

EXPLORING EIGHTH-GRADE MINORITY AND FEMALE STUDENTS' STATISTICAL LITERACY SKILLS USING TIMSS

Do Differences Exist?

Jamie D. Mills and Marsha E. Simon

University of Alabama

This study investigated the performance of U.S. eighth-grade female and minority students' achievement in learning statistics concepts using the Data and Chance content domain from the Trends in International Mathematics and Science Study 2007(2007). Using variables that have been linked to mathematics and statistics achievement in the literature, a hierarchical level modeling approach revealed significant achievement differences between boys and girls, as well as lower academic performances for minority students. The implications from this study suggest the need for additional research related to the teaching and learning of statistics concepts, with a particular focus on how to better engage and improve the achievement of female and minority students in PK–12 education.

INTRODUCTION

In today's changing world, numbers are a part of everyday life. If we expect our students to be able to read, understand, and make informed decisions about the numbers in which they encounter on a daily basis, a basic understanding of statistics concepts become important skills to have. We are constantly inundated with numbers and statistics on television, the Internet, and on many other popular forms of the social media (i.e., Facebook,

Twitter, etc.). Oftentimes, claims from published data such as these can have serious or important implications for everyone in our society. Understanding what an average daily balance means for one's checking account or even what the most recent Centers for Disease Control statistics suggest about the rate at which minority females are contracting the HIV infection are two telling examples. Therefore, to be able to understand, evaluate, and react appropriately to valid claims is necessary if our ultimate goal is for students to be

• Correspondence concerning this article should be addressed to: Janice D. Mills, jmills@ua.edu

Middle Grades Research Journal, Volume 10(3), 2015, pp. 1–22
Copyright © 2015 Information Age Publishing, Inc.

ISSN 1937-0814
All rights of reproduction in any form reserved.

informed and productive citizens as they enter into adulthood.

The ability to be able to critically evaluate and understand the implications of statistics involves one being “statistically” literate. According to Gal (2002), statistical literacy involves one being able to interpret basic statistical information or data and also being able to talk about what it means and offer opinions about the implications or conclusions. Perhaps the goal of many teachers might be to develop students’ skills as a mathematician or statistician; however, educating students to be good consumers of statistical information appears to be one reasonable objective to strive for in many classrooms.

The notion that students have struggled learning basic statistical concepts and ideas has been well documented in the literature. Research has shown that students have had a difficult time understanding sampling distributions (delMas, Garfield, & Chance, 1999; Saldanha & Thompson, 2001), measures of variability (delMas & Liu, 2005), and probability and chance ideas and concepts (Garfield, 2003; Konold, 1995; Shaughnessy, 1992). Other studies have shown that students have challenges reasoning about distributions and graphical representations of distributions (Ben-Zvi, 2004; Biehler, 1997; Hammerman & Rubin, 2004). These topics encompass necessary and important concepts for students to understand and use if a major goal is for the student to be able to critically analyze the information in the world around them.

Another major area of concern in statistics education today that is related to student achievement is teacher attitudes and training. How are PK–12 teachers managing in the classroom? Previous research has shown that many teachers who are teaching statistics have had very little formal training or they have never even taken a basic statistics course (Begg & Edwards, 1999; Franklin, 2000; Mills & Holloway, 2013). According to the Mathematical Education of Teachers II Report (Conference Board of the Mathematical Sciences, 2012), teachers at the middle and secondary

levels need to study more in the areas of statistics and probability (p. 68), while further research has shown that most elementary teachers have little or no experience in these fields (Conference Board of the Mathematical Sciences, 2001, p. 87). Additionally, previous studies have reported that teacher attitudes toward statistics are neutral to negative (Estrada & Batanero, 2008; Onwuegbuzie, 1998, 2003).

Empirical research related to student achievement in statistics is scarce in the literature. Most teachers recommend actively engaging students with hands-on activities (Garfield, 1993; Lovett & Greenhouse, 2000; Roseth, Garfield, & Ben-Zvi, 2008; Smith, 1998), as well as integrating technology into instruction to enhance teaching and learning (Garfield & Everson, 2009; Harrington, 1999; Mills & Raju 2011; Ward, 2004). Other variables shown to be related to achievement have included the number of previously taken mathematics courses (Gonul & Solano, 2013; Johnson & Kuennen, 2006; Lunsford & Poplin, 2011), school attendance (Gonul & Solano, 2013, and noncognitive factors, such as beliefs and attitudes about learning statistics (Dempster & McCorry, 2009; Gal & Ginsburg, 1994; Vanhoof, 2006).

In response to the growing need for improved statistics learning as well as training for teachers in PK–12 education, the Guidelines for Assessment and Instruction in Statistic Education (GAISE) report was published in 2007 by the American Statistical Association. These guidelines outline four major problem-solving components, which are consistent with the components in the National Council of Teachers of Mathematics (NCTM) Data Analysis and Probability Standard (1989, 1991, 2000). They also delineate specific student skills and objectives within each component to facilitate a student’s progression toward statistical literacy. The College Board (2006) has also published standards that focus on improving students’ data analysis and probability skills, not only for students who are entering college, but also for students who are expected

to use these skills in their work and everyday life. Thus, a recent shift toward refining and advancing statistics teaching and learning is apparent, particularly in PK–12 education.

Trends in International Mathematics and Science Study

The Trends in International Mathematics and Science Study, 2007 (TIMSS, 2007) is a study that is conducted every 4 years for students in the fourth and eighth grades. It provides extensive data on mathematics and science achievement for 59 participating countries, including the United States. Within the mathematics portion of this study, statistics achievement can be further analyzed using the Data and Chance content domain. Within this domain, students are evaluated on their ability in three cognitive domains: (1) applying, (2) knowing, and (3) reasoning. The topic areas include data interpretation, chance, and data organization and representation, all areas that are consistent with the objectives specified in the Data Analysis and Probability Standard (NCTM, 1989, 1991, 2000). It also provides background information on participating teachers, including their years of teaching experiences, education level, their perceptions about teaching statistical concepts, and many other demographic and contextual variables. Thus, this data can be useful in examining important student and teacher information as it relates specifically to statistics learning and teaching.

Numerous research studies have been conducted regarding TIMSS students' performance in mathematics (Caponera & Russo, 2010; Chen, Ferron, Thompson, Gorin, & Tatsuoka, 2010; Koretz, McCaffrey, & Sullivan, 2001; Martin, Mullis, Gregory, Hoyle, & Shen, 2000; Min-Hsiung, 2009; Neuschmidt, Barth, & Hastedt, 2008; O'Dwyer, 2005). However, statistics is thought to be a related content domain in mathematics, yet very little is known about student achievement since a fairly new data analysis and probability standard was first implemented in PreK–12 educa-

tion in most states in the United States (NCTM, 1989, 1991, 2000).

Specifically, only two TIMSS-related studies were found in the literature. One study by Kien-Kheng and Idris (2010) reported how eighth-grade students from Malaysia performed on the Data and Chance items compared to other students in other countries from 1999–2007. They found that students in Malaysia performed significantly lower than students in other countries, including students in the United States. More recently (Mills & Holloway, 2013), explored the relationship between student achievement in statistics and factors at the student and classroom level, including investigating teacher training as well as teacher perceptions and attitudes about teaching statistics. One important finding from their study revealed a statistically significant difference in achievement between eighth-grade boys and girls in the United States. Other findings showed that the majority of the students in the sample were reported as economically disadvantaged, as reported by their principals; and overall, these students appeared to lag behind in terms of learning statistics concepts and ideas as measured by the NCTM (1989, 1991, 2000) data analysis standards. The finding regarding the performance of girls and boys as well as how minority students are achieving in learning statistics concepts provided the basis for this research study. First though, it is important to provide a brief summary of how these populations have achieved in mathematics in the past as well as provide some notion as to where they stand today.

Previous Mathematics Achievement Research

It is well documented in the literature the latter finding regarding boys outperforming girls in mathematics as well as the fact that minority students continue to lag behind in terms of their mathematics achievement in America's schools. Previous research indicates that minority student achievement climbed

substantially during the 1970s and 1980s with a significant close in the gap (i.e., compared to White students); however, beginning in the early 1990s, a gradual reversal of achievement has been reported (Education Trust, 2003; Riegle-Crumb & Humphries, 2012).

