DOCUMENT RESUME

ED 441 678 SE 063 564

AUTHOR	Calvert, Tasha; Engelhard, George, Jr.
TITLE	Gender Differences on Eighth Grade Mathematics Items: A Cross-Cultural Comparison of the United States and Spain.
PUB DATE	2000-04-28
NOTE	47p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Cross Cultural Studies; Foreign Countries; Grade 8; *Item Analysis; Junior High Schools; *Mathematics Achievement; Mathematics Education; Multiple Choice Tests; *Sex Differences
IDENTIFIERS	*Spain; *United States

ABSTRACT

Rather than describe mean gender differences in mathematics across different cultures, this study instead focused on an in-depth item analysis across two countries, the United States and Spain. The purpose of this study was to investigate gender differences on multiple-choice mathematics items across two countries by using the data gathered in the Third International Mathematics and Science Study (TIMSS). Although there were no mean gender differences on the total scores in the United States as there were in Spain, micro-level analysis of item characteristics must be considered when interpreting results. Even within categories, the direction of gender differences varied depending on other characteristics of the item. It was concluded that item difficulty was indeed related to gender differences in both countries. (Contains 50 references.) (ASK)

Running head: GENDER DIFFERENCES ON EIGHTH GRADE

Gender Differences on Eighth Grade Mathematics Items: A Cross-Cultural Comparison of the United States and Spain

Tasha Calvert
Emory University

George Engelhard, Jr.
Emory University

Paper presented at the annual meeting of the American Educational Research Association
in New Orleans, Louisiana

on April 28, 2000

Tasha Noel Calvert
Division of Educational Studies
Emory University
1784 North Decatur Rd., Suite 240
Atlanta, GA 30322

Phone: 404-728-4342
FAX: 404-727-2799
Email: tcalver@emory.edu

Gender Differences on Eighth Grade Mathematics Items: A Cross-Cultural Comparison of the United States and Spain

Mathematics has been referred to as a "critical filter" (Fennema, 1990, p. 2) that can inhibit individuals' occupational choices and later career advancement and change. Although some researchers (e.g., Noddings, 1998) criticize the structure of society that values mathematics more than other domains, individuals in technological societies must be able to understand and apply mathematical concepts. However, boys continue to outperform girls in mathematics achievement, particularly by the end of secondary school (Fierros, 1999; Frost, Hyde, & Fennema, 1994; Leder, 1992). This finding has recently been challenged by other researchers who have suggested that gender differences in the patterns of problem-solving strategies of students in early elementary school appear earlier than originally thought (Fennema, Carpenter, Jacobs, Franke, & Levi, 1998a; 1998b). Furthermore, it is unclear whether gender differences in mathematics have narrowed over time (Leder, 1992; Willingham & Cole, 1997). If equity is to be achieved, it is important to continue a line of inquiry into the nature of gender differences, particularly at the international level (Fierros, 1999).

The nature of the gender gap is also an increasingly international concern. More reliable samples of data, particularly large samples that are representative of students nationally, are needed to examine the patterns in gender differences across cultures (Willingham & Cole, 1997). Large data samples are well-suited for the study of gender differences because they are less subject to sampling variation and to other extraneous factors that might be found in smaller published reports. Although Benbow and Lubinski (1997) argued that mathematical talent appeared to have biological co-variates in gifted mathematics students, Stromquist (1989) argued that the assumption of innate differences between the sexes has led to attention being taken away from the study of environmental or cultural factors on achievement in school. In addition, Hyde (1997) concluded that Benbow and Lubinski's findings could not be generalized to the general population. Because women have achieved more in physics, an area that traditionally favors men in the United States, in other countries such as Belgium, Brazil, France, Hungary, and the Philippines (Dresselhaus, Franz, & Clark, 1994), Hyde noted that gender differences in mathematics are not due to biological factors. "Culture, not biology, shapes women's success in science" (Hyde, 1997, p. 287).

Because gender differences vary from country to country (Beller & Gafni, 1996; Hanna, 1988, 1994; Karunuñgan & Engelhard, 1999), some researchers (Hanna 1988; Leder, 1992) have argued that socio-cultural models, rather than biological factors, might explain the gender differences observed in eighth grade mathematics achievement. Baker and Jones (1993) showed that there was variation in the size and direction of gender differences in mathematics performance in the Second International Mathematics Study (SIMS). Furthermore, sociological factors such as variations in the gender stratification of educational and occupational opportunities in adulthood were related to the gender differences observed in mathematics performance. Because mathematics has elements of universality and truth, the subject has widely been regarded as "culture-free" (Bishop, 1988, p. 179). However, mathematics has increasingly been addressed as a socio-cultural, value-laden phenomenon. In general, researchers (e.g. Leder, 1990; Reyes & Stanic, 1988; Stromquist, 1989) proposed theoretical models that relate gender differences in mathematics achievement to the social and cultural environment rather than genetics or biology. According to Leder (1992), these models share a number of features in common:

> the emphasis on the social environment, the influence of other significant people in that environment, students' reactions to the cultural and more immediate context in which learning takes place, the cultural and personal values placed on that learning and the inclusion of learner-related affective, as well as cognitive, variables. (p. 609)

Cross-cultural data can assist researchers in understanding the differential influence of cultural variables on mathematics attainment and gender differences in that achievement.

However, many large-scale national and international studies on gender differences have been limited to reports of overall mean differences. In a meta-analysis Frost, Hyde, & Fennema (1994) found that girls scored higher than boys during elementary and middle school but by high school and college, gender differences favored boys. Women usually performed as well or better than men in areas such as computation, but only in the early years, while men performed better in content areas such as geometry and problems solving. However, the respective effect sizes of the mean gender differences were generally small.

Similarly, in a secondary analysis of the 1991 International Assessment of Educational Progress in Mathematics and Sciences (IAEP), the largest gender differences that favored boys occurred in the content areas of geometry and measurement for both 9-year-olds and 13-year-olds (Beller & Gafni, 1996). Furthermore, the differences tended to increase with age, but these differences varied according to the country. For example, there were no statistically significant gender difference in terms of effect sizes in Hungary, Scotland, and the United States for either the 9-year-old or the 13-year-old participants. However, statistically significant gender effects were found in Israel, Spain, Korea, and Ireland. In Ireland and Spain the effect size increased with age, whereas the effect size in Korea decreased with age. For example, in Spain the gender effect size in mathematics was 0.01 at age 9 but increased to 0.18 by age 13. Despite the effect size being statistically significant in favor of boys at age 13, an effect size of 0.20 is considered relatively small (Cohen, 1992; Willingham & Cole, 1997).

In the Third International Mathematics and Science Study (TIMSS), researchers (Beaton, et al., 1996; Fierros, 1999) found that few countries had statistically significant mean gender differences in the eighth grade. When gender differences were statistically significant, both content category with the exception of algebra (Beaton, et al., 1996) and higher cognitive demand (Fierros, 1999) tended to favor boys. The number of gender differences increased in the 12th grade mathematics literacy and advanced mathematics assessments (Fierros, 1999). In a majority of countries, boys outperformed girls at higher performing levels on both the Knowing and Procedures and the Reasoning and Problem Solving items as well as the multiple-choice and short answer items.

Item Characteristics

Because gender differences in mathematics achievement are complex (Tate, 1997), the nature of gender differences can be masked when comparing mean scores (Willingham & Cole, 1997). Consequently, the focus of research has turned to examining the characteristics of mathematics items rather than simply comparing mean scores (Engelhard, 1990; Garner & Engelhard, 1999). According to Hanna (1988, 1994), the Second International Mathematics Study (SIMS) data indicated that there were variations in gender differences at the item level among countries. When gender differences did exist, the algebra and computation subtests tended to favor girls, whereas the measurement and geometry subtests favored boys. Using Differential Item Functioning (DIF) analysis on items from SIMS, Engelhard (1990) found that gender differences tended to be more favorable toward boys as the complexity of the mathematics items increased and as the content changed from arithmetic through algebra to geometry.

Differential bundle functioning (DBF), a collection of DIF items with a common characteristic such as item content or cognitive complexity, is another method of detecting gender bias by producing a "bundle" of items that are differentially easier for one matched group of test takers (Ryan & Fan, 1996). Using DBF in a secondary analysis of SIMS, Ryan and Fan like Engelhard (1990) also found that algebra, arithmetic, and computation item sets were differentially easier for eighth grade girls in the United States and geometry and applied items were easier for boys in the United States sample. In addition, Ryan and Fan suggested that other areas such as ratio, proportion, and percent should be examined in order to detect relationships between gender differences and subcontent domains. Although Garner and Engelhard (1999) found a similar relationship between content and gender on the 1994

Georgia High School Graduation Test (GHSGT), boys had an unexpected advantage on number and computation items in high school. Garner and Engelhard suggested that the male advantage on number and computation problems might be related to the inclusion of word problems and the use of calculators. In addition to finding gender differences in content areas, the researchers found that the type of item mattered. Multiple-choice items favored men while constructed response items favored women. Some research (e.g., Lane, Wang, & Magone, 1996) found that women performed better on constructed response items perhaps because their responses were more complete. However, the explanations for these findings are still inconsistent and emerging.

