (1.46)	6	14.52 (4.61)	23	3.28 (1.11)	5	4.99 (2.48)	12
Conceptual	7.89 (2.23)	11	3.44 (1.58)	6	8.15 (2.94)	16	4.13 (2.10)	9
Problem Solving	7.38 (3.20)	13	2.94 (1.38)	6	10.52 (4.25)	24	8.04 (3.24)	15
Male Procedural	4.10 (1.49)		14.55 (4.72)		3.27 (1.14)		5.09** (2.57)	
Conceptual	7.92 (2.29)		3.51** (1.59)		8.23* (3.07)		4.14 (2.19)	
Problem Solving	7.53** (3.24)		2.91 (1.39)		10.77** (4.32)		8.10 (3.44)	
Female Procedural	4.13 (1.42)		14.49 (4.49)		3.28 (1.09)		4.89 (2.37)	
Conceptual	7.86 (2.16)		3.36 (1.57)		8.07 (2.81)		4.11 (2.00)	
Problem Solving	7.22 (3.15)		2.98* (1.36		10.27 (4.17)		7.99 (3.03)	

*p<.05

**p<.01

Table 6
Multiple-Choice Results for Cognitive Categories:
Narrow Interpretation of Problem Solving

Grade	3		5		8		10	
	mean (SD)	# items	mean (SD)	# items	mean (SD)	# items	mean (SD)	# items
Total Procedural	12.08 (4.17)	19	4.65 (1.62)	7	15.25 (4.97)	29	10.14 (4.19)	21
Conceptual	5.52 (1.67)	8	7.80 (3.07)	14	3.58 (1.59)	7	3.81 (1.86)	8
Problem Solving	1.79 (.95)	3	8.46 (2.89)	14	2.66 (1.72)	8	3.40 (1.79)	7
Male Procedural	12.18* (4.26		4.65 (1.64)		15.43** (5.12)		10.26* (4.42)	
Conceptual	5.57** (1.70)		7.90** (3.13)		3.65** (1.65)		3.78 (1.96)	
Problem Solving	1.81 (.96)		8.42 (2.92)		2.75** (1.75)		3.45* (1.85)	
Female Procedural	11.98 (4.08)		4.65 (1.60)		15.06 (4.80)		10.02 (3.96)	
Conceptual	5.47 (1.62)		7.69 (3.01)		3.52 (1.53)		3.83 (1.76)	
Problem Solving	1.76 (.93)		8.49 (2.85)		2.57 (1.69)		3.35 (1.73)	

*p<.05
**p<.01

Appendix A

Item Characterization Protocol
1995 Delaware Mathematics Assessment

Mathematical Thinking and Processes

Each item will be categorized into one of three mutually exclusive categories: Procedural, Conceptual, and Problem Solving, using the following definitions.

PROCEDURAL
Demands routine computation or the application of a routine procedure; may have multiple steps. Students demonstrate knowledge when they select and apply appropriate procedures, or extend or modify procedures. Procedural knowledge includes the various numerical algorithms. It also encompasses the abilities to read and produce graphs and tables, execute geometric constructions, and perform noncomputational skills such as rounding and ordering. These latter activities can be differentiated from conceptual understanding by the task context or presumed student background--that is, an assumption that the student has the conceptual understanding of a representation and can apply it as a tool to create a product or to achieve a numerical result. In these settings, the assessment question is how well the student executed a procedure or how well the student selected the appropriate procedure to effect a given task.

CONCEPTUAL
The primary focus of the item is on understanding a concept; may require explaining a concept, or could require the application of the concept in a limited way. Students demonstrate conceptual understanding when they provide evidence that they can recognize, label, and generate examples and nonexamples of concepts; use and interrelate models, diagrams, manipulatives, and varied representations of concepts; identify and apply principles (i.e., valid statements generalizing relationships among concepts in conditional form); know and apply facts and definitions; compare, contrast, and integrate related concepts and principles to extend the nature of concepts and principles; recognize, interpret, and apply the signs, symbols, and terms used to represent concepts; or interpret the assumptions and relations involving concepts in mathematical settings. Conceptual understanding reflects a student's ability to reason in settings involving the careful application of concept definitions, relations, or representations of either. Such an ability is reflected by student performance that indicates the production of examples, common or unique representation, or communication indicating the ability to manipulate central ideas about the understanding of a concept in a variety of ways.