Furthermore, research devoted to minority student achievement in mathematics has been controversial in the literature for several decades. Research has shown lower academic performances for African American students compared to their White counterparts, with evidence that this gap is evident as early as kindergarten (Cooper & Schleser, 2006; Heubert & Hauser, 1999; Hill & Craft, 2003; Jencks & Phillips, 1998; Kim & Hocevar, 1998; Lee, Autry, Fox, & Williams, 2008; Lee & Burkam, 2002; Penner & Paret, 2007). However, factors contributing to these differences have been linked to socioeconomic status (SES) (Kao & Thompson, 2003; Levine, Vasilyeva, Lourenco, Newcombe, & Huttenlocher, 2005; Riegle-Crumb, 2006; Ryan & Ryan, 2005), participation in mathematics courses (Campbell, 1989; Johnson, 1984; Stiff & Harvey, 1988), bias related to instructional practices (Stiff & Harvey, 1988; Tate, 1994), and attitudinal/psychological factors (Campbell, 1989; Ryan & Ryan, 2005), among many other explanations. Other performance trends have revealed that both White and Asian students achieve higher than Hispanic students (Lockhead, Thorpe, Brooks-Gunn, & McAloon, 1985; Penner & Paret, 2007) but that all three groups tend to outperform African American students (Heubert & Hauser, 1999; Jencks & Phillips, 1998; Lockhead et al., 1985; National Center for Educational Statistics, 2001).

A great portion of the early research on gender differences as it relates to mathematics performance has shown no differences in boys and girls at the elementary level (Armstrong, 1981; Callahan & Clements, 1984; Feingold, 1988; Fennema & Sherman, 1974; Hyde, Fennema, & Lamon, 1990; Siegel, 1968). Other research has indicated slightly better performance for younger girls than boys (Arm-

strong, 1981; Fennema & Carpenter, 1981; Marshall, 1984; Potter & Levy, 1968) while the greatest differences between the sexes have been reported at the middle and high school levels (Backman, 1972; Connor & Serbin, 1985; Hyde et al., 1990; Moore & Smith, 1987).

More recently however, Levine et al. (2005) found that gender differences do exist at an earlier age, with middle and high SES males in the second and third grades outperforming their female counterparts. Research by Riegle-Crumb (2006) supported these claims, but additionally, her research suggests that gender gaps vary across race and SES at the high school as well as at the postsecondary levels. Lee, Moon, and Hegar (2011) also found a gender gap favoring boys as early as the third grade. In general, other researchers have also reported gender differences in more recent years (Lindberg, Hyde, Petersen, & Linn, 2010; Robinson, & Lubienski, 2011), despite claims that gender differences in mathematics have declined (Else-Quest, Linn, & Hyde, 2010; Hyde, 2005; Spelke, 2005).

Considering these previous and sometimes conflicting findings, one major purpose of this study is to investigate how these populations are achieving in learning statistics concepts. Furthermore, using the Data and Chance content domain from the TIMSS (2007) study, student background, teacher and school data can also be considered. The next section will discuss briefly why statistics might be considered a separate discipline from mathematics, and thus, provides clear evidence why research in this area is needed. Detailed and descriptive information about the TIMSS (2007) sampling design will then lead into the methodology used in this study.

Statistical Versus Mathematical Reasoning

Many statistics educators would argue that statistics is a related content domain in mathematics, yet, there is a growing body of research that suggests that statistics should be treated as

a distinct and separate discipline (i.e., from mathematics), which requires a very different approach to teaching and learning (Cobb, 1992; Garfield & Ben-Zvi, 2008; Moore, 1997). According to delMas (2004), statistics is similar to a discipline such as physics that uses mathematics, for example, but the developed methods and concepts are different than that of mathematical inquiry. This is due to the fact that quite often, a statistical investigation is grounded within a context (Cobb & Moore, 1997), it is dependent on data (Chance, 2002), and the use of mathematics comes into play only after much consideration has been given to the research questions under investigation, a design for the collection of data, and the exploration and analysis of the data (delMas, 2004).

Furthermore, delMas (2004) asserts that statistical and mathematical reasoning might appear similar when a student is required to reason with highly abstract concepts and relationships. However, a student is dependent on the characteristics of the context in a statistics scenario in order to accurately develop and select an appropriate model. According to delMas (2004), these cognitive demands require abstract reasoning, which have been shown to result in a variety of conceptual errors in many students. Many of these errors stem from the heuristic and associative processes that were learned in previous mathematics courses, yet lead to erroneous conclusions or interpretations in a real-world statistics scenario (delMas, 2004). In addition, a final step in solving a statistics problem is to make a decision and a final conclusion within the context of the original scenario. This presents another difficulty for statistics learning, as delMas contends that "This translation or mapping represents another potential source of error as multiple relationships must be tracked and validated, and context once again has an opportunity to influence reasoning" (delMas, 2004, p. 91). Thus, in many ways, delMas proclaims that the practice of statistics can be a much more difficult task than the purely mental activity required in mathematical reasoning.

Given the inherent cognitive differences and skills that are necessary for mathematical and statistical reasoning, it makes sense to consider how students are achieving in learning statistics concepts. Specifically, research on how female and minority students are performing is of particular interest in this study, as the literature related to how these students are achieving in statistics is nonexistent.

Purpose

With these broader ideas in mind, many questions might come to mind regarding how minority and female students are progressing in learning statistics concepts. Therefore, this study seeks to describe these relationships, by considering previous variables shown to be related to mathematics and statistics achievement at the student and classroom levels, using the U.S. eighth-grade Data and Chance content domain from the Trends in International Mathematics and Science Study, 2007. In particular, the following research questions were considered:

1. What student and teacher/classroom variables are related to achievement in statistics learning for eighth-grade students in the United States?
2. To what extent do these variables account for the variation in statistics achievement at the student and classroom levels for female and minority students?

METHOD

Data Source: TIMSS/Data and Chance

This study used student achievement, student background, and teacher background data files from the U.S. National Database (public-use data set) at the eighth-grade level. The TIMSS (2007) International Database does not include all of the variables collected in the U.S. National Database. The U.S. National Database conforms to the international specifica-

tions common to the data files for all countries, but it also includes specific adaptations made to the questionnaire items (e.g., such as the race/ethnicity variable added to the student questionnaire). Due to the additional data collected, and because this study focused on the performance of eighth-grade students in the United States, the U.S. National Database was utilized in this study.

Sampling

The U.S. sample of students, which included both private and public schools, were randomly selected and weighted to be representative of students in the nation. In total, 239 eighth-grade schools and 7,377 eighth-grade students participated. Because the TIMSS (2007) study used a two-stage sampling procedure, in which a random sample of schools was selected at the first stage and one or two intact classes (fourth or eighth grade) were selected at the second stage, sampling weights were applied in order to estimate accurate population estimates. Specifically, because students were not selected randomly (i.e., simple random sample), every student did not have an equal chance of being selected. Therefore, sampling weights took this unequal probability into consideration. Sampling weights were also applied due to disproportional sampling of subgroups as well as to address school nonresponse (Foy & Olson, 2009).

Of the 239 eighth-grade schools, two eighth-grade classrooms were selected within each school in an equal probability sample. The mathematics and science teachers of the students sampled for the TIMSS (2007) study were administered at least one questionnaire for each TIMSS class taught. These teachers completed one set of background questions and a separate set of class-specific questions for each TIMSS class they taught. Therefore, a teacher may have taught more than one TIMSS class (and would have completed a separate class-specific questionnaire for each class). In addition, data files were created so that the teachers were linked with their students. For

example, if two teachers were linked to one student, there would be two entries in the data file corresponding to that student. It is also important to point out that the teachers in this study did not constitute a random sample. Instead, they were selected because they taught a representative sample of students in the TIMSS (2007) study (Williams et al., 2009); as a result, it would not be appropriate to generalize any findings to all U.S. teachers.

Assessment, Design, Scoring, and Plausible Values

A TIMSS (2007) assessment booklet, which consisted of both multiple-choice and constructed-response items, was administered to students near the end of the eighth grade in a paper-and-pencil format. A scoring rubric was used to evaluate the accuracy of the constructed-response items. TIMSS (2007) questionnaires were administered similarly to the principals and teachers.

