The Development of Gender Achievement Gaps in Mathematics and Reading During Elementary and Middle School: Examining Direct Cognitive Assessments and Teacher Ratings

A Critical Analysis

Inquiry Exam

Shelli L. Casler-Failing

University at Albany

April 12, 2012

Robinson, J. P., & Theule Lubienski, S. (2011, April). The development of gender achievement gaps in mathematics and reading during elementary and middle school: Examining direct cognitive assessments and teacher rating. *American Educational Research Journal, 48*(2), 268-302.

Overview of Study

This study sought to investigate when gender gaps in mathematics and reading occur among both male and female participants in kindergarten through eighth grade, where the gaps are most pronounced, and if the measurement instrument used to assess the students impacted the results. The researchers also examined teacher assessments of the students in an attempt to find a connection between teacher expectations and student achievement, or lack thereof. Although the authors provide a wealth of background information regarding reading and mathematical skills for both males and females in grades Kindergarten through eighth grade they do not use a theoretical framework to guide their research.

The researchers gathered their data from the Early Childhood Longitudinal Study, Kindergarten Class of 1998-1999 (ECLS-K) reports, which followed students from kindergarten through eighth grade. There were 7,075 students represented in the data at each wave (test period) and a total of six waves – fall and spring of kindergarten and the spring of first, third, fifth, and eighth grades. Many types of statistical analysis were used to examine the data (male and female data analyzed separately) including, but not limited to, mean differences, least squares regression, and quantile regressions; each analysis used data from the 10^{th}, 50^{th}, and 90^{th} percentiles.

The researchers found that a gap does exist between male and females after they complete kindergarten and continues through eighth grade. The gap was found to favor males in mathematics and females in reading. The teacher ratings tended to be an

inaccurate form of cognitive ability due to the inconsistent gaps found at each wave as compared to the assessed cognitive ability.

Critical Analysis

For the purpose of evaluating this article I have chosen to follow the guidelines set forth by Isaac and Michael (1995) and will evaluate (a) the problem posed - its significance and limitations; (b) the research design – methodology, validity, and data analysis; and (c) the results of the research. I will follow this in-depth analysis with a personal reflection on the article.

The Problem

Robinson and Lubienski designed this research to determine if a gender gap exists in the content areas of mathematics and reading at any time period between the fall of kindergarten and the spring of eighth grade. The title and abstract of this article are very well written; the title clearly states the focus, sample space, and data collection and the abstract provides a detailed overview of the study and its findings.

The authors of this study provide a brief introduction to the study by writing a short depiction of educational myths centered on gender achievement. The information provided is well written and cites information from 1999 and 2000, but also creates conflict for the reader with a more recent citation from 2008. Upon further reading it is clear the conflicting information was included to provide purpose for the research set forth. Although the authors include the type of longitudinal cognitive assessments evaluated in the study, they continue to refer to “the teachers’ own assessments” (p. 269) rather than providing the actual type of assessment used by the teachers.

The authors of this study provided a detailed literature review which was broken down into four subsections: math and gender, reading and gender, possible causes of gaps, and unanswered questions, which ended with the statement of the research questions that guided their research. The questions were:

1. What are the achievement scores of males and females in reading and math from kindergarten through eighth grade? What types of skills does each group demonstrate at various points?
2. When do math and reading gender gaps first appear in elementary school, and do they widen or narrow as children progress from kindergarten to eighth grade? Are gender gaps concentrated in a particular achievement range (e.g., among low-achieving students), or are they consistent across the score distribution? Does the metric of the achievement measure (scale score, standardized score) affect the answers to these questions?
3. Are teachers' assessments of the relative progress of males and females similar to those of formal cognitive assessments?
4. How do K-8 patterns in gender gaps in reading achievement and teacher assessments compare to those in math? And what does this comparison suggest for future research in to the causes of these gaps? (p. 276).

It is my opinion that these questions are well written and answerable through the data collected. The authors used a mixture of descriptive, relationship, and comparison questions (Creswell, 2008) to evaluate the data which provides the reader with an easy-to-read format for understanding the results. I would, however, alter question three due to the fact it allows for a yes/no answer; an appropriate replacement would be to insert the word "how" at the start of the question to prompt an explanatory answer rather than a one-word response.

The United States has been evaluating and re-designing its educational curriculum for over a decade now. The first major overhaul of the twenty-first century was the No Child Left Behind Act (NCLB) of 2001 which aimed to close achievement gaps among students (No Child Left Behind Act, 2002). The NCLB was followed by the Race to the Top legislation of 2009 which aimed to improve teacher and school effectiveness by rewarding schools and states showing the most academic growth (Hamilton, 2009). At a time when student achievement is at the forefront of education a study such as this is necessary to inform the many stakeholders in the content

areas and/or grade levels in need of restructuring to ensure all children are afforded the same opportunity to meet with success.