The difficulty of items may explain the inconsistencies in research findings on different content areas like computation (Bielinski and Davison, 1998). Although there were no gender differences on mean scores, Bielinski and Davison found that boys performed better on the mathematics subtests that included application problems involving ratios, proportions, and percents and estimation problems in real-life contexts. Girls, on the other hand, performed better on the subtest that required students to read, use, and interpret graphs. However, Bielinski and Davison suggested that these differences were not due to the content of the items, but rather, the difficulty of the items. In other words, boys performed better on more difficult items, whereas girls scored higher on easier items such as those found in the data interpretation subtest. Bielinski and Davison proposed that there is a shift in mathematics ability for boys and girls as mathematics items become more difficult. The gender-by-item difficulty interaction as well as the differences found in male-female variances (Feingold, 1992) may be the result of this hypothesized shift in ability (Bielinski & Davison, 1998). Bielinski and Davison, however, did not examine the complex relationships between difficulty and other item characteristics such as content or cognitive complexity. Kupermintz and Snow (1997) attempted to describe the relationship between cognitive complexity and difficulty by suggesting that differences between levels of cognitive performance, such as the difference between mathematical knowledge and reasoning, are not simply distinctions of difficulty. The distinction between knowledge and reasoning is instead a “qualitative, psychological distinction between kinds of cognitive functions” (p. 143).

Spain

To extend previous research about gender differences in mathematics, cross-cultural data are needed to explore the pattern of relationships between gender differences and item characteristics. Spain was selected for investigation for several reasons. First, Spain’s geographical position and history have made the country a major crossroad at which many cultures have met (Gil, 1994). Its diverse traditions and languages--Catalonian, Galician, Valencian, Basque, and Castilian--have resulted from the mixture and coexistence of several different cultures. Second, the Spanish educational system has undergone significant changes over the last three decades (Alberdi & Alberdi, 1991). Until 1970 co-education had been declared illegal, and compulsory education was limited to primary education until the introduction of the 1970 General Education Act (LGE), which was in force until 1990. Once the objective of providing at least eight years of schooling had been achieved (Gil, 1994), the General Arrangement of Education System Act of October 1990 (LOGSE) was introduced to ensure higher quality teaching levels). The reforms in 1990 further raised the compulsory school age from 14 to 16 years.

In addition to increasing the level of student education, Spain has been changing its nationally centralized educational system to one that is regionally centralized with high responsibility at the school level (Beaton, et al., 1996). The central administration in Spain continues to be responsible for basic legislation, the regulation of certificates and degrees, the organization of the school system's levels, the subject matter, the requirements for passing from one grade to another, and general planning (Barrio, 1999). The rest of the responsibilities have been transferred to some of the "Autonomous Communities," which include Andalucia, the Canary Islands, Catalunya, Galicia, Basque Country, Navarra, and Valencia. These communities administer the educational system while the other communities continue to be managed by the Ministry of Education and Science. Few responsibilities, which include the maintenance of preschools and elementary schools and the additional pedagogic services, are delegated to the municipal governments.

Moreover, research on girls and educational equality in Spain has received scant attention (Alberdi & Alberdi, 1991). Like the United States, girls perform better than boys throughout primary school in Spain. For example, in eighth grade girls attained higher grades in arithmetic, reading, spelling, and comprehension. However, boys performed better in aptitude tests, except in abstract reasoning, a cognitive demand that typically favors boys in the United States. In TIMSS (Beaton, et al., 1996) as well as in the IAEP (Beller & Gafni, 1996), boys had higher mean mathematics achievement than girls in eighth grade and tended to perform better in measurement, whereas there were no statistically significant mean differences between boys and girls in the United States. In a national survey conducted in Spain, other researchers (Instituto Nacional de Calidad y Evaluacion, 1997) also reported small gender differences in mathematics between 14-year-old boys and girls. This difference in favor of boys appeared to increase as students moved through secondary school.

In addition, Spain was selected in this analysis because Spanish students performed lower than students in the United States in overall mathematics achievement in TIMSS (U.S. Department of Education, 1996). Overall, the mathematics assessment was more difficult for Spanish students. Consequently, exploring patterns in gender differences would be of interest because gender differences may be related to the difficulty of the mathematics item (Bielinski & Davison, 1998).

Purpose of the Study

Rather than describe mean gender differences in mathematics across different cultures, this study instead focused on an in-depth item analysis across two countries. Few researchers (Engelhard, 1990; Hanna, 1988) have conducted studies that examine gender differences in mathematics at the item level in different cultures. Because Bielinski and Davison (1998) suggested that task difficulty moderated gender differences, the interaction between item difficulty and gender differences within item characteristics was also be investigated.

The purpose of the present study was to investigate gender differences on multiple-choice mathematics items across two countries: United States and Spain. A secondary analysis of the data in the Third International Mathematics and Science Study (TIMSS) was used to address the following research questions:

1. Is there a relationship between gender differences and item difficulty?
2. Is there a relationship between gender differences and mathematics content after controlling for item difficulty?
3. After controlling for both item difficulty and content, is there a relationship between gender differences and cognitive demand?
4. Does the type of item difficulty index and estimate of gender difference affect the relationship between gender differences and item characteristics (difficulty, content, and cognitive demand)?
5. Do these relationships between item characteristics and gender differences in questions 1, 2, 3, and 4 replicate across cultures?

After exploring the relationship between item difficulty and gender differences, item difficulty was controlled before the relationship between content and gender differences was investigated because item difficulty might moderate the gender differences observed in mathematics (Bielinski and Davison, 1998). The relationship between gender differences and cognitive demand was explored after controlling for both item difficulty and content because the cognitive demand of the item was intended to be an indicator of the expected behavior within a content area (Robitaille, et al., 1993). Furthermore, the relationships observed between item characteristics and gender differences were not affected by whether difficulty and cognitive demand were controlled before exploring the relationship between gender differences and content or whether difficulty and content were controlled before examining the relationship between gender differences and cognitive demand.

Gender differences in each research question were operationally defined by both an Impact Index, which does not control for student achievement, and a Differential Item Functioning (DIF) Index, which controls for students' achievement in mathematics within each country. To examine relationships between gender differences and item characteristics, two types of difficulty indices were used and analyzed separately. Item difficulty was defined by both the TIMSS international difficulty index estimated from item response theory scaling (IRT) and a computed proportion-correct index based on the difficulty of each item calculated separately for both Spain and the United States. This study of the relationship between gender differences and item characteristics at the micro-level was intended to be descriptive in nature.

Method

Participants

Participants included 7,087 eighth grade students from the United States (3,561 girls and 3,526 boys) and 3,855 students from Spain (2,007 girls and 1,848 boys) who participated in TIMSS (Martin & Kelly, 1997). Population 2 within each country was defined as the two adjacent grades, which corresponded to seventh and eighth grade classrooms in the United States and Spain, containing the most 13-year old students (Martin & Kelly, 1996). For this study only the upper grade level of Population 2, eighth grade in both countries, was studied.

The TIMSS sample design was a two-stage cluster sample, with schools as the first stage of selection and classrooms within these schools as the second stage of sample selection (Foy, Rust, & Schleicher, 1996; Gonzalez & Smith, 1997). Because certain populations (e.g., African American and Hispanic students in the United States) were oversampled, scores were weighted. The probability of an individual student being selected was calculated by multiplying three selection probabilities--school, classroom, and student--and their respective adjustment factors (Gonzalez & Smith, 1997). Inverting the probability provided the sampling weight for each student. Sampling weights are necessary so that different subgroups of a population are proportionally represented when techniques other than simple random sampling are used (Foy, 1997). Three types of sample weights (total student weight, house weight, and senate weight) that have different properties but yield similar results were employed in TIMSS

(Gonzalez & Smith, 1997). For example, the sum of the total weights within a sample provided an estimate of the size of that population. In this case, each country would contribute proportionally to its population size so that analyses would be affected by the size of the particular population. On the other hand, the sum of the senate weight, proportional to the total weight, in each country would add to 1,000. In this instance, the contribution of each country is the same when researchers require international estimates. Although three sampling weights were provided in the TIMSS database, we used the house weight, which was designed to preserve the actual sample size of each population tested when performing significance tests.

United States sampling. In the United States researchers followed the international specifications with a few differences. First, an additional sampling stage preceded the school sampling stage. Primary sampling units (PSUs), defined as metropolitan statistical areas, single counties, or groups of counties, were sampled during this first stage (Gonzalez & Smith, 1997). In TIMSS there was a total of 1,027 PSUs on the sampling frame covering the 50 states. Eleven of the PSUs were taken as certainty selections because they represented the 11 largest metropolitan areas while 48 noncertainty PSUs, their probability of being selected would be proportionate to the 1990 population, were drawn from the remaining 1,016 PSUs. For the 11 certainty PSUs, the school sample was the first stage of selection. In the 48 sampled noncertainty PSUs, the measures of size of the school were proportional to the target grade size in the school divided by the PSU probability of selection. Furthermore, in both certainty and noncertainty PSUs, schools with high percentage of blacks and Hispanics (greater than 15 percent) were oversampled by a factor of two to allow for more detailed data analysis of patterns among minority groups in the United States. In addition, one lower-grade classroom and two upper-grade classrooms were sampled in each school (Martin & Kelly, 1997).

Spain sampling. In Spain explicit stratification by eight regions, two types of schools (public and private), and three levels of school size were created for a total of 43 strata (Martin & Kelly, 1997). However, because 15 of these strata were small, proportional allocation of the 150 schools was limited to the remaining 28 explicit strata. Other schools where the language of instruction was Euskera and very small schools were also excluded.