To minimize the testing burden as well as to ensure a broad coverage of the subject matter, a rotated block design of each assessment booklet was utilized. This required that each assessment booklet included both mathematics and science but it was also created so that no student responded to all of the items. As a result, there would be missing data for most items. To accommodate this missing data, the TIMSS (2007) study used the “plausible values” methodology to represent what the true performance of a student might have been had they taken all of the items. This method generates five possible scale scores for each student based on a random selection of scale scores from students with similar backgrounds, who also answered the assessment items in a similar way (Williams et al., 2009). Therefore, each student in this study had five plausible values (i.e., dependent variables) which were taken into consideration in the analyses. Finally, there were 40 data and chance items on the data and chance content domain, which served as the dependent variable in this study (i.e., average of the 40 items).

Variable Selection

The TIMSS (2007) questionnaires contain numerous independent variables available for selection and many of these variables might be related to an outcome of interest. In this study, 76 variables were initially chosen based on: (1) previous research and theory linked to teaching and learning as reported in the mathematics and statistics literature, and (2) the practical implications of this study (i.e., the researcher used sound and reasonable judgment when choosing variables, selecting variables that “make sense,” etc.). Once selected, the intercorrelations among the variables were generated to investigate the relationship among the variables as well as to reduce the variables even further, with the criterion that variables statistically related to the dependent variable were chosen for further analyses. Recall that the average statistics achievement (on the 40 data and chance items) served as the dependent variable. This resulted in 63 variables to consider for the second phase of the variable selection process (see Appendix 2 for a complete description of these variables).

In addition, this study utilized a two-phase variable selection process with regard to selecting variables that might be related to a student's achievement in statistics. The first phase identified variables based on theory, previous research, and the statistical relationship between a student's average achievement score on the data and chance content domain and a number of TIMSS independent or predictor variables. The second phase involved selecting a “good” smaller subset of variables that emerged as a result of determining the minimum number and nature of underlying hypothetical common factors. Therefore, an exploratory factor analysis was conducted using a principal axis factoring method. Two statistical criteria were used to determine the number of extracted factors: (1) eigenvalue-greater-than-one rule and (2) an evaluation of the scree plot. Because the factors were assumed to be uncorrelated, the Varimax orthogonal rotation method was used. Approx-

imately $n = 7,189$ valid cases considered a 2 to 8 factor solution, which explained from 21 to 67% of the total variance. A loading of at least .50 was used to define a factor (Crocker & Algina, 1986, p. 299) with items that did not meet this criteria being eliminated. When items were deleted, the analysis was conducted again as suggested by Benson and Nasser (1998), with the goal of sharpening the factor pattern as well as to achieve simple structure (i.e., variables primarily loading high on only one factor). A four-factor solution, which explained 65.2% of the variance was chosen based on the total amount of variance explained, theory, and interpretability of the loadings. This resulted in 15 student and teacher variables/items that were selected to be included in the final hierarchical linear modeling model. Table 1 presents the descriptive statistics for these variables. (For a complete discussion of the two-stage variable selection process, see Mills & Holloway, 2013).

Table 2 presents other student, teacher, and school background variables which were retained due to their statistical or marginal significance, their relationship to mathematics and statistics learning in previous research, and their practical implications in this study. These variables for the student included: gender, number of books in the home, parent's education level, whether or not there was a computer, or Internet access in the home, percentage of economically disadvantaged students in the school, and percentage of students on free lunch. Teacher variables included gender, number of years teaching, level of formal education, area of study (i.e., mathematics education, mathematics, or neither), whether the teacher was certified or not, and whether or not the teacher had participated in professional development or mathematics pedagogy within the last 2 years.

The sample consisted of almost an equal number of boys (47.7%) and girls (48.7%) (Note that when percentages did not total to 100, observations were missing, omitted, or not administered). Of these students, almost 41 percent reported that their parents' highest

TABLE 1
Descriptive Statistics for Statistics Achievement-Related Variables

| <i>TIMSS Variable (Students)</i> | <i>**Mean (SD)</i> | <i>Percent</i> | <i>n*</i> |
|---|--------------------|----------------|-----------|
| I would like to take more math | 2.62 (1.24) | | |
| Agree a lot | | 18.3 | 1,392 |
| Agree a little | | 29.6 | 2,244 |
| Disagree a little | | 26.4 | 2,001 |
| Disagree a lot | | 21.5 | 1,635 |
| <i>TIMSS Variable (Students)</i> | | <i>Percent</i> | <i>n*</i> |
| I enjoy learning math | 2.49 (1.31) | | |
| Agree a lot | | 20.4 | 1,551 |
| Agree a little | | 36.5 | 2,773 |
| Disagree a little | | 20.9 | 1,586 |
| Disagree a lot | | 17.5 | 1,326 |
| I like math | 2.45 (1.25) | | |
| Agree a lot | | 22.8 | 1,729 |
| Agree a little | | 33.7 | 2,558 |
| Disagree a little | | 20.7 | 1,569 |
| Disagree a lot | | 18.7 | 1,423 |
| Math is boring | 2.51 (1.30) | | |
| Agree a lot | | 23.4 | 1,777 |
| Agree a little | | 28.5 | 2,163 |
| Disagree a little | | 25.2 | 1,913 |
| Disagree a lot | | 18.5 | 1,403 |
| * <i>n</i> = 7,377. When sample sizes do not total 7,377, observations are missing, omitted, or not administered. | | | |
| **1 = agree a lot, 2 = agree a little, 3 = disagree a little, 4 = disagree a lot | | | |
| <i>TIMSS Variable (Teachers)</i> | | <i>Percent</i> | <i>n*</i> |
| How well-prepared are you able to teach: | | | |
| Reading and displaying data | | | |
| Not applicable | | .2 | 1 |
| Very well | | 83.2 | 341 |
| Somewhat | | 3.4 | 14 |
| Not well | | 1.2 | 5 |
| Interpreting data sets | | | |
| Not applicable | | .2 | 1 |
| Very well | | 79.5 | 326 |
| Somewhat | | 6.8 | 28 |
| Not well | | 1.2 | 5 |
| <i>TIMSS Variable (Teachers)</i> | | <i>Percent</i> | <i>n*</i> |
| Judging chances | | | |
| Not applicable | | .2 | 12 |
| Very well | | 76.8 | 315 |
| Somewhat | | 9.0 | 37 |
| Not well | | 1.7 | 7 |

(Table continues on next page)

TABLE 1
(Continued)

| <i>TIMSS Variable (Teachers)</i> | <i>Percent</i> | <i>n*</i> |
|---|----------------|-----------|
| When did you first teach: | | |
| Interpreting data | | |
| Mostly taught before this year | 32.7 | 134 |
| Mostly taught this year | 45.1 | 185 |
| Not yet taught or just introduced | 12.2 | 50 |
| Data displays | | |
| Mostly taught before this year | 31.7 | 130 |
| Mostly taught this year | 34.4 | 141 |
| Not yet taught or just introduced | 23.9 | 98 |
| Using data to predict chance | | |
| Mostly taught before this year | 27.8 | 114 |
| Mostly taught this year | 34.4 | 141 |
| Not yet taught or just introduced | 27.8 | 114 |
| Using chance to solve problems | | |
| Mostly taught before this Year | 26.1 | 107 |
| Mostly taught this year | 32.9 | 135 |
| Not yet taught or just introduced | 31.0 | 127 |
| How would you characterize: | | |
| Parental support for achievement | | |
| Very high | 6.3 | 26 |
| High | 20.0 | 82 |
| Medium | 36.1 | 148 |
| Low | 19.3 | 79 |
| Very low | 8.5 | 35 |
| Parental involvement in school activities | | |
| Very high | 6.6 | 27 |
| High | 14.4 | 59 |
| Medium | 36.6 | 150 |
| Low | 23.2 | 95 |
| Very low | 9.5 | 39 |
| Students' regard for school property | | |
| Very high | 2.0 | 8 |
| High | 18.8 | 77 |
| Medium | 40.7 | 167 |
| Low | 21.5 | 88 |
| Very Low | 7.3 | 30 |
| <i>TIMSS Variable (Teachers)</i> | <i>Percent</i> | <i>n*</i> |
| Students' desire to do well | | |
| Very high | 2.9 | 12 |
| High | 20.5 | 84 |
| Medium | 44.6 | 183 |
| Low | 17.3 | 71 |
| Very low | 4.9 | 20 |