A major limitation of this study, in my opinion, is that the reader is not given the demographics of the data collected. Robinson and Lubienski state the data included is "a nationally representative sample" (p. 276), however, of the 21,400 students first assessed in 1998-1999, only 7,075 students were included in this study. Without knowing the percentages of student ethnicity it is impossible to know if the final sample truly is representative of all school age children. For this reason, although the results are informative, they may not be reliable. Another limitation is that these children began kindergarten in 1998 and completed grade 8 in 2007, therefore at the printing of this article the results were four years old and may not provide a true depiction of the reading and mathematics gaps that may/may not exist today in response to the educational reforms of the past decade.

The Design

Methodology. Robinson and Lubienski presented their methodology in a very clear, concise manner to guide the reader through the steps taken to evaluate the data and the reasons for the types of evaluation used. Since the authors were not conducting research to collect their data (data was obtained from the Department of Education) their Methods section was written to link each evaluation process with the question(s) it was chosen to answer and to substantiate their processes with previous literature. The authors were transparent and made the reader aware of areas in which limitations may have altered their findings (e.g., "One limitation of these proficiency levels is that they convey a hierarchy of math or reading knowledge that might not always hold," p. 279).

A major flaw in the design process, I believe, was the collection of the teacher ratings. The authors imply the ratings were included in this report in an attempt to determine if teachers' ratings accurately reflected students' abilities and achievements. However, Robinson and Lubienski state "…to lesson teacher burden, math teacher survey data were collected for only half of the ECLS-K fifth and eighth graders (the other half were assigned to science" (p. 277). Although the authors ensure the reader that no bias exists, in my opinion the data is invalid. A student may perform well in science and poorly in math, thus affecting the teacher rating. I do not believe the teacher data to be reliable in response to this issue.

Validity. Internal validity will not be assessed as Robinson and Lubienski did not participate in the submission of the assessments when the data was originally collected by the Department of Education. It is important to note, though, that the authors did speak to the construct validity and reliability of the assessments. I will, however, evaluate the external validity of this research. Isaac and Michael (1995) define external validity as "the generalizability or representativeness of the experimental findings" (p. 69). The information presented within this research could impact current educational practices; therefore the external validity is an important factor to analyze.

As previously mentioned, it is unclear whether the sample used in the research was truly representative of the students currently enrolled in the U.S. educational system. Although the original sample of 21,400 kindergarten students was said to be representative, it is unknown if the final 7,075 student sample included in this study is also representative. However, the authors do state the sample is representative of the population of English-proficient students in kindergarten at the start of ECLS-K data collection due to longitudinal sample weights. The question now becomes whether English-proficient students properly represent all students and I do not believe that to be the case. As of 2009, the number of English Language Learners aged 5-

17 was 11.2 million students (Fast facts: English language learners, 2011); for this reason, I remain convinced that this sample limits the generalizability of the findings.

Data Analysis. Robinson and Lubienski used several quantitative methods to evaluate the data such as mean difference, least squares regression, quantile regression, and a metric-free measure developed by the authors. Robinson and Lubienski provide a strong background and cite many literary sources for the choice of evaluation methods, which provides the reader with a sense of reliability in their methods. Robinson and Lubienski chose to evaluate the achievement data of males and females at the 10^{th}, 50^{th}, and 90^{th} percentiles (p.281), but when they used quantile regression to assess gaps also included the 25^{th} and 75^{th} percentiles without providing an explanation as to the reason for the addition of percentiles.

Robinson and Lubienski provide detailed figures and tables of their findings that allow the reader to easily comprehend their findings. Figures 1 and 2 are shown, and described, in order to answer the first research question of mathematics and reading achievement, respectively. The authors provide a thorough interpretation of the figures and explain at which wave and percentile gaps begin to occur. The math achievement scores of males tend to be higher than that of females at each percentile, whereas the reading achievement scores favor females at only the 10^{th} and 50^{th} percentiles.

Tables 1 and 2 provide the reader with data regarding the gaps in achievement scores for both the reading and mathematics assessments using mean differences and quantile regression, respectively; however, I found these tables extremely difficult to interpret since the authors were not explicit in showing whom the scores favored - females or males. The authors refer to Table 1 in the reading on page 283, but they speak of standard deviations which are not represented in the table. On page 286 I found a reference to Table 2 stating "…shows that the gap at the 90^{th}

percentile is 0.12 SDs (in favor of males);" I used this statement to interpret Table 1 as I assumed a negative value favored females and a positive value favored males in all data. I was not able to understand the information in either table without the explanation in the body of the article referencing Table 2, which is a shortfall in my opinion.