Instruments

All mathematics test items were grouped into 23 mutually exclusive item clusters; in other words, each item appeared in only one of the 23 clusters. Although multiple-choice, short answer, and extended response items were included in TIMSS (Adams & Gonzalez, 1996), only the 124 multiple-choice items were analyzed in the present study. The TIMSS items were first prepared in English and later translated into other languages (Maxwell, 1996). In Spain the TIMSS test was in four languages: Catellano, Catalan, Gallego, and Valenciano (Gonzalez & Smith, 1997).

Because the total testing time in Population 2 was not to exceed 90 minutes, students could not take all mathematics items (Adams & Gonzalez, 1996). Although there was a small subset of items common to all test booklets, students were given different booklets that were approximately parallel in content and difficulty. The design of the TIMSS test was based on a mutually exclusive cluster of items and then assigning these clusters to eight test booklets in a systematic fashion. An item cluster was defined as a small group of items that were collected together and treated as a block for the purposes of the test design. The number of items within each cluster varied according to the type of cluster and item (multiple choice, short answer, and extended response) administered. These clusters allowed for items to be rotated within test booklets. Of the 23 item clusters in mathematics, one cluster appeared in all booklets, some in four, some in three, some in two, and some in only one booklet. Each test booklet for Population 2 was comprised of up to seven item clusters of both science and mathematics items and was divided into two parts administered in two consecutive testing sessions.

Variables

Difficulty. This statistic, which reflected the difficulty level estimated from item response theory scaling (IRT), was developed from the performance of students in both grades in all countries (TIMSS, 1996). The higher the international difficulty index, the more difficult the item. The international difficulty of the multiple-choice items ranged from 326 to 693. A new difficulty index, based on the proportion correct for each item within each country, was computed separately for Spain and the United States. The international difficulty index was correlated to a computed proportion-correct scale--a conventional p-value--for each country. Pearson correlations between the international difficulty index and the conventional item difficulty indices for both the United States and Spain were $r(122) = -.91$, $p < .0001$ and $r(122) = -.88$, $p < .0001$, respectively. The associations were negative because smaller p-values corresponded to more difficult items. In other words, as the international index increased, the computer proportion-correct scale for each country decreased. The correlation between the conventional p-values for each country was $r(122) = .83$, $p < .0001$.

Content. The content categories referred to the subject matter content of the mathematics items (Robitaille et al., 1993). Although TIMSS was designed to permit a detailed analysis of student performance in many content categories, many of the detailed categories had to be collapsed into a few reporting categories (Garden & Orpwood, 1996). Multiple-choice items reported in the Population 2 tests covered six different content areas in mathematics that included: (a) Fractions and Number Sense (N = 41); (b) Geometry (N = 22); (c) Algebra (N = 22); (d) Data Representation, Analysis, and Probability (N = 18); (e) Measurement (N = 13); and (f) Proportionality (N = 8).

Cognitive demand. Items were originally classified according to their performance expectations as follows: (a) Knowing; (b) Routine Procedures; (c) Complex Procedures; (d) Solving Problems; (e) Justifying and Proving; and (f) Communicating (Fierros, 1999). The performance expectations were a reconceptualization of the cognitive-behavior dimension that had been used in earlier large-scale studies (Robitaille, et al., 1993). The purpose of the performance expectations was to describe in a non-hierarchical way the kind of performance that students would be expected to demonstrate within a content area. For this study the performance expectations (knowing, routine procedures, and complex procedures and reasoning, problem solving, and communicating) for each item were collapsed into two cognitive demands (knowing/procedures and reasoning/problem solving) as done in previous large-scale assessments (e.g., Fierros, 1999; Kuppermintz & Snow, 1997). Knowledge and reasoning have been identified as two meaningful dimensions for investigating mathematics achievement in large-scale assessments (Kuppermintz & Snow, 1997). In the present study, knowing and procedure items were reclassified as the knowing/procedures cognitive demand (N = 89) while reasoning, problem solving, and communicating items were reclassified as the reasoning/problem solving cognitive demand (N = 35).

Procedures

Reponses to each item were weighted to ensure that the results represented the student populations in the United States and Spain (Gonzalez & Smith, 1997). A SAS macro was used to score the multiple choice items from the TIMSS database as either correct or incorrect by gender. In this secondary analysis of data from TIMSS, the item was used as the unit of analysis for detecting gender differences in mathematics. Researchers (Engelhard, 1990; Holland & Thayer, 1988) have recommended that the Mantel-Haenszel (MH) Procedure be used to examine differential item functioning between selected groups such as gender. The values obtained from the MH Procedure generally range from –2.6 to 2.6. The scales were set up to indicate that girls were more likely to succeed on items that were positive while boys were more likely to succeed on items that were negative. Two estimates of gender differences on the multiple-choice items were obtained: an Impact Index and a Differential Item Functioning (DIF) Index. Gender difference estimates at the item level were calculated without controlling for the students' overall level of mathematics achievement in the Impact Index. The DIF Index, on the other hand, provided a parametric estimate of gender differences after controlling for the students' overall level of achievement. Because students did not complete every item on the mathematics section of the TIMSS assessment, student achievement was calculated using the first plausible value of the students' overall mathematics score. Although a plausible value should not be considered an individual test score (Gonzalez & Smith, 1997), this statistic provides the only estimate of a student's overall achievement in mathematics given that students received a selected set of the test items. To calculate the DIF Index, 10 score groups on the basis of the students' overall mathematics scores were created to control for achievement (Appendix). Although the number of students in each score group was not evenly distributed, particularly the number of students in the upper and lower extremes, collapsing the score groups did not change the results of the DIF estimates for either country. Both the Impact Index and the DIF Index were calculated separately for each country.

After values for gender differences were calculated using the MH Procedure, separate ANOVAs were used to examine the relationship between item characteristics (item difficulty, content, and cognitive demand) and the Impact and DIF Indices within each country. To study the effect of the type of difficulty index on the relationship between gender differences and item characteristics, two types of difficulty indices were used in the analysis: the international difficulty index and the computed proportion-correct scale for each country. Descriptive statistics were also calculated for both the Impact and DIF Indices for the United States and Spain to address whether mean gender differences for each content and cognitive category were significantly different from 0 after controlling for item difficulty.

Results

Impact Index

United States. The summary for the United States ANOVA, based on the Impact Index, is presented in Table 1. International item difficulty had a statistically significant effect on gender differences. After controlling for the international item difficulty, content category also had a statistically significant effect. However, there was no statistically significant relationship between gender differences and cognitive demand after controlling for the international item difficulty and content category. The interaction between difficulty and content category was statistically significant. This interaction was related to the size and the direction of the Pearson correlation between item difficulty and the Impact Index in each content category (Table 2). The correlation between item difficulty and gender differences within the two content areas of fractions and number sense and data representation, analysis, and probability were statistically significant, $r(39) = -.53$, $p < .001$ and $r(16) = -.52$, $p < .05$, respectively. In other words, a negative correlation indicated that more difficult items favored boys. Although the correlations between item difficulty and gender differences within the other content categories were not statistically significant, the direction of the correlation between item difficulty and gender differences within proportionality changed, $r(6) = .43$, n.s. The interactions between content and cognitive demand and between difficulty, content, and cognitive demand were not statistically significant.

In Tables 3 and 4, mean gender differences based on the Impact Index after controlling for item difficulty in each content category and cognitive demand are presented. The scales were defined so that boys were more likely to succeed on items with a negative value, whereas girls were more likely to succeed on items with a positive value. The mean gender differences on the Impact Index were significantly different from 0 for the measurement items with boys having the advantage in the United States (Table 3). Furthermore, the contrast between measurement and all other content categories with the exception of proportionality was statistically significant. Similarly, boys had an advantage on reasoning/problem solving items (Table 4). The difference between knowing/procedures items with the reasoning/problem solving items was also statistically significant. Within algebra, girls had an advantage on knowing/procedures items, ($M = .28$, $SE = .13$, $p < .05$), but in measurement, boys outperformed girls within both knowing/procedures and reasoning/problem solving ($M = -.46$, $SE = .16$, $p < .01$; $M = -.82$, $SE = .37$, $p < .05$, respectively).

The United States difficulty index had a greater effect on gender differences than did the international difficulty index (Table 1). Although content category did not have a statistically significant relationship to the Impact Index after controlling for the country specific difficulty index, the effect approached significance, $F(1,100) = 2.09$, $p = .0727$. Similarly, the interaction between content and the United States item difficulty approached statistical significance, $F(5,100) = 2.10$, $p = .0715$. Although the directions of the correlations between gender difference and United States item difficulty remained the same, more difficult items favored boys in algebra, rather than data representation, analysis, and probability (Table 2). United States item difficulty continued to be related to Impact within fractions and number sense with more difficult items favoring boys.

Descriptive statistics for mean gender differences, based on the Impact Index in each content category and cognitive demand, are presented after controlling for the United States item difficulty in Tables 3 and 4. When controlling for the United States item difficulty, the mean gender differences for the Impact Index were no longer statistically significant for either content or cognitive demand. However, within algebra, the advantage that girls had on knowing/procedures items approached statistical significance ($M = .22$, $SE = .12$, $p < .08$), whereas boys continued to outperform girls on knowing/procedures items within geometry ($M = -.43$, $SE = .16$, $p < .01$).