**n* = 410. When sample sizes do not total 410, observations are missing, omitted, or not administered.

TABLE 2
Descriptive Statistics for Participating Students and Teachers

| <i>TIMSS Variable</i> | <i>Percent</i> | <i>n*</i> |
|---|----------------|-----------|
| Students | | |
| Gender | | |
| Boys | 47.7 | 3,622 |
| Girls | 48.7 | 3,697 |
| <i>TIMSS Variable</i> | <i>Percent</i> | <i>n*</i> |
| Parents' highest educational level | | |
| University degree | 40.8 | 3,095 |
| Completed postsecondary but not university | 7.0 | 530 |
| Completed upper secondary | 20.3 | 1,538 |
| Completed lower secondary | 7.6 | 574 |
| Less than lower secondary | 2.3 | 175 |
| Do not know | 17.9 | 1,362 |
| Computer in home? | | |
| Yes | 92.9 | 7,055 |
| No | 5.9 | 450 |
| Access to Internet? | | |
| Yes | 82.9 | 6,293 |
| No | 13.0 | 988 |
| Number of books in the home | | |
| None or very few (0 to 10) | 16.9 | 1,281 |
| One shelf (11 to 25) | 19.8 | 1,500 |
| One bookcase (26 to 100) | 27.1 | 2,059 |
| Two bookcases (101 to 200) | 15.5 | 1,178 |
| Three or more bookcases (Over 200) | 16.4 | 1,243 |
| Percentage of economically disadvantaged in school | | |
| 0 to 10 | 13.6 | 1,033 |
| 11 to 25 | 18.4 | 1,400 |
| 26 to 50 | 23.2 | 1,760 |
| >50 | 32.4 | 2,459 |
| Percentage on free lunch | | |
| <10% | 9.9 | 749 |
| 10-24.9% | 18.4 | 1,399 |
| 25-49.9% | 25.2 | 1,916 |
| 50-74.9% | 21.5 | 1,632 |
| >=75% | 19.1 | 1,447 |
| * <i>n</i> = 7,377. When sample sizes do not total 7,377, observations are missing, omitted, or not administered. | | |
| Teachers | | |
| Gender | | |
| Male | 29.5 | 121 |
| Female | 61.5 | 252 |
| Years taught | | |
| ≤ 10 | 42.3 | 174 |
| 11-15 | 17.5 | 72 |
| 16-20 | 7.7 | 32 |
| >20 | 20.9 | 86 |

(Table continues on next page)

TABLE 2
(Continued)

| <i>TIMSS Variable</i> | <i>Percent</i> | <i>n*</i> |
|----------------------------|----------------|-----------|
| Level of formal education | | |
| Associate or bachelors | 38.3 | 157 |
| Masters or higher | 52.4 | 215 |
| Major area/mathematics? | | |
| Yes | 37.8 | 155 |
| No | 51.7 | 212 |
| Major area/mathematics ed? | | |
| Yes | 43.2 | 177 |
| No | 46.6 | 191 |
| Teaching certificate? | | |
| Yes | 87.1 | 357 |
| No | 3.4 | 14 |
| Professional development? | | |
| Yes | 71.2 | 292 |
| No | 16.6 | 68 |
| Mathematics pedagogy? | | |
| Yes | 66.3 | 272 |
| No | 21.1 | 87 |

**n* = 410. When sample sizes do not total 410, observations are missing, omitted, or not administered.

level of education was a university degree (40.8%), while 37.2% reported lower educational levels (17.9% reported they did not know). Approximately 82.9% of the students reported that they had Internet access in their home, 92.9% reported having a computer. However, there appeared to be a larger variation of percentages in terms of the number of books in their homes: 27.1% revealed they had about one bookcase, 19.8% reported one shelf of books, and 16.9% reported none or very few books. The principals reported that 32.4% of the participating schools had 50% or more economically disadvantaged students. They also reported that about 40% of the students in the participating schools were on free lunch.

Most of the TIMSS teachers in this study were female (61.5%); 29.5% were male. The majority of these teachers (42.3%) reported that they have been teaching for 10 years or less, 17.5% reported between 11–15 years, 7.7% reported 16–20, and 20.9% indicated 21 or more years of teaching experience. Only 38.3% of the teachers reported finishing a first

degree (i.e., associate or bachelor) while the majority (52.4%) indicated that they finished a second degree or higher (i.e., masters, PhD or EdD or law). While 87.1% of the teachers reported that they do have a teaching certificate, 46.6% reported that their major area of study was not mathematics education (51.7% reported that their major area of study was also not general mathematics). The majority of these teachers however, reported that in the last 2 years, they have participated in professional development in mathematics content (71.2%) and mathematics pedagogy/instruction (66.3%), respectively.

One question addressed how well-prepared teachers felt there were to teach certain data and chance topics (see Table 1). In general, approximately 83.2% of the teachers reported that they felt “very well prepared” to teach the topics related to reading and displaying data using tables (i.e., pictographs, bar graphs, pie charts and line graphs), 79.5% felt “very well prepared” to teach topics related to interpreting data sets (i.e., drawing conclusions, making

predictions, and estimating values between and beyond given data points), and 76.8% felt “well prepared” to teach the topics related to judging, predicting, and determining the chances of possible outcomes.

Table 3 presents the average statistics achievement on the Data and Chance content domain by race and gender. There were 40 items which measured students’ ability in knowing, applying, and reasoning within the topic areas of data organization and representation, data interpretation and chance. Sample item tasks included: true/false statements, probability of drawing a blue bead, number of tickets sold, making a bar chart, completing and labeling a pie chart, mean and median number of staff members, and how likely it will rain. A full description of a sample of these items can be found at TIMSS (2007).

Hierarchical Linear Modeling

A two-level hierarchical linear model was estimated due to the nesting and dependency of the data (Raudenbush & Bryk, 2002). The level-1 and level-2 equations, respectively are:

$$Y_{ij} = \beta_{oj} + r_{ij}$$

$$\beta_{oj} = \gamma_{00} + \mu_{oj}$$

where Y_{ij} is the statistics score of student i in classroom j , β_{oj} is the average statistics achievement for classroom j , γ_{00} is the overall average statistics achievement over all of the classrooms, μ_{oj} is the random effect for classroom j , and r_{ij} is the random effect of student i in classroom j .

When building a hierarchical linear model, the first step usually involves estimating an unconditional (null) model, which is the equivalent to a one-way ANOVA with random effects. The intraclass correlation coefficient for this model revealed that 27.4% of the variation in statistics achievement could be explained at the classroom level. This left a considerable amount of variation that might be explained at the student level. The next steps involved adding previous variables shown to be related to mathematics and statistics achievement at the student and classroom levels, to further examine the extent to which these factors contribute to statistics learning for minority and female middle school students in the United States today. The variables discussed in Tables 1 and 2 above were selected as potential candidates for the final model.

A plausible values analysis was utilized in the hierarchical linear model so that all 5 scores (i.e., dependent variables) for every stu-

TABLE 3
Average Statistics Achievement by Race and Gender*

| <i>TIMSS Variable</i> | <i>Mean</i> | <i>SD</i> | <i>n**</i> |
|-----------------------|-------------|-----------|------------|
| Gender | | | |
| Male | 532.9 | 107.3 | 3,622 |
| Female | 523.4 | 103.0 | 3,806 |
| Race | | | |
| White | 565.4 | 93.9 | 3,873 |
| Black | 465.3 | 94.5 | 949 |
| Hispanic | 479.2 | 99.5 | 1,787 |
| Asian | 563.7 | 97.0 | 243 |
| Native American | 515.5 | 93.8 | 90 |
| Pacific Islander | 527.4 | 90.2 | 58 |
| Two or more races | 518.8 | 85.7 | 282 |

*TIMSS scale average: 500. See Appendix 2 for interpretation of scale scores and benchmarks.