PROBLEM SOLVING
Item demands that student combines concepts or applies them in a new way to a novel situation. In problem solving, students are required to use their accumulated knowledge of mathematics in new situations. Problem solving requires students to recognize and formulate problems;

determine the sufficiency and consistency of data; use strategies, data, models, and relevant mathematics; generate, extend, and modify procedures; use reasoning (i.e., spatial, inductive, deductive, statistical, or proportional) in new settings; and judge the reasonableness and correctness of solutions. Problem solving situations require students to connect all of their mathematical knowledge of concepts, procedures, reasoning, and communication/representational skills in confronting new situations.

COMMUNICATION

Each item will be categorized a second time, according to whether or not the item demands communication, using the definition below:

COMMUNICATION
The item requires that students display their mathematical thinking, make a proof give an argument, write out the steps of a process, or make a graph, table, drawing, or construction to explain an answer. The focus is on clarity, completeness, and mathematical accuracy of an explanation, argument, conjecture, etc. Scoring rubric should take into account the communication of mathematical ideas, whether it be through explaining an answer, showing work, drawing a graph, making a table, drawing a picture, writing an equation, or making an argument.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM028443

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: A Cognitive Analysis of Gender Differences on Constructed-Response and Multiple-Choice Assessments in Mathematics

Author(s): Linda Dager Wilson and Liru Zhang

Corporate Source:	Publication Date:
Presented at AERA San Diego	4-16-98

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1	Level 2A	Level 2B
[X]	[ ]	[ ]
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, please

Signature: Linda Dager Wilson

Printed Name/Position/Title: LINDA DAGER WILSON / ASST. PROF.

Organization/Address: University of Delaware, Willard Hall, Newark DE 19716

Telephone: 302-831-1663 FAX:

E-Mail Address: ldwilson@udel.edu Date: 4/17/98

(over)

ERIC® Clearinghouse on Assessment and Evaluation

University of Maryland
1129 Shriver Laboratory
College Park, MD 20742-5701

Tel: (800) 464-3742
(301) 405-7449
FAX: (301) 405-8134
ericae@ericae.net
http://ericae.net

March 20, 1998

Dear AERA Presenter,

Congratulations on being a presenter at AERA[1]. The ERIC Clearinghouse on Assessment and Evaluation invites you to contribute to the ERIC database by providing us with a printed copy of your presentation.

Abstracts of papers accepted by ERIC appear in *Resources in Education (RIE)* and are announced to over 5,000 organizations. The inclusion of your work makes it readily available to other researchers, provides a permanent archive, and enhances the quality of *RIE*. Abstracts of your contribution will be accessible through the printed and electronic versions of *RIE*. The paper will be available through the microfiche collections that are housed at libraries around the world and through the ERIC Document Reproduction Service.

We are gathering all the papers from the AERA Conference. We will route your paper to the appropriate clearinghouse. You will be notified if your paper meets ERIC's criteria for inclusion in *RIE*: contribution to education, timeliness, relevance, methodology, effectiveness of presentation, and reproduction quality. You can track our processing of your paper at http://ericae.net.

Please sign the Reproduction Release Form on the back of this letter and include it with **two** copies of your paper. The Release Form gives ERIC permission to make and distribute copies of your paper. It does not preclude you from publishing your work. You can drop off the copies of your paper and Reproduction Release Form at the **ERIC booth (424)** or mail to our attention at the address below. Please feel free to copy the form for future or additional submissions.

Mail to: AERA 1998/ERIC Acquisitions
University of Maryland
1129 Shriver Laboratory
College Park, MD 20742

This year ERIC/AE is making a **Searchable Conference Program** available on the AERA web page (http://aera.net). Check it out!

Sincerely,

Lawrence M. Rudner, Ph.D.
Director, ERIC/AE

[1] If you are an AERA chair or discussant, please save this form for future use.

CUA
The Catholic University of America