A summary of the number of items that favored boys and girls within each content area and cognitive demand is reported in Table 5. For the Impact index, almost twice as many items favored boys as girls in the United States. Boys were more likely to succeed on an item within fractions and number sense, measurement, and geometry. Girls, in contrast, were more likely to succeed on items within the content categories of algebra and data representation. Furthermore, boys were also more likely to succeed on items within both cognitive demands.

Spain. International item difficulty also had a statistically significant effect on gender differences in Spain (Table 6). Content category had a statistically significant effect on Impact after controlling for the international item difficulty, but cognitive level did not have an effect after controlling for item difficulty and content category. Unlike the United States, the interaction between the international item difficulty and content category was not statistically significant, whereas the interaction between the international item difficulty and cognitive demand was statistically significant. Even though there was no interaction between item difficulty and content, there was a statistically significant correlation between item difficulty and gender differences within the content category of data representation, analysis, and probability as observed in the United States, $r(39) = -.48$, $p < .05$ (Table 2). The interaction between item difficulty and gender differences within cognitive demand appeared to be attributable to the change in direction and magnitude of the Pearson correlation between knowing/procedures and reasoning/problem solving (Table 7). Whereas gender differences within knowing/procedures were significantly correlated to item difficulty, $r(87) = -.29$, $p < .01$, gender differences within reasoning/problem solving were not related to item difficulty, $r(33) = .04$, n.s. More difficult items within the lower cognitive demand favored Spanish boys. The interactions between content and cognitive demand and between difficulty, content, and cognitive demand were not statistically significant.

After controlling for the international item difficulty, mean differences were significantly different from 0 for both data representation, analysis, and probability and knowing/procedures (Tables 3 and 4). In both cases, these categories favored boys. Within the content categories of measurement; geometry; and data representation, analysis, and probability, boys had the advantage in the knowing/procedures items ($M = -.72$, $SE = .20$, $p < .001$; $M = -.36$, $SE = .18$, $p < .05$; $M = -.62$, $SE = .31$, $p < .05$, respectively).

Like the United States, the relationship between gender difference and the Spanish difficulty index was stronger (Table 6). Furthermore, the relationship between gender differences and content, after controlling for the Spanish item difficulty, was not statistically significant. However, this effect approached significance $\underline{F}(5,100) = 2.09$, $\underline{p} = .0734$. The interaction between cognitive demand and Spanish item difficulty, however, continued to be statistically significant. Item difficulty and Impact were related within the knowing/procedures cognitive demand (Table 7); a positive correlation indicated that boys performed better on more difficult items within this cognitive demand for the country specific difficulty index. Although there was not a statistically significant relationship interaction between content and Spanish item difficulty, there were statistically significant relationships between item difficulty and gender differences within the content categories of fractions and number sense and data representation, analysis, and probability (Table 2). In fact, these correlations between gender differences and Spanish item difficulty were stronger than the relationships between gender differences and the international difficulty index within these content areas.

After controlling for item difficulty in Spain, the mean gender differences based on Spain's Impact Index are presented in Tables 3 and 4. Within the content category of data representation, analysis, and probability, boys had the advantage. In addition, the advantage that boys had in geometry approached statistical significance ($\underline{M} = -.31$, $\underline{SE} = .16$, $\underline{p} < .06$). The contrast of the data representation, analysis, and probability items with the fractions and number sense items was statistically significant, whereas the contrast between cognitive demands was not. However, Spanish boys were more likely to succeed on Knowing/Procedure items (Table 4). Within the content categories of measurement; geometry; and data representation, analysis, and probability, boys continued to have an advantage on mathematics items that were classified as knowing/procedures ($\underline{M} = -.71$, $\underline{SE} = .18$, $\underline{p} < .001$; $\underline{M} = -.38$, $\underline{SE} = .17$, $\underline{p} < .05$; $\underline{M} = -.65$, $\underline{SE} = .26$, $\underline{p} < .05$, respectively).

For the Impact Index, the number of items that favored boys and girls within each content category and cognitive demand is presented in Table 5. Spanish boys outperformed Spanish girls on over three times as many mathematics items. Boys, in general, were more likely to succeed on items within all content categories, except algebra, and the lower cognitive demand, knowing/procedures.

DIF Index

United States. The results for the third ANOVA, based on the DIF Index, which controls for student achievement in mathematics, are reported in Table 8. The international item difficulty index did not have an effect on the second measure of gender difference, DIF, in the United States although the relationship approached significance, $F(1,100) = 3.34$, $p = .0704$. After controlling for the international item difficulty, content category continued to have a statistically significant effect. In addition, the interaction between content category and international item difficulty remained from Impact to DIF. Within the content category of fractions and number sense, the DIF index was related to the international difficulty index (Table 9). Boys continued to perform better on more difficult items within this category. No other statistically significant relations were detected although the relationship between item difficulty and gender difference within data representation, analysis, and probability approached significance, $r(16) = -.44$, $p < .07$.

After controlling for the international difficulty, the mean difference was significantly different from 0 within the content area of measurement (Table 10); again, this difference favored boys within both cognitive demands, knowing/procedures and reasoning/problem solving, respectively ($M = -0.39$, $SE = .19$, $p < .05$; $M = -.93$, $SE = .45$, $p < .05$). Measurement differed significantly from the other content categories with the exception of proportionality while reasoning/problem solving differed significantly from the lower cognitive demand, knowing/procedures. Within algebra, girls outperformed boys on only knowing/procedures items ($M = .47$, $SE = .16$, $p < .01$). Overall, boys continued to succeed on the higher cognitive demand (Table 11).

The United States item difficulty had a statistically significant effect on gender differences even when student achievement was controlled (Table 8). The relationship between the DIF index and United States item difficulty was stronger than the relationship observed between DIF and the international index as expected. Content continued to have an effect on gender differences even though the United States item difficulty was controlled. The interaction, however, between United States item difficulty and DIF was not statistically significant in this case although there was a statistically significant relationship between DIF and item difficulty within data representation, analysis, and probability (Table 9).

Whereas the magnitude of the predicted mean score, after controlling for United States difficulty, was relatively large within both measurement and reasoning/problem solving, the gender differences in favor of boys were not statistically significant due to the large variability (Tables 10 and 11). Girls outperformed boys in algebra within knowing/procedures ($M = .43$, $SE = .15$, $p < .01$) while boys succeeded on measurement items within knowing/procedures

($M = -.42$, $SE = .20$, $p < .05$). Although the magnitude of the boys' performance in geometry within reasoning/problem solving was higher than that of knowing/procedures ($M = -1.43$, $SE = 1.14$, $p < .22$), there was no statistically significant difference due to the large variability.

When controlling for the achievement of the student, girls were more likely to succeed on a mathematics item (Table 12). In general, girls and boys were equally likely to succeed on mathematics items within most content categories and both cognitive demands. However, girls were still more likely to succeed on algebra items, whereas boys were more likely to succeed on measurement items. Interestingly, more geometry items favored girls when controlling for student achievement.

Spain. The results for the ANOVA based on the DIF index in Spain are presented in Table 13. The international difficulty index had no statistically significant effect on gender differences when controlling for student achievement, but the relationship between difficulty and DIF approached significance, $F(1,100) = 3.51$, $p = .0639$. After controlling for the international item difficulty, the content of the item continued to have an effect on gender differences. No interactions between variables were detected.

Although there were no statistically significant gender differences within each content area after controlling for international item difficulty, boys continued to outperform girls within knowing/procedures (Tables 11). Within geometry, Spanish boys also outperformed Spanish girls in the lower cognitive demand, knowing/procedures ($M = -.59$, $SE = .23$, $p < .01$).

The Spanish item difficulty had a strong effect on DIF (Table 13). After controlling for the Spanish item difficulty, the relationship between gender differences and content category approached significance, $F(5,100) = 2.30$, $p = .0508$. The interaction between the Spanish item difficulty and cognitive demand was also statistically significant. Once more, gender differences, when controlling for achievement, and item difficulty had a statistically significant association with knowing and procedures (Table 13). Boys performed better on more difficult items within this cognitive demand. In addition, boys were more likely to outperform girls within the content categories of fractions and number sense and data representation, analysis, and probability when the items became more difficult (Table 9).

When controlling for Spanish item difficulty, boys were more likely to succeed on items within data representation, analysis, and probability even when student achievement was controlled (Table 10). Moreover, the difference between data representation, analysis, and probability and fractions and number sense was statistically significant. Overall, Spanish boys were more likely to succeed on items that were classified as knowing/procedures in contrast to boys in the United States (Table 11). Furthermore, Spanish boys were more likely to perform better on knowing/procedures items within the content areas of measurement and data representation, analysis, and probability ($\underline{M} = -.58$, $\underline{SE} = .22$, $\underline{p} < .01$; $\underline{M} = -.60$, $\underline{SE} = .31$, $\underline{p} < .05$, respectively).

After controlling for student achievement, the number of items that favored boys and girls in Spain is presented in Table 12. As in the United States, girls performed relatively better when they were matched with boys on mathematics achievement. Girls and boys performed similarly in all content areas with the exception of measurement and knowing/procedures, both this content category and cognitive demand favored boys.