** $n = 7,737$. When sample sizes do not total 7,737, observations are missing or omitted.

dent were used in the analysis. Variables that were statistically significant remained in the model. The appropriate weights were also used in the analysis at both levels in order to take into consideration the complex sampling design. To improve the interpretation of the results, all of the predictor variables, except gender, were grand-mean centered. Finally, an examination of the missing data was conducted at both levels for this study, as missing data can be a potential problem in any large-scale study. The listwise deletion method was used to eliminate any missing data at both levels; therefore, parameter estimates were only computed for complete data.

RESULTS

The parameter estimates for the final model are presented in Table 4. Recall that all other

student and teacher/classroom variables that were not statistically significant, including cross-level interaction terms were removed from the model.

The first finding, which has been reported in other mathematics education studies, was also confirmed in this study. That is, there was a positive and significant relationship between the number of books found in a student's home and their statistics achievement. The predicted statistics achievement increased by 11 points for each higher category of books reported, controlling for all other variables in the model. This result supports previous research reported in the literature (Koretz et al. 2001; Martin et al. 2000; Mills & Holloway, 2013; O'Dwyer, 2005; Phan, Sentovich, Kromey, Dedrick, & Ferron, 2010). Similarly, an even stronger relationship was obtained for access to the Internet—students who had access to the Internet scored 15 points higher than those who did

TABLE 4
HLM Final Model Results

| <i>Fixed Effect</i> | <i>Coefficient*</i> | <i>SE</i> | <i>p Value</i> |
|---------------------------|---------------------|-----------|----------------|
| Intercept | 525.73 | 5.14 | <.001 |
| Percentage/free lunch | -20.66 | 2.96 | <.001 |
| I like math | -13.74 | 1.41 | <.001 |
| Gender of student | 12.95 | 2.53 | <.001 |
| Number of books | 11.01 | 1.28 | <.001 |
| Parent ed level | 4.05 | .96 | .001 |
| Percentage/free lunch | .11 | .02 | <.001 |
| Internet access | 15.31 | 3.99 | <.001 |
| Black | -26.81 | 7.10 | <.001 |
| Hispanic | -16.09 | 7.32 | .032 |
| Asian | 26.63 | 9.97 | .008 |
| Level of formal education | 12.94 | 6.33 | .042 |
| Parents' highest ed level | 3.84 | .94 | <.000 |
| <i>Random Effect</i> | <i>Variance</i> | <i>df</i> | <i>p Value</i> |
| Intercept | 1,078.38 | 34 | <.000 |
| Book | 31.94 | 35 | .026 |
| Black | 595.79 | 35 | .015 |
| Hispanic | 571.68 | 35 | .039 |
| Asian | 954.23 | 35 | .011 |
| σ^2 | 6,330.55 | | |

*Unstandardized regression coefficient.

not have access. The relationship between Internet access and statistics achievement has also been confirmed in a previous study by Mills and Holloway (2013).

Another finding revealed that for students who disagreed that they like math, their statistics achievement was 13.7 points higher than those who agreed (that they disliked math)—not a surprising outcome and consistent with previous mathematics achievement findings (Caponera & Russo, 2010; Koretz et al. 2001; Mills & Holloway, 2013). There was also a positive and statistically significant relationship between achievement in statistics and a parent's educational level, another finding confirmed in the literature (Martin et al. 2000; Mills & Holloway, 2013). Specifically, this finding revealed that statistics achievement will likely improve by more than 4 points as the parent's education (category) level increases, controlling for the other variables included in this model. Other studies have also shown a similar relationship for variables related to parental education level (i.e., SES, percentage of economically disadvantaged students, free/reduced lunch, etc.) (Koretz et al. 2001; Martin et al. 2000; Mills & Holloway, 2013; O'Dwyer, 2005; Phan et al. 2010).

Findings related to the teachers were interesting yet worthy of further discussion. One finding revealed that the more education a teacher had obtained, the better students performed in statistics. Presumably related to this finding, almost 53% of these teachers reported earning an advanced degree; however, recall that these degrees were in fields other than mathematics. The next finding might also be related—although the majority of teachers felt they were prepared to teach statistics concepts, almost half of them reported that they do not have a degree in either mathematics or mathematics education. Therefore, it is possible that many of the TIMSS teachers may not have had any or very little (i.e., inadequate) formalized training in statistics. Previous studies have also reported similar findings; that is, teachers who teach statistics in the lower grades either have never even taken a statistics course or have not

been properly trained (Begg & Edwards, 1999; Bryce, 2005; Franklin, 2000; Sotos, Vanhoof, Van den Noortgate, & Onghena, 2007).

Furthermore, these findings appear to reflect some of the challenges experienced by mathematics educators. According to Ingersoll and Perda (2009), there is a widespread shortage of mathematics and science teachers not only in the United States but internationally as well. Teacher job dissatisfaction as well as teachers choosing to seek other career paths has also been related to both the number and quality of available elementary and secondary mathematics classroom teachers (Ingersoll & Perda, 2009). As defined by Ingersoll and Perda (2009), a teacher is "qualified" if he or she holds an undergraduate or graduate degree in that or a related field (e.g., mathematics, mathematics education, or statistics). Clearly in this study, most of the teachers did not appear to meet this criterion.

One cautionary note—these results should be considered based on the fact that the teachers in this study did not constitute a random sample. Instead, they were selected because they taught a representative sample of students in the TIMSS study (Williams et al., 2009). Therefore, although it would not be appropriate to generalize these results to all U.S. teachers, the size of the sample alone will allow some revealing yet informative conclusions to be drawn about demographics and instructional practices of teachers teaching statistics concepts in the middle school in the United States.

In terms of exploring what relationships might exist for minority and female students and their statistics achievement, the findings revealed statistically significant but important differences. First, at the mean for all of the other predictors in the model, the predicted statistics achievement for girls was 525.94; for boys, it was approximately 13 points higher. As mentioned in the previous literature earlier in this article, findings about the differences in mathematics performance between boys and girls have long been reported (Armstrong, 1981; Callahan & Clements, 1984; Feingold,

1988; Hyde et al., 1990; Siegel, 1968). Our findings confirmed that differences do exist between the two groups on the TIMSS-related data and chance content domain items. Other findings revealed that minority students in general performed lower statistically than White students overall. Furthermore, compared to White students, African American students scored more than 26 points lower while Hispanic students scored 16 points lower than did their White counterparts. Other studies in the mathematics education literature have also shown these performance trends (Cambell, 1989; Heubert & Hauser, 1999; Jencks & Phillips, 1998; Lockhead et al., 1985; NCES, 2001; Penner & Paret, 2007).

Factors such as motivation and/or other psychological/attitudinal factors, varying instructional practices and biases, lack of opportunities or exposure to mathematics, and socioeconomic factors have been used to explain these differences (Heubert & Hauser, 1999; Jencks & Phillips, 1998; Lee et al., 2011; NCES, 2001; Penner & Paret, 2007; Riegle-Crumb, 2006; Ryan & Ryan, 2005). It is difficult to know if these factors are still contributing to lower academic performances for students learning statistics concepts, but our guess would be that they are. What then, are some strategies to combat these challenges? Well, technology has changed the way we teach and learn forever, and it will likely play an even larger role in education as many K–12 institutions, colleges and universities utilize more of it to achieve their own teaching and learning objectives. Thus, the use of stimulating, audience-appropriate, and engaging course materials might offer one remedy to economically disadvantaged students in particular—those who might not have access to good course content, or teachers who are trained and willing to teach statistics (or mathematics) concepts and ideas in their schools. Especially for a course like statistics, actively engaging students with real-world data and using technology to enhance conceptual understanding are critical instructional practices supported by previous research (American Sta-

tistical Association, 2007; Garfield, 1993; Garfield & Ben-Zvi, 2008; Harrington, 1999; NCTM, 1989, 1991, 2000; Ward, 2004). Statistics content or even entire courses can be accessed online—even the poorest schools in the United States manage to have access to the Internet. It is not known the extent to which the teachers in this study utilized technology and hands-on activities in their statistics teaching. However, as technology begins to change the landscape of education, it will become more important to conduct the research needed to answer what practices are best for optimal learning not only in statistics, but in every field and level of education.