I found Figures 3 and 4 to be the most useful in terms of showing when the gaps occur in mathematics and reading, respectively, and which gender is favored; these figures were created using the metric-free measure developed by the authors. Although I found the developed measure extremely difficult to understand, the end product produced a very easy-to-understand graphical representation, with a 95% confidence level also shown, at each percentile evaluated.

The teacher ratings were collected using a five-point scale ranging from the low of "child has not yet demonstrated the skill, knowledge or behavior" (p.277) to the high of "child demonstrates skill, knowledge, or behavior competently and consistently" (p. 277). The data was evaluated using Rasch analyses; the scores were then standardized to coincide with the cognitive assessment data. After researching this method I found that Rasch analysis is a commonly used tool to assess a range of disciplines, including education. Rasch analysis is the simplest and "the only item response theory (IRT) model in which the total score across items characterizes a person totally" (Rasch analysis, 2010); for this reason I believe this was an appropriate method to evaluate the teacher ratings.

The final three figures (Figures 5, 6, and7) represent the findings from the teacher ratings. Table 5 presents a comparison of teacher ratings in relation to the cognitive assessments for both reading and mathematics. In both instances females were rated higher than males, however the gap is less pronounced for reading in kindergarten and first grade. Again it is important to understand the standardization of the scores in order to interpret the figures – the more negative a

value the more it favors females and the more positive the more it favors males, which is not stated in the description of the figure. Figures 6 and 7 represent the metric-free gender rating for mathematics and reading, respectively. When viewing these graphs it is evident teachers rate females higher than males in both content areas, however when I compared them to Figures 3 and 4 the differences between assessment and teacher rating is not as clear, thus showing the need for Figure 5.

Results

Robinson and Lubienski found that although males and females begin kindergarten nearly identical in mathematical cognitive ability, males tend to acquire new information at a higher rate than females until middle school, at which time females tend to "catch up" to males. The formation of gaps appears to be widest in the 50th and 90th percentiles. The results of the reading assessments produced different outcomes although males and females again began kindergarten with similar abilities. Males and females in the 90th percentile tend to remain consistent with one another, however females outscore males on reading assessments by the spring of kindergarten in the 50th and 10th percentiles and the gaps continue to expand at these percentiles through eighth grade; the greatest gap is seen in the 10th percentile between fifth and eighth grades.

When interpreting the findings from the teacher ratings Robinson and Lubienski found teachers tend to rate female skills higher than that of males in both mathematics and reading at the start of kindergarten even though the cognitive assessments show males and females possess similar skill sets. Although teachers continue to rate females higher than males, in mathematics, in first grade, the males are rated closer to females' ratings, which contrast the cognitive assessments; it is in first grade when males start outperforming females in mathematics.

Although teachers consistently rated females higher than males in mathematics, the ratings shadowed the cognitive assessments as shown in Figure 5, which implies that teachers were aware of areas of growth, but were biased in which gender presented the growth.

As with mathematics, teachers continued to rate females higher than their male counterparts, however the teacher ratings were higher than the cognitive assessments. This is evident in Figure 5, where both lines favor females, but the teacher ratings are very inconsistent with the assessments.

Personal Reflection

Robinson and Lubienski created this research article in a very linear method. They began with a background on equity and social justice issues in the areas of mathematics and reading assessments by breaking the article into parts. These parts were developed, I feel, to better define the resulting research questions. As the authors worked through the data, methods, and results they again broke the findings into parts, which I feel was to ensure the reader connected the findings with the question(s) to which it answered, even though this was not explicitly stated by the authors.

The results were very clearly stated by the authors and provided the reader with much important information, although its use may be limited due to the aforementioned validity factors. I found this article to be relevant to my own area of interest, writing in mathematics, and would be interested in the outcomes of a similar study on more recent student data. The limitation of this, however, is the fact that the data is longitudinal and therefore requires nine years to collect.

I found the background and literature review to be extremely informative, but I do feel that a theoretical lens through which to analyze the data would have been useful and may have

provided insight into the teacher rating differences as compared to the cognitive differences. Another shortfall is the fact that I, the reader, was forced to believe the authors when they stated the data was a national representation of students without any evidence of such. The results would be more reliable and valid had I known the percentage of minority and/or English Language Learners. Another area that would have impacted the findings is to know the teacher sample. It is implied the teacher ratings were completed by the teachers of the students assessed, including science teachers, but it is unclear if the teachers' ratings were unbiased and representative of their genuine interpretations of students' abilities.

Overall, I do feel this research sheds light on an important shortcoming of our educational system and cites the need for curriculum and teacher reform to minimize the gaps so all students will be able to achieve success. Robinson and Lubienski make two strong points in