Discussion

The purpose of the study was to investigate the relationships between gender differences at the item level and item characteristics across Spain and the United States. The first objective of the study was to determine whether observed gender differences on mathematics items were related to item characteristics (item difficulty, content, and cognitive demand). The results indicated that gender differences were related to item characteristics in both the United States and Spain. In general, gender differences in both countries were related to item difficulty and content category, controlling for item difficulty, as found in previous research (Engelhard, 1990). However, unlike the findings of Engelhard, cognitive demand was not related to gender differences after controlling for both item difficulty and content category. We should point out, however, that the cognitive demand in the present study was collapsed into two categories unlike the previous study, which contained three cognitive levels: computation, comprehension, and application.

Although Bielinski and Davison (1998) indicated that mathematics content did not explain gender differences, these results suggested that the relationship between gender differences and difficulty also depended on the content category of the item. Indeed, the strength of the association between the difficulty of the mathematics item and gender differences was stronger, but, in general, gender differences continued to be related to the content of the item even after controlling for item difficulty. Furthermore, the relationships between gender differences and item difficulty within each content category varied. Interactions between item difficulty and other item characteristics--content in the United States and cognitive demand in Spain--were also detected. Overall, gender differences in both countries were related to item difficulty within the content areas of fractions and number sense and data representation, analysis, and probability. In both content areas, boys performed relatively better than girls as the difficulty of those items increased. However, this was not true within the other content areas with the exception of the relationship between United States item difficulty and Impact. Furthermore, the most difficult subtests did not always correspond with boys' success in a particular content category as suggested by Bielinski and Davison. Although not statistically significant, the results suggested that girls performed better on more difficult items within proportionality in the United States sample. In fact, the most difficult item, which was within proportionality, for both the international and United States samples favored girls. In addition, the most internationally difficult content categories were algebra and proportionality while the easiest category for both difficulty indices was data representation, analysis, and probability. Interestingly, the easiest content category, for both the international and the Spanish sample, favored boys in Spain. Because there were a limited number of items in the categories of proportionality, caution must be used in the interpretation of these results. Although narrowly defined sub-categories could permit some types of analysis, a highly specified content classification would not have allowed researchers to describe trends and changes in curriculum as well as making the framework applicable to all participating countries (Robitaille et al., 1993). These results would suggest that the relationship between item difficulty and gender differences is more complex than initially suggested by Bielinski and Davison (1998).

For the most part, however, the pattern of gender differences in the United States, related to content category and cognitive demand, confirmed previous research (Engelhard, 1990; Frost, Hyde, & Fennema, 1994; Garner and Engelhard, 1999; Hanna, 1988). For example, boys in the United States performed better in measurement after item difficulty was controlled. The number of measurement items that statistically favored boys even when student achievement was controlled also confirmed this relationship. Although the overall adjusted mean score in algebra did not statistically favor girls, girls outperformed boys on almost a quarter of the algebra items while boys outperformed girls on only one of the algebra items. Interestingly, the algebra item in which boys were more likely to succeed was also labeled reasoning/problem solving, the higher cognitive demand. In addition, the adjusted mean score in reasoning/problem solving within algebra favored boys while the algebra items labeled knowing/procedures favored girls.

As expected, boys in the United States sample were more likely to succeed on cognitively complex items. Although girls in the United States were more likely to be successful on the computation items in the Second International Mathematics Study (SIMS) (Engelhard, 1990), a similar comparison could not directly be made in the Third International Mathematics Science Study (TIMSS) because computation and comprehension problems were both found in the lower cognitive demand. Even though the adjusted means did not show the content area favoring either gender, boys performed better on more items within the fractions and number sense category for Impact in the United States. For this gender difference estimate, the items that favored girls tended to be items related to computation such as the subtraction of decimals or multiplication of decimals. Boys in contrast performed better on items that required students to either order or relate fractions, decimals, and percents and those items that integrated measurement. This may explain the inconsistencies in the direction of gender differences in this content area found in other research on gender differences in mathematics (Garner & Engelhard, 1999).

The second objective of the study was to explore whether the type of difficulty index and estimate of gender difference affected the relationship between item characteristics and gender differences. Although the country specific indices were highly correlated to the international difficulty index, gender differences in both countries tended to have stronger relationships to their respective country difficulty indices. In general, after controlling for the country specific difficulty index, the relationships between gender differences and content, as well as the interactions between item characteristics, were similar to those reported using the international difficulty index. Even when the relationship between gender differences and item characteristics were not statistically significant, the relationships generally approached statistical significance with the exception of the interaction between item difficulty and content in the United States for DIF. Although item difficulty is strongly related to gender differences as Bielinski and Davison (1998) suggested, other characteristics such as the content classification of the item are also related to the observed gender differences in mathematics.

In both countries, controlling for student achievement in mathematics appeared to reduce the differences between boys and girls. However, content was still associated with gender differences after controlling for both types of item difficulty. Furthermore, the adjusted mean score in measurement continued to favor boys in the United States after controlling for international item difficulty, although the difference disappeared due to the large variability when using the country specific difficulty index. Nonetheless, within the lower cognitive demand, the gender differences in measurement favored boys. In measurement, more items favored boys even when achievement was controlled in both countries while algebra items continued to favor girls in the United States. Similarly, the gender differences in algebra were in favor of girls within the lower cognitive demand. Girls, on the other hand, performed better than boys on algebra items that required the translation of words into algebraic symbols. Again, the only algebra item that favored boys in the United States was one that demanded higher cognitive functioning.

Moreover, this algebra item required students to use a balance to solve an algebraic equation, a visual method for finding the value of an unknown variable. The number of items that favored boys and girls in both measurement and algebra confirmed these observations.

For DIF, the number of items that favored either boys or girls was approximately equal within fractions and number sense. However, boys performed better on more difficult items within this content category in both countries and within data representation, analysis, and probability in Spain. Furthermore, the items that continued to favor boys in fractions and number sense tended to be related to measurement; fraction, decimal, and percent concepts; or application problems. Those items that tended to favor girls related to computation, which confirms earlier research (Engelhard, 1990; Frost, Hyde, & Fennema, 1994; Ryan & Fan, 1996) and may explain the discrepancies found in other studies (e.g., Garner & Engelhard, 1999). Perhaps the exploration of a secondary classification of the item content would provide an alternative way for detecting gender differences on mathematics items.

The final objective of the study was to explore whether the relationships between item characteristics and gender varied across cultures. Although the study was descriptive in nature, the results suggest that gender differences vary across countries (Beller & Gafni, 1996; Hanna, 1988, 1994; Karunuñgan & Engelhard, 1999). Even though the relationship between gender differences and item difficulty and between gender differences and content after controlling for difficulty in Spain were statistically significant, there was an interaction between difficulty and gender differences within cognitive demand rather than within content. Interestingly, Spanish boys were more likely to outperform Spanish girls within the lower cognitive demand, knowing/procedures, for both Impact and DIF, whereas in the United States, as expected, the higher cognitive demand tended to favor boys. Furthermore, more difficult items, using either measure of item difficulty within the lower cognitive demand, tended to favor boys in Spain for both Impact and DIF. Although this was true in the United States for the Impact Index, there was no statistically significant relationship between item difficulty and gender difference for the DIF index.

Moreover, the pattern between gender differences and content category differed from that of the United States. In Spain the results indicated that gender differences, using the Impact index and both measures of item difficulty, become more favorable toward boys as the content category moved from fractions and number sense to algebra; geometry; proportionality; measurement; and data representation, analysis, and probability. Although the arrangement changed from Impact to DIF, the results were similar. Data representation, analysis, and probability items were still more likely to favor boys after controlling for Spain's level of item difficulty. These differences may be due to curricular differences (Hanna, 1988) or other cultural variables. For example, in an analysis of eighth grade mathematics books, Howson (1995) noted that Spain omitted the study of probability and statistics in grade 8. Other researchers, using regression analysis (Byrnes & Takahira, 1993), have noted that prior knowledge and strategy use explained nearly 50% of the variance in Scholastic Aptitude Test (SAT) scores, whereas gender explained no unique variance even though male students outperformed female students overall on the SAT mathematics items. Perhaps the higher performance of boys in Spain on data representation, analysis, and probability is related to this topic's omission in the eighth grade curriculum.

Karunuñgan & Engelhard (1999) suggested that girls also outperform boys in content areas other than algebra. In a recent study that examined gender differences in Singapore, Karunuñgan & Engelhard found that girls in general succeeded on more items in every content category except measurement in TIMSS. Even within measurement, a category that internationally favors boys, boys scored higher than girls on only 2 of the 13 items within that category. In addition, girls in Singapore continued to score better on more algebra items. Because students in Singapore scored higher than students in the United States, the items were relatively easier in Singapore. Future studies should examine the relationship between item characteristics and gender differences in other countries, particularly looking at countries that performed at different levels of achievement and at the role of item difficulty in gender differences. This may raise suspicions about the widely accepted belief that boys perform better than girls in mathematics, specifically in measurement, geometry, and higher level problem solving (Engelhard, 1990; Frost, Hyde, & Fennema, 1994; Hanna, 1988; Ryan & Fan, 1996).