CONCLUDING REMARKS

There are limitations and alternative explanations that may help to explain the conclusions and findings in our study. First, although the NCTM Data Analysis and Probability standard was implemented in PreK–12 education in many states in the United States in 1989, these standards were recently revised as early as 2000. In addition, it is also not known whether the participating schools in this study adopted this standard. Furthermore, a great deal of work in statistics education in the lower grades began to increasingly emerge in the mid-2000s (i.e., American Statistical Association, 2007; College Board, 2006). Therefore, the results from the TIMSS (2007) study might not be an adequate measure of the most recent progress made in statistics teaching and learning. Nevertheless, we trust that the results and findings are indeed useful today, and offer the type of information we as statistics educators desire in order to monitor the teaching and learning of statistics in the upcoming years.

Second, the complex sampling design from this study required a sample of only two classrooms from each school. For the TIMSS (2003) study, low-income schools were oversampled for the fourth grade but the TIMSS (2007) study reported no oversampling for either grade in 2007. Based our findings

however, the demographic statistics revealed that the majority of participating students were in fact economically disadvantaged (as reported by the principals). As a result, it is questionable whether or not this sample of students is truly representative of the average United States eighth-grade minority and female student's achievement in statistics. Yet again, this data is also valuable, as information about how students at lower SES levels are performing in statistics is an area that is also lacking in the literature. There appears to be an abundance of research of lower SES student performance in mathematics but additional research about their performance in statistics is also informative and useful.

Finally, two issues are related to our hierarchical level modeling approach. First, we specified students at level-1 for the model while we defined the teacher and any other classroom variables at level-2. We also included one school-level variable, as reported by the principals (i.e., percentage of free/reduced lunch), which we "treated" as a classroom variable. Therefore, this variable served as an estimate of the SES composition at the classroom level (An alternative way to describe this data would be to consider a three-level model, with students at level-1, teacher/classroom at level-2, and school variables at level-3). Second, misspecification of the final hierarchical linear model when important variables are omitted can lead to bias when estimating the level 2 predictors of the intercept (Raudenbush & Bryk, 2002, p. 259). Related to choosing a final model, finding the "best subset" has long been a topic of discussion in the educational research literature, as researchers often desire to reduce a large number of variables to a smaller set, whose goal is to explain almost as much variance as the total set. The recommendation of using sound theory, previous research, and statistics (Mills, Olejnik, & Marcoulides, 2005) was considered in this study. It is quite likely that a different "best subset" may emerge depending on the criteria and methods selected by the researchers.

In conclusion, this study revealed significant differences in the statistics achievement between boys and girls as well as minority students versus their White counterparts using data collected from the TIMSS (2007) U.S. National Database. Overall, students tend to struggle to understand statistics concepts in general and a "shift" in the way we teach and learn statistics has evolved into a movement to improve all students' statistical literacy in every level of education (Cobb, 1992; Moore, 1997). This has resulted in much more of a focus on teacher training as well as addressing changes related to content, pedagogy, and the use of technology in teaching and learning. However, research on how to improve the teaching and learning obstacles for female and minority students seem to also need further study. If statistical literacy is a goal for our students, future research in PreK-12 statistics education will be important in order to better monitor and evaluate progress in this field, for all students.

REFERENCES

- American Statistical Association. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report*. Retrieved from <http://www.amstat.org/education/gaise/>
- Armstrong, J. M. (1981). Achievement and participation of women in mathematics: Results of two national surveys. *Journal for Research in Mathematics Education*, 12, 356-372.
- Backman, M. E. (1972). Patterns of mental abilities: Ethnic, socioeconomic, and sex differences. *American Educational Research Journal*, 9, 1-12.
- Begg, A., & Edwards, R. (1999). *Teachers' ideas about teaching statistics*. Paper presented at the combined annual meeting of the Australian Association for Research in Education and the New Zealand Association for Research in Education, Melbourne, Australia.
- Benson, J., & Nasser, F. (1998). On the use of factor analysis as a research tool. *Journal of Vocational Education Research*, 23, 13-33.
- Ben-Zvi, D. (2004). Reasoning about data analysis. In D. Ben-Zvi & J. B. Garfield (Eds.), *The chal-*

- lence of developing statistical literacy, reasoning, and thinking (pp. 121–146). Dordrecht, Netherlands: Kluwer.
- Biehler, R. (1997). Students' difficulties in practicing computer-supported data analysis: Some hypothetical generalizations from results of two exploratory studies. In J. B. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics: Proceedings of the 1996 IASER round table conference* (pp. 169–190). Voorburg, The Netherlands: International Statistical Institute.
- Bryce, G. R. (2005). Developing tomorrow's statisticians. *Journal of Statistics Education*, 13(1). Retrieved from <http://www.amstat.org/publications/jse/v13n1/bryce.html>
- Caponera, E., & Russo, P. M. (2010). Performance in TIMSS mathematics test: The influence of self-concept and socio-economic background. *CADMO*, 18(2), 91–105.
- Callahan, L. G., & Clements, D. H. (1984). Sex differences in rote-counting ability on entry to first grade: Some observations. *Journal of Research in Mathematics Education*, 15, 378–382.
- Campbell, P. B. (1989). So what do we do with the poor, non-White female? Issues of gender, race, and social class in mathematics and equity. *Peabody Journal of Education*, 66(2), 95–112.
- Chance, B. L. (2002). Components of statistical thinking and implications for instruction and assessment. *Journal of Statistics Education* 10(3). Retrieved from <http://www.amstat.org/publications/jse/>
- Chen, Y., Ferron, J. M., Thompson, M. S., Gorin, J. S., & Tatsuoka, K. K. (2010). Group comparisons of mathematics performance from a cognitive diagnostic perspective. *Educational Researcher and Evaluation*, 16(4), 325–343.
- Cobb, G. W. (1992). Teaching statistics. In L. Steen (Ed.), *Heeding the call for change: Suggestions for curricular action* (pp. 3–33). Washington, DC: Mathematical Association of America.
- Cobb, G. W., & Moore, D. (1997). Mathematics, statistics, and teaching. *American Mathematical Monthly*, 104, 801–823.
- College Board. (2006). *College Board standards for college success: Mathematics and statistics*. New York, NY: Author.
- Conference Board of the Mathematical Sciences. (2012). *The mathematical education of teachers II*. Washington, DC: American Mathematical Society and Mathematical Association of America.
- Connor, J. M., & Serbin, L. A. (1985). Visual-spatial skill: Is it important for mathematics? Can it be taught? In S. F. Chipman, L. R. Brush, & D. M. Wilson (Eds.), *Women and mathematics: Balancing the equation* (pp. 151–174). Hillsdale, NJ: Erlbaum.
- Cooper, J. A., & Schleser, R. (2006). Closing the achievement gap: Examining the role of cognitive developmental level in academic achievement. *Early Childhood Education Journal*, 33(5), 301–306.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Fort Worth, TX: Harcourt.
- delMas, R. C. (2004). A comparison of mathematical and statistical reasoning. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 79–97). The Netherlands: Kluwer.
- delMas, R., Garfield, J., & Chance, B. (1999). A model of classroom research in action: Developing simulation activities to improve students' statistical reasoning. *Journal of Statistics Education*, 7(3). Retrieved from <http://www.amstat.org/publications/jse/secure/v7n3/delmas.cfm>
- delMas, R., & Liu, Y. (2005). Exploring students' conceptions of the standard deviation. *Statistics Education Research Journal*, 2(1), 55–82.
- Dempster, M., & McCorry, N.K. (2009). The role of previous experience and attitudes toward statistics in statistics assessment outcomes among undergraduate psychology students. *Journal of Statistics Education*, 17(2). Retrieved from <http://www.amstat.org/publications/jse/v17n2/dempster.html>
- Education Trust. (2003). African American achievement in America. Retrieved from <http://www.edtrust.org/dc/publication/african-american-achievement-in-america>
- Else-Quest, N. M., Linn, M. C., & Hyde, J. S. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, 136(1), 103–127.
- Estrada, A., & Batanero C. (2008). Explaining teachers' attitudes toward statistics. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE Study: Teaching Statistics in School Mathematics: Challenges for Teaching and Teacher Education*. Retrieved from http://www.stat.auckland.ac.nz/~iase/publications/rt08/T2P4_Estrada.pdf

- Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist*, *43*, 95–103.
- Fennema, E., & Carpenter, T. P. (1981). Sex-related differences in mathematics: Results from the national assessment. *Mathematics Teacher*, *74*, 554–559.
- Fennema, E., & Sherman, J. (1977). Sex-related differences in mathematics achievement, spatial visualization, and sociocultural factors. *American Educational Research Journal*, *14*(1), 51–71.
- Foy, P., & Olson, J. F. (Eds.). (2009). *TIMSS 2007 user guide for the international database*. Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Boston College.
- Franklin, C. (2000). Are our teachers prepared to provide instruction in statistics at the K–12 levels? *Dialogues*, *10*. Retrieved from <http://www.nctm.org/resources/content.aspx?id=1776>
- Gal, I. (2002). Adults' statistical literacy: Meanings, components, and responsibilities. *International Statistical Review*, *70*(1), 1–25.
- Gal, I., & Ginsburg, L. (1994). The role of beliefs and attitudes in learning statistics: Towards an assessment framework. *Journal of Statistics Education*, *2*(2). Retrieved from <http://www.amstat.org/publications/jse/v2n2/gal.html>
- Garfield, J. (2003). Assessing statistical reasoning. *Statistics Education Research Journal*, *2*(1), 22–38.
- Garfield, J., & Ben-Zvi, D. (2008). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review*, *75*(3), 372–396.
- Garfield, J. B. (1993). Teaching statistics using small-group cooperative learning. *Journal of Statistics Education*, *1*, 1. Retrieved from <http://www.amstat.org/publications/jse/v1n1/garfield.html>
- Garfield, J. B., & Everson, M. (2009). Preparing teachers of statistics: A graduate course for future teachers. *Journal of Statistics Education*, *17*(2). Retrieved from <http://www.amstat.org/publications/jse/v17n2/garfield.html>
- Gonul, F., & Solano, R. A. (2013). Innovative teaching: An empirical study of computer-aided instruction in quantitative business courses. *Journal of Statistics Education*, *21*(1). Retrieved from <http://www.amstat.org/publications/jse/v21n1/gonul.html>
- Hammerman, J. K., & Rubin, A. (2004). Strategies for managing statistical complexity with new software tools. *Statistics Education Research Journal*, *3*(2), 17–41.
- Harrington, D. (1999). Teaching statistics: A comparison of traditional classroom and programmed instruction/distance learning approaches. *Journal of Social Work*, *35*, 343–352.
- Heubert, J. P., & Hauser, R. M. (Eds.). (1999). *High stakes: Testing for tracking, promotion, and graduation*. Washington, DC: National Academy Press.
- Hill, N. E., & Craft, S. A. (2003). Parent-school involvement and school performance: Mediated pathways among socioeconomically comparable African American and Euro-American families. *Journal of Educational Psychology*, *95*(1), 74–83.
- Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, *60*, 581–592.
- Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, *107*, 139–155.
- Ingersoll, R. M., & Perda, D. (2009). The mathematics and science teacher shortage: Fact and myth. *Consortium for Policy Research in Education* (Research Report No. RR-6). Philadelphia, PA: University of Pennsylvania.
- Jencks, C., & Phillips, M. (Eds.). (1998). *The Black-White test score gap*. Washington, DC: Brookings Institution Press.
- Johnson, M. L. (1984). Blacks in mathematics: A status report. *Journal for Research in Mathematics Education*, *15*(2), 145–153.
- Johnson, M., & Kuennen, E. (2006). Basic math skills and performance in an introductory statistics course. *Journal of Statistics Education*, *14*(2). Retrieved from <http://www.amstat.org/publications/jse/v14n2/johnson.html>
- Kao, G., & Thompson, J. S. (2003). Racial and ethnic stratification in educational achievement and attainment. *Annual Review of Sociology*, *29*, 417–441.
- Kien-Kheng, F., & Idris, N. (2010). A comparative study on statistics competency level using TIMSS data: Are we doing enough? *Journal of Mathematics Education*, *3*(2), 126–138.
- Kim, S., & Hocevar, D. (1998). Racial differences in eighth-grade mathematics: Achievement and opportunity. *Clearing House*, *71*(3), 75–79.

- Konold, C. (1995). Issues in assessing conceptual understanding in probability and statistics. *Journal of Statistics Education*, 3(1). Retrieved from <http://www.amstat.org/publications/jse/v3n1/konold.html>
- Koretz, D., McCaffrey, D., & Sullivan, T. (2001). *Using TIMSS to analyze correlates of performance variation in mathematics*. Retrieved from <http://nces.ed.gov/pubs2001/200105.pdf>
- Lee, J., Autry, M. M., & Williams, C. (2008). Investigating children's mathematics readiness. *Journal of Research in Childhood Education*, 22(3), 316–328.
- Lee, V., & Burkam, D. (2002). *Inequality at the starting gate: Social background in achievement as children begin school*. Washington, DC: Economic Policy Institute.
- Lee, J., Moon, S., & Hegar, R. L. (2011). Mathematics skills in early childhood: exploring gender and ethnic patterns. *Child Ind Res*, 4, 353–368.
- Levine, S. C., Vasilyeva, M., Lourenco, S. F., Newcombe, N. S., & Huttenlocher, J. (2005). Socioeconomic status modifies the sex difference in spatial skill. *Psychological Science*, 16, 841–845.
- Lindberg, S. M., Hyde, J., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin*, 136(6), 1123–1135.
- Lockhead, M., Thorpe, M., Brooks-Gunn, J., Casserly, P., & McAloon, A. (1985). *Gender and ethnic differences in middle school mathematics, science, and computer science. What do we know?* Princeton, NJ: Educational Testing Service.
- Lovett, M., & Greenhouse, J. (2000). Applying cognitive theory to statistics education. *The American Statistician*, 54(3), 196–206.
- Lunsford, M. L., & Poplin, P. (2011). From research to practice: Basic mathematics skills and success in introductory statistics. *Journal of Statistics Education*, 19(1). Retrieved from <http://www.amstat.org/publications/jse/v19n1/lunsford.html>
- Marshall, S. P. (1984). Sex differences in children's mathematics achievement: Solving computations and story problems. *Journal of Educational Psychology*, 76, 194–204.
- Martin, M. O., Mullins, I. V. S., Gregory, K. D., Hoyle, C., & Shen, C. (2000). *Effective schools in science and mathematics: IEA's third international mathematics and science study*. Chestnut Hill, MA: TIMSS International Study Center Boston College.
- Mills, J. D., & Holloway, C. E. (2013). The development of statistical literacy skills in the 8th grade: Exploring the TIMSS data to evaluate student learning and teacher factors in the U.S. *Educational Research and Evaluation: An International Journal on Theory and Practice*, 19(4), 323–345.
- Mills, J. D., Olejnik, S. F., & Marcoulides, G. A. (2005). The TABU search procedure: An alternative to the variable selection methods. *Multivariate Behavioral Research*, 40(3), 351–371.
- Mills, J. D., & Raju, D. (2011). Teaching statistics online: A decade's review of the literature about what works. *Journal of Statistics Education*, 19(2). Retrieved from <http://www.amstat.org/publications/jse/v19n2/mills.html>
- Min-Hsiung, H. (2009). Classroom homogeneity and the distribution of student math performance. *Social Science Research*, 38(4), 781–791.
- Moore, D. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review*, 65, 123–137.
- Moore, E. G., & Smith, A. W. (1987). Sex and ethnic group differences in mathematics achievement: results from the national longitudinal study. *Journal for Research in Mathematics Education*, 18, 25–36.
- Mullis, I. V. S., Martin, M. O., & Foy, P. (with Olson, J. F., Preuschoff, C., Erberber, E., Arora, A., & Galia, J.). (2008). *TIMSS 2007 International mathematics report: Findings from the IEA's trends in international mathematics and science study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- National Council of Teachers of Mathematics (1989). *Curriculum and evaluation standards for school mathematics*, Reston, VA: Author.
- National Council of Teachers of Mathematics. (1991). *Professional standards for teaching mathematics*, Reston, VA: Author
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*, Reston, VA: Author.
- National Center for Educational Statistics. (2001). *Digest for educational statistics, 2001*. Washington, DC: Government Printing Office.
- Neuschmidt, O., Barth, J., & Hastedt, D. (2008). Trends in gender differences in mathematics and