Because the culture in the United States is not homogeneous, it would, therefore, be prudent to investigate gender differences using the United States sample. Racial-ethnic background are rarely examined in the context of gender differences in mathematics achievement (Tate, 1997). Previous research (e.g., Fan, Chen, & Matsumoto, 1997) indicated that gender differences do not occur in similar ways among different ethnic groups within the United States. Although trends in gender differences in mathematics were consistent for Whites, Asian, and Hispanics, African American students showed an opposite pattern: Girls had a slight advantage over boys in the 8th, 10th, and 12th grades. Due to the oversampling of minority students in the United States population, it would be interesting to investigate patterns in gender differences within various ethnic groups.

Because this study was intended to be descriptive in nature, certain limitations need to be addressed. The methodology of both this study and large-scale surveys of academic performance in general do not support causal inferences. Although one of the major strengths in conducting secondary analyses on large representative samples is the sensitivity of statistical tests to detect small gender differences, these differences need to be interpreted cautiously. Although some of the statistically significant differences were small, the patterns in gender differences that emerged suggested that gender differences in mathematics vary across cultures. Furthermore, international comparisons need to be made judiciously because of the multiple variables that need to be taken into account (Robitaille & Travers, 1992). The present study did not address possible explanatory variables such as the curriculum, opportunity to learn, and other cultural and student variables that might relate to these gender differences. Another potential limitation of this study is that these items went through a prior bias review (Garden & Orpwood, 1996); the sensitivity of the items to detect gender differences may be attenuated to an unknown degree. In addition, the secondary data researcher has no control over the design of the instruments in the study. Some of the items were originally classified into two content areas, but only the principal label was retained in the reporting of results (Garden & Orpwood, 1996). The original classification system may have clarified the inconsistencies and provided more information about the relationship between mathematics and gender differences within content areas.

Because of the nature of large-scale studies, it is difficult to assess the strategies used to solve complex problems. In a recent study, Fennema, Carpenter, Jacobs, Franke, & Levi (1998a, 1998b) found that gender differences in problem-solving strategies were observed as early as grades 1-3. No gender differences were found in the ability to solve any problems with the exception of an extension problem, which favored the boys in grade 3. However, girls tended to use concrete solution strategies while the boys were more likely to employ invented algorithms in their problem solving strategies. Those students, both boys and girls, who were able to use invented algorithms in the earlier grades were better able to solve the extension problems by grade 3. Although the TIMSS data provided some information on the processes used by students in solving certain extended response and performance assessment problems, qualitative interviews of students about their problems solving strategies, particularly the strategies employed by students engaged in these items, would provide further information about the nature of gender differences.

Conclusion

Despite the limitations, these results have important implications for interpreting gender differences in mathematics achievement. First, although there were no mean gender differences on the total scores in the United States as in Spain (Beaton et al., 1996), micro-level analysis of item characteristics must be considered in interpreting results. Even within categories, the direction of gender differences varied depending on other characteristics of the item. When using either Impact or DIF as a measure of gender difference, the examined item characteristics accounted for a large percentage of the variance in gender differences, approximately 30% of the variance after controlling for student achievement in both countries. Furthermore, the role of item difficulty in gender difference research is receiving more attention in the literature (Bielinski & Davison, 1998). Item difficulty was indeed related to gender differences in both countries, particularly when the country specific difficulty indices were used. Nevertheless, the results suggest that difficulty may interact with other item characteristics such as content and cognitive complexity. Both the definition of difficulty, whether it is specific to a particular population or independent of the population, and the nature of the concept of difficulty in assessment need to be further addressed to clarify its relationship with other item characteristics and opportunity to learn. Is an item inherently difficult or is its difficulty related to the degree to which it is taught in school or at home? If an item's difficulty is related to a student's opportunity to learn, changes in curriculum and instruction practices may be part of the solution to gender differences in mathematics performance.

Whereas the results from the United States sample tended to replicate previous research, results from the Spanish sample indicated that boys outperformed girls on items that typically favor girls in the United States. Because gender differences may vary across cultures, socio-cultural models, rather than biological factors, might explain the gender differences observed in mathematics achievement. Although most studies on gender differences have been investigated in the United States, cross-cultural studies can help to clarify the complex nature of gender differences in mathematics achievement. If researchers are to continue to explore gender differences in mathematics achievement so as to inform educational policy, develop teaching strategies, and clarify the theoretical basis of gender differences, they will need to address the role of cultural variables and opportunity to learn on gender differences in mathematics.

References

Adams, R. J., & Gonzalez, E. J. (1996). The TIMSS test design. In M. O. Martin & D. L. Kelly, D. L. (Eds.), TIMSS technical report, volume I: Design and development. Chestnut Hill, MA: Boston College.

Alberdi, I., & Alberdi, I. (1991). Spain. In M. Wilson (Ed.), Girls and young women in education: A European perspective. (pp. 153-170). Oxford, England: Pergamon Press.

Baker, D. P., & Jones, D. P. (1993). Creating gender equality: Cross-national gender stratification and mathematical performance. Sociology of Education, 66, 91-103.

Barrio, J. F. (1999). The division of educational responsibilities. Retrieved September 26, 1999 from the World Wide Web: http://www.docuweb.ca/SiSpain/english/educatio/.

Beaton, A., Mullis, I. V. S., Martin, M., Gonzalez, E. J., Kelly, D., & Smith, T. (1996). Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study. Chestnut Hill, MA: Boston College.

Beller, M., & Gafni, N. (1996). The 1991 International Assessment of Educational Progress in Mathematics and Sciences: The gender differences perspective. Journal of Educational Psychology, 88, 365-377.

Benbow, C. P., & Lubinski, D. (1997). Psychological profiles of mathematically talented: Some sex differences and evidence supporting their biological basis. In M. R. Walsh (Ed.), Women, men, and gender. (pp. 274-282). New Haven, CT: Yale University Press.

Bielinski, J., & Davison, M. L. (1998). Gender differences by item difficulty interactions in multiple-choice items. American Educational Research Journal, 35, 455-476.

Bishop, A. J. (1988). Mathematics education in its cultural context. In A. J. Bishop (Ed.), Mathematics education and culture. (pp. 179-191). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Byrnes, J. P., & Takahira, S. (1993). Explaining gender differences on SAT-math items. Developmental Psychology, 29, 805-810.

Cohen, J. (1992). A power primer. Psychological Bulletin, 112, 155-159.

Dresselhaus, M. S., Franz, J. R., & Clark, B. C. (1994). Interventions to increase the participation of women in physics. Science, 263, 1392-1393.

Engelhard, G. (1990). Gender differences in performance on mathematics items: Evidence from the United States and Thailand. Contemporary Educational Psychology, 15, 13- 26.

Fan, X., Chen, M., Matsumoto, A. R. (1997). Gender differences in mathematics achievement: Findings for the National Education Longitudinal Study of 1988. Journal of Experimental Education, 65, 229-242.

Feingold, A. (1992). Sex differences in variability in intellectual abilities: A new look at an old controversy. Review of Educational Research, 62, 61-84.

Fennema, E. (1990). Justice, equity, and mathematics education. In E. Fennema & G. C. Leder (Eds.), Mathematics and gender. (pp. 1-9). New York: Teachers College Press.

Fennema, E., Carpenter, T. P., Jacobs, V. R., Franke, M. L., Levi, L. W. (1998a). A longitudinal study of gender differences in young children's mathematical thinking. Educational Researcher, 27, 6-11.

Fennema, E., Carpenter, T. P., Jacobs, V. R., Franke, M. L., Levi, L. W. (1998b). New perspectives on gender differences in mathematics: A reprise. Educational Researcher, 27, 19-21.

Fierros, E. G. (1999, April). Examining gender differences in mathematics achievement on the Third International Mathematics and Science Study (TIMSS). Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec.

Foy, P. (1997). Calculation of sampling weights. In M. O. Martin and D. L. Kelly (Eds.), TIMSS technical report volume II: Implementation and analysis. (pp. 71-79). Chestnut Hill, MA: Boston College.

Foy, P., Rust, K. & Schleicher, A. (1996). Sample design. In M. O. Martin & D. L. Kelly, D. L. (Eds.), TIMSS technical report, volume I: Design and development. Chestnut Hill, MA: Boston College.

Frost, L. A., Hyde, J. S., & Fennema, E. (1994). Gender, mathematics performance, and mathematics-related attitudes and affect: A meta-analytic synthesis. International Journal of Educational Research, 21, 373-384.

Garden, R. A., & Orpwood, G. (1996). Development of the TIMSS achievement tests. In M. O. Martin & D. L. Kelly, D. L. (Eds.), TIMSS technical report, volume I: Design and development. Chestnut Hill, MA: Boston College.

Garner, M., & Engelhard, G. (1999). Gender differences in performance on multiple-choice and constructed response mathematics items. Applied Measurement in Education, 12, 29-51.

Gil, G. A. (1994). Spain: System of education. In T. Husen & T. N. Postlethwaite (Eds.), International encyclopedia of education. (pp. 901-911). Oxford, England: Pergamon.

Gonzalez, E., & Smith, T. A. (Eds.). (1997). User guide for the TIMSS international database: Primary and middle school years. Chestnut Hill, MA: Boston College.

Hanna, G. (1988, February). Gender differences in mathematics achievement among eighth graders: Results from twenty countries. Paper presented at the annual meeting of the American Association for the Advancement of Science, Boston.

Hanna, G. (1994). Cross-cultural gender differences in mathematics education. International Journal of Educational Research, 21, 417-425.

Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel Procedure. In H. Wainer & H. I. Braun (Eds.), Test validity (pp. 129-145). Hillsdale, NJ: Lawrence Erlbaum.

Howson, G. (1995). TIMSS monograph no. 3: Mathematics textbooks-A comparative study of grade 8 texts. Vancouver, BC: Pacific Educational Press.

Hyde, J. S. (1997). Gender differences in math performance: Not big, not biological. In M. R. Walsh (Ed.), Women, men, and gender. (pp. 283- 287). New Haven, CT: Yale University Press.

Instituto Nacional de Calidad y Evaluacion (INCE). (1997). Preliminary report of the general diagnosis of the educational system. Retrieved September 26, 1999 from the World Wide Web: http://www.ince.mec.es/diag/ejem.htm.

Karunuñgan, M. L., & Engelhard, G. (1999, October). Mother's education and the mathematics performance of eighth grade girls and boys in Singapore and the United States. Paper presented at the annual meeting of the Georgia Educational Research Association, Morrow, Georgia.

Kupermintz, H., & Snow, R. (1997). Enhancing the validity and usefulness of large-scale educational assessments: III. NELS:88 mathematics achievement to 12th grade. American Educational Research Journal, 34, 124-150.

Lane, S., Wang, N., & Magone, M. (1996). Gender-related differential item functioning on a middle-school mathematics performance assessment. Educational Measurement: Issues and Practice, 15, 21-27, 31.

Leder, G. C. (1990). Gender differences in mathematics: An overview. In E. Fennema & G. C. Leder (Eds.), Mathematics and gender. (pp. 1-9). New York: Teachers College Press.

Leder, G. C. (1992). Mathematics and gender: Changing perspectives. In D. A. Grouws (Ed.), Handbook of research on mathematics teaching and learning. (pp. 597-622) New York: Macmillan.

Martin, M. O., & Kelly, D. L. (Eds.). (1996). TIMSS technical report, volume I: Design and development. Chestnut Hill, MA: Boston College.

Martin, M. O., & Kelly, D. L. (Eds.). (1997). TIMSS technical report, volume II: Implementation and analysis-Primary and middle school years. Chestnut Hill, MA: Boston College.

Maxwell, B. (1996). Translation and cultural adaptation of the survey instruments. In M. O. Martin and D. L. Kelly (Eds.), TIMSS technical report volume I: Design and development. Chestnut Hill, MA: Boston College.

Noddings, N. (1998). Perspectives from feminist philosophy. Educational Researcher, 27, 17-18.

Reyes, L. H., & Stanic, G. M. (1988). Race, sex, socioeconomic status and mathematics. Journal for Research in Mathematics Education, 19, 26-43.

Robitaille, D. F., McKnight, C. C., Schmidt, W. H., Britton, E. D., Raizen, S. A., and Nicol, C. (1993). TIMSS monograph no. 1: Curriculum frameworks for mathematics and science. Vancouver, BC: Pacific Educational Press.

Robitaille, D. F., & Travers, K. J. (1992). International studies of achievement in mathematics. In D. A. Grouws (Ed.), Handbook of research on mathematics teaching and learning. (pp. 687-709). New York: Macmillan.

Ryan, K. E., & Fan, M. (1996). Examining gender DIF on a multiple-choice test of mathematics: A confirmatory approach. Educational Measurement: Issues and Practice, 15-20, 38.

Stromquist, N. P. (1989). Determinants of educational participation of women in the Third World: A review of the evidence and a theoretical critique. Review of Educational Research, 59, 143-183.

Tate, W. F. (1997). Race-ethnicity, SES, gender, and language proficiency trends in mathematics achievement: An update. Journal for Research In Mathematics Education, 28, 676-677.

TIMSS (1996). TIMSS mathematics items-Released set for population 2 (seventh and eighth grades). Chestnut Hill, MA: Boston College.

U.S. Department of Education. (1996). National Center for Educational Statistics. Pursuing excellence: A study of U.S. eighth-grade mathematics and science teaching, learning, curriculum, and achievement in international context. (NCES 97-198). Washington, DC: U.S. Government Printing Office.

Willingham, W. W., & Cole, N. S. (1997). Gender and fair assessment. Hillsdale, NJ: Lawrence Erlbaum.

Table 1
Analysis of Variance Summary for United States (Impact Index)

Source of variation	SS	df	F	p
International difficulty index				
Item difficulty (D)	3.250	1	11.74	.0009
Content category (A)	4.242	5	3.07	.0129
Cognitive demand (B)	0.663	1	2.40	.1248
D x A	3.825	5	2.76	.0221
D x B	0.352	1	1.27	.2621
A x B	0.675	5	0.49	.7848
D x A x B	1.318	5	0.95	.4506
Error	27.676	100		
USA difficulty index				
Item difficulty (D)	7.767	1	29.87	.0001
Content category (A)	2.719	5	2.09	.0727
Cognitive demand (B)	.450	1	1.73	.1911
D x A	2.731	5	2.10	.0715
D x B	.236	1	0.91	.3434
A x B	.879	5	0.68	.6427
D x A x B	1.216	5	0.94	.4615
Error	26.003	100		

Note. Sequential sums of squares (Type I SS) are reported here. For the international difficulty index and the United States difficulty index, the r^2 was .34 and .38, respectively.

Table 2
Correlation Between Gender Differences and Difficulty Within Content Category for the United States and Spain (Impact)

	Content Category					
	Number Sense (N = 41)	Algebra (N = 22)	Measurement (N = 13)	Geometry (N = 22)	Data Representation (N = 18)	Proportionality (N = 8)
International difficulty index						
United States	-.53***	-.23	-.17	-.09	-.52*	.43
Spain	-.28	-.07	.01	-.03	-.49*	-.48
Country difficulty index						
United States	.54***	.49*	.10	.30	.35	-.05
Spain	.47**	.22	-.02	-.06	.63**	.48

Note. A negative correlation between gender differences and international difficulty within content indicates that boys perform better on more difficult items. A positive correlation between gender differences and the derived country difficulty index within content indicates that boys perform better on more difficult items. The country difficulty index refers to the computed proportion correct difficulty index for each country.

*$p < .05$. **$p < .01$. ***$p < .001$.

Table 3

Adjusted Mean Scores by Content Category for United States and Spain (Impact Index)

	Content Category					
	Data Representation (N = 18)	Algebra (N = 22)	Number Sense (N = 41)	Geometry (N = 22)	Proportionality (N = 8)	Measurement (N = 13)
International difficulty index						
United States						
M	.01	-.04	-.08	-.16	-.35	-.64**
SE	.16	.15	.09	.13	.20	.20
Spain						
M	-.45*	-.08	-.07	-.31	-.34	-.41
SE	.20	.18	.12	.17	.25	.25
Country difficulty index						
United States						
M	-.06	-.10	-.10	-.08	-.27	-.40
SE	.16	.16	.09	.15	.19	.47
Spain						
M	-.49***	-.13	-.06	-.31	-.32	-.38
SE	.18	.18	.11	.16	.24	.24

Note. Scores are adjusted least square means. Girls are more likely to succeed on items with positive values on the Impact Index while boys are more likely to succeed when the values are negative. The asterisk indicates mean gender differences that are significantly different from 0. The country difficulty index refers to the computed proportion correct difficulty index for each country.

*p < .05. **p < .01. ***p < .001.

Table 4
Adjusted Mean Scores by Cognitive Demand for United States and Spain (Impact Index)

	Cognitive demand	
	Knowing/Procedures (N = 89)	Reasoning/Problem Solving (N = 35)
International difficulty index		
United States		
M	-.07	-.35**
SE	.07	.11
Spain		
M	-.38**	-.17
SE	.09	.13
Country difficulty index		
United States		
M	-.05	-.29
SE	.07	.18
Spain		
M	-.23*	.03
SE	.10	.15

Note. Scores are adjusted least square means. Girls are more likely to succeed on items with positive values on the Impact Index while boys are more likely to succeed when these values are negative. The asterisk indicates mean gender differences that are significantly different from 0. The country difficulty index refers to the computed proportion correct difficulty index for each country.

* $p < .05$. **$p < .01$. **$p < .0001$

Table 5

Number of Mathematics Items that Statistically Favored Boys and Girls in the United States and Spain (Impact Index)

		United States		Spain	
Content		Boys	Girls	Boys	Girls
Fractions/Number Sense	41	10	4	6	3
Algebra	22	1	4	3	2
Measurement	13	4	0	6	0
Geometry	22	3	1	4	1
Data Representation	18	1	2	4	2
Proportions	8	2	1	3	0
Total	124	21	12	26	8
Cognitive Demand					
Knowing/Procedures	89	14	9	22	5
Problem Solving	35	7	3	4	3
Total	124	21	12	26	8

Note: Calculations were based on Impact. Values were statistically significant at $p < .05$.

Table 6
Analysis of Variance Summary for Spain (Impact Index)

Source of variation	SS	df	F	p
International difficulty index				
Item difficulty (D)	2.540	1	5.92	.0168
Content category (A)	4.978	5	2.32	.0488
Cognitive demand (B)	0.511	1	1.19	.2779
D x A	2.419	5	1.13	.3511
D x B	1.878	1	4.37	.0390
A x B	0.809	5	0.38	.8636
D x A x B	1.168	5	0.95	.7426
Error	42.932	100		
Spain difficulty index				
Item difficulty (D)	7.173	1	18.65	.0001
Content category (A)	4.022	5	2.09	.0726
Cognitive demand (B)	0.566	1	1.47	.2281
D x A	3.338	5	1.74	.1333
D x B	2.064	1	5.37	.0226
A x B	0.710	5	.37	.8687
D x A x B	0.896	5	.47	.8007
Error	38.464	100		

Note. Sequential sums of squares (Type I SS) are reported here. For the international difficulty index and the Spanish difficulty index, the r^2 was .25 and .33, respectively.

Table 7

Correlation Between Gender Differences and Difficulty Within Cognitive Demand for the United States and Spain (Impact)

	Cognitive demand	
	Knowing/Procedures (N = 89)	Reasoning/Problem Solving (N = 35)
International difficulty index		
United States	-.22*	-.43**
Spain	-.29**	.04
Country difficulty index		
United States	.39****	.58***
Spain	.45****	.03

Note. A negative correlation between gender differences and international difficulty within cognitive demand indicates that boys perform better on more difficult items. A positive correlation between gender differences and the derived country difficulty index within cognitive demand indicates that boys perform better on more difficult items. The country difficulty index refers to the computed proportion correct difficulty index for each country.

*$p < .05$. **$p < .01$. ***$p < .001$. ****$p < .0001$.

Table 8
Analysis of Variance Summary for United States (DIF Index)

Source of variation	SS	df	F	p
International difficulty index				
Item difficulty (D)	1.355	1	3.34	.0704
Content category (A)	6.048	5	2.99	.0149
Cognitive demand (B)	0.865	1	2.14	.1470
D x A	5.033	5	2.48	.0364
D x B	0.137	1	0.34	.5625
A x B	1.054	5	0.52	.7605
D x A x B	1.409	5	0.70	.6280
Error	56.414	100		
USA difficulty index				
Item difficulty (D)	5.305	1	13.39	.0004
Content category (A)	4.750	5	2.40	.0425
Cognitive demand (B)	0.645	1	1.63	.2049
D x A	3.370	5	1.70	.1412
D x B	0.153	1	0.39	.5358
A x B	1.228	5	0.62	.6849
D x A x B	1.346	5	0.68	.6399
Error	39.616	100		

Note. Sequential sums of squares (Type I SS) are reported here. For the international difficulty index and the United States difficulty index, the r^2 was .28 and .30, respectively.

Table 9

Correlation Between Gender Differences and Difficulty Within Content Category for the United States and Spain (DIF)

	Content Category					
	Number Sense (N = 41)	Algebra (N = 22)	Measurement (N = 13)	Geometry (N = 22)	Data Representation (N = 18)	Proportionality (N = 8)
International difficulty index						
United States	-.45**	-.08	-.17	.02	-.44	.43
Spain	-.21	-.11	.10	-.03	-.48*	-.34
Country difficulty index						
United States	.41**	.32	-.18	.21	.25	-.05
Spain	.37**	.23	-.07	-.03	.63**	.38

Note. A negative correlation between gender differences and international difficulty within content indicates that boys perform better on more difficult items. A positive correlation between gender differences and the derived country difficulty index within content indicates that boys perform better on more difficult items. The country difficulty index refers to the computed proportion correct difficulty index for each country.

*$p < .05$. **$p < .01$.

Table 10

Adjusted Mean Scores by Content Category for United States and Spain (DIF Index)

	Content Category					
	Algebra (N = 22)	Number Sense (N = 18)	Data Representation (N = 41)	Geometry (N = 22)	Proportionality (N = 8)	Measurement (N = 13)
International difficulty index						
United States						
M	.09	.02	.00	-.07	-.28	-.66**
SE	.18	.11	.19	.16	.24	.25
Spain						
M	.18	.16	-.34	-.14	-.14	-.24
SE	.21	.14	.23	.19	.29	.29
Country difficulty index						
United States						
M	.03	-.01	-.05	-.01	-.19	-.93
SE	.19	.11	.20	.18	.24	.58
Spain						
M	.11	.16	-.40*	-.13	-.12	-.20
SE	.21	.13	.21	.19	.28	.28

Note. Scores are adjusted least square means. Girls are more likely to succeed on items with positive values on the DIF Index while boys are more likely to succeed when the values are negative. The asterisk indicates mean gender differences that are significantly different from 0. The country difficulty index refers to the computed proportion correct difficulty index for each country.

$^*p < .05$. $^{**}p < .01$.

Table 11
Adjusted Mean Scores by Cognitive Demand for United States and Spain (DIF Index)

	Cognitive demand	
	Knowing/Procedures (N = 89)	Reasoning/Problem Solving (N = 35)
International difficulty index		
United States		
M	.01	-.31*
SE	.09	.13
Spain		
M	-.21*	.04
SE	.11	.16
Country difficulty index		
United States		
M	.03	-.40
SE	.09	.22
Spain		
M	-.23*	.03
SE	.10	.15

Note. Scores are adjusted least square means. Girls are more likely to succeed on items with positive values on the DIF Index while boys are more likely to succeed when these values are negative. The asterisk indicates mean gender differences that are significantly different from 0.The country difficulty index refers to the computed proportion correct difficulty index for each country.

*p < .05.

Table 12
Number of Mathematics Items that Statistically Favored Boys and Girls in the United States and Spain (DIF)

		United States		Spain	
Content		Boys	Girls	Boys	Girls
Fractions/Number Sense	41	10	7	4	5
Algebra	22	1	5	2	3
Measurement	13	3	1	5	0
Geometry	22	2	4	2	2
Data Representation	18	0	2	2	3
Proportions	8	1	1	1	1
Total	124	17	20	16	14
Cognitive Demand					
Knowing/Procedures	89	13	17	15	11
Problem Solving	35	4	3	1	3
Total	124	17	20	16	14

Note: Calculations were based on DIF. Values were statistically significant at $p < .05$.

Table 13
Analysis of Variance Summary for Spain (DIF Index)

Source of variation	SS	df	F	p
International difficulty index				
Item difficulty (D)	2.021	1	3.51	.0639
Content category (A)	7.152	5	2.49	.0364
Cognitive demand (B)	0.636	1	1.10	.2959
D x A	3.156	5	1.10	.3671
D x B	1.672	1	2.91	.0914
A x B	1.211	5	0.42	.8333
D x A x B	1.651	5	0.57	.7200
Error	57.551	100		
Spain difficulty index				
Item difficulty (D)	7.340	1	14.08	.0003
Content category (A)	5.988	5	2.30	.0508
Cognitive demand (B)	0.705	1	1.35	.2477
D x A	4.360	5	1.67	.1482
D x B	2.215	1	4.25	.0419
A x B	1.003	5	0.38	.8583
D x A x B	1.294	5	0.50	.7785
Error	52.146	100		

Note. Sequential sums of squares (Type I SS) are reported here. For the international difficulty index and the Spanish difficulty index, the r^2 was .23 and .31, respectively.

Table 14

Correlation Between Gender Differences and Difficulty Within Cognitive Demand for the United States and Spain (DIF)

	Cognitive demand	
	Knowing/Procedures (N = 89)	Reasoning/Problem Solving (N = 35)
International difficulty index		
United States	-.11	-.29
Spain	-.23*	.04
Country difficulty index		
United States	.09	.22
Spain	.39**	-.02

Note. A negative correlation between gender differences and international difficulty within cognitive demand indicates that boys perform better on more difficult items. A positive correlation between gender differences and the derived country difficulty index within cognitive demand indicates that boys perform better on more difficult items. The country difficulty index refers to the computed proportion correct difficulty index for each country.

*$p < .05$. **$p < .001$.

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

AERA
SE063504
ERIC®

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Gender Differences on Eighth Grade Mathematics Items: A Cross-Cultural Comparison of the United States and Spain	
Author(s): Tasha Calvert, George Engelhard, Jr.	
Corporate Source: Emory University	Publication Date: 4/28/00

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 ↑ ☐
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: Tasha N. Calvert	Printed Name/Position/Title: Tasha N. Calvert	
Organization/Address: Division of Educational Studies Emory University Atlanta, GA 30322	Telephone: (404) 728-4342	FAX: 404-727-2799
	E-Mail Address: tcalver@emory.edu	Date: 5/5/00

(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 2000

Dear AERA Presenter,

Congratulations on being a presenter at AERA. The ERIC Clearinghouse on Assessment and Evaluation would like you to contribute to ERIC by providing us with a written copy of your presentation. Submitting your paper to ERIC ensures a wider audience by making it available to members of the education community who could not attend your session or this year's conference.

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, electronic, and internet versions of *RIE*. The paper will be available **full-text, on demand through the ERIC Document Reproduction Service** and through the microfiche collections housed at libraries around the world.

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

To disseminate your work through ERIC, you need to sign the reproduction release form on the back of this letter and include it with **two** copies of your paper. You can drop of the copies of your paper and reproduction release form at the ERIC booth (223) or mail to our attention at the address below. **If you have not submitted your 1999 Conference paper please send today or drop it off at the booth with a Reproduction Release Form.** Please feel free to copy the form for future or additional submissions.

Mail to:

AERA 2000/ERIC Acquisitions
The University of Maryland
1129 Shriver Lab
College Park, MD 20742

Sincerely,

Lawrence M. Rudner

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

ERIC/AE is a project of the Department of Measurement, Statistics and Evaluation at the College of Education, University of Maryland.