- science (TIMSS 1995–2003). *Studies in Educational Evaluation*, 34(2), 56–72.
- Onwuegbuzie, A. J. (1998). Teachers' attitudes toward statistics. *Psychological Reports*, 83, 1008–1010.
- Onwuegbuzie, A. J. (2003). Modeling statistics achievement among graduate students. *Educational and Psychological Measurement*, 63(6), 1020–1038.
- O'Dwyer, L. (2005). Examining the variability of mathematics performance and its correlates using data from TIMSS '95 and TIMSS '99. *Educational Research and Evaluation*, 11(2), 155–177.
- Penner, A. M., & Paret, M. (2007). Gender differences in mathematics achievement: exploring the early grades and the extremes. *Social Science Research* 37, 239–253.
- Phan, H., Sentovich, C., Kromey, J., Dedrick, R., & Ferron, J. (2010). *Correlates of mathematics achievement in developed and developing countries: An HLM analysis of TIMSS 2003 eighth-grade mathematics scores*. Paper presented at the annual meeting of the American Educational Research Association, Denver, CO.
- Potter, M. C., & Levy, E. (1968). Spatial enumeration without counting. *Child Development*, 39, 265–272.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: SAGE.
- Riegle-Crumb, C. (2006). The path through math: course sequences and academic performance at the intersection of race-ethnicity and gender. *American Journal of Education*, 113, 101–122.
- Riegle-Crumb, C. C., & Humphries, M. M. (2012). Exploring bias in math teachers' perceptions of students' ability by gender and race/ethnicity. *Gender and Society*, 26(2), 290–322.
- Robinson, J., & Lubienski, S. (2011). The development of gender achievement gaps in mathematics and reading during elementary and middle school: Examining direct cognitive assessments and teacher ratings. *American Educational Research Journal*, 48(2), 268–302.
- Roseth, C. J., Garfield, J. B., & Ben-Zvi, D. (2008). Collaboration in learning and teaching. *Journal of Statistics Education*, 16, 1. Retrieved from <http://www.amstat.org/publications/jse/v16n1/roseth.html>
- Ryan, K. E., & Ryan, A. M. (2005). Psychological processes underlying stereotype threat and standardized math test performance. *Educational Psychologist*, 40(1), 53–63.
- Saldanha, L.A., & Thompson, P. W. (2001). Students' reasoning about sampling distributions and statistical inference. In R. Speiser & C. Maher (Eds.), *Proceedings of the twenty-third annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Columbus, OH: ERIC Clearinghouse.
- Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465–494). New York, NY: Macmillan.
- Siegel, L. S. (1968). The development of the ability to process information. *Journal of Experimental Child Psychology*, 6, 368–383.
- Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science? *American Psychologist*, 60, 950–958.
- Smith, G. (1998). Learning statistics by doing. *Journal of Statistics Education*, 6(3). Retrieved from <http://www.amstat.org/publications/jse/v6n3/smith.html>
- Sotos, A. E. C., Vanhoof, S., Van den Noortgate, W., & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review*, 2, 98–113.
- Stiff, L. V., & Harvey, W. B. (1988). On the education of Black children in mathematics. *Journal of Black Studies*, 19(2), 190–203.
- Tate, W. F. (1994). Race retrenchment, and the reform of school mathematics. *The Phi Delta Kappan*, 75(6), 477–480, 482–484.
- TIMSS. (2003). *The trends in international mathematics and science study*. Retrieved from <http://timss.bc.edu/timss2003.html>
- TIMSS. (2007). *The Trends in International Mathematics and Science study*. Retrieved from <http://timss.bc.edu/timss2007/index.html>
- Vanhoof, S., Sotos, A., Onghena, P., Verschaffel, L., Dooren, W., den Noortgate, W., & Leuven, K. (2006). Attitudes toward statistics and their relationship with short- and long-term exam results. *Journal of Statistics Education*, 14(3). Retrieved from <http://www.amstat.org/publications/jse/v14n3/vanhoof.html>
- Ward, B. (2004). The best of both worlds: A hybrid statistics course. *Journal of Statistics Education*,

12(3). Retrieved from <http://www.amstat.org/publications/jse/v12n3/ward.html>

Williams, T., Ferraro, D., Roey, S., Brenwald, S., Kastberg, D., Leslie, J., Smith, C., Stearns, P., & Gonzales, P. (2009). *TIMSS 2007 U.S. Technical Report and User Guide* (NCES 2009–012). Retrieved from <http://nces.ed.gov/pubs2009/2009012.pdf>

Appendix 1: Pool of Variables: Second Phase of Variable Selection

**Student Variables*

-
- | | |
|--|---|
| <ul style="list-style-type: none"> • Sex • Calculator in home? • Study desk in home? • Highest education/mother • How far in school expect to go? • Like to take more math? • Enjoy learning math? • Learn quickly in math? • Like math? • Need math for other subjects? • Need math to get job • How often write equations? • How often relate to daily life? • How often use computer? • How often math homework given? | <ul style="list-style-type: none"> • Number of books in home? • Computer in home? • Internet connection in home? • Highest education/father • Usually do well in math? • Math difficult? • Math a strength? • Math boring? • Math will help in daily life? • Need math for university study? • How often interpret data in table/chart? • How often explain answers? • How often use calculator? • Time spent doing homework? |
|--|---|

**Teacher Variables*

- | | |
|---|---|
| <ul style="list-style-type: none"> • Years of teaching • Major area of study/math • Teaching certificate? • Prepared to teach/interpreting data? • How often teach/interpreting data? • How often ask students/apply concepts? • How often ask students/relate to life? • When students taught/reading data? • When students taught/descriptives? • When students taught/data displays? • When students taught/using chances? • School in safe neighborhood? • School security policies sufficient? • Classrooms overcrowded? • Characterize/job satisfaction • Characterize/parental support • Characterize/student regard/property | <ul style="list-style-type: none"> • Level of formal education • Major area of study/math education • Prepared to teach/reading/displaying data? • Prepared to teach/judging chances? • How often ask students/write equations? • How often ask students/explain answers? • Percentage of time teaching data and chance? • When students taught/organizing data? • When students taught/interpreting data? • When students taught/predicting chances? • Class size for math instruction? • Feel safe at school? • Building needs repair? • Have adequate workspace? • Characterize/expectations student achievement • Characterize/parental involvement • Characterize/student desire to do well |
|---|---|

*The complete variable descriptions are available in TIMSS (2007) Questionnaires which can be accessed at: <http://timss.bc.edu/timss2007/context.html>

Appendix 2: TIMSS 2007: Average Achievement in the Mathematics Content Domain

| | <i>Number</i> | <i>Algebra</i> | <i>Geometry</i> | <i>Data and Chance</i> |
|------|---------------|----------------|-----------------|------------------------|
| U.S. | 510 (2.7) | 501(2.7) | 480 (2.5) | 531(2.8) |

*TIMSS Scale Average: 500.

Advanced International Benchmark—625

Students can organize and draw conclusions from information, make generalizations, and solve nonroutine problems. They can solve a variety of ratio, proportion, and percent problems. They can apply their knowledge of numeric and algebraic concepts and relationships. Students can express generalizations algebraically and model situations. They can apply their knowledge of geometry in complex problem situations. **Students can derive and use data from several sources to solve multistep problems.**

High International Benchmark—550

Students can apply their understanding and knowledge in a variety of relatively complex situations. They can relate and compute with fractions, decimals, and percents, operate with negative integers, and solve word problems involving proportions. Students can work with algebraic expressions and linear equations. Students use knowledge of geometric properties to solve problems, including area, volume, and angles. **They can interpret data in a variety of graphs and tables and solve simple problems involving probability.**

Intermediate International Benchmark—475

Students can apply basic mathematical knowledge in straightforward situations. They can add and multiply to solve one-step word problems involving whole numbers and decimals. They can work with familiar fractions. They understand simple algebraic relationships. They demonstrate understanding of properties of triangles and basic geometric concepts. **They can read and interpret graphs and tables. They recognize basic notions of likelihood.**

Low International Benchmark—400

Students have some knowledge of whole numbers and decimals, operations, and **basic graphs.**

Source: Mullis, Martin, and Foy (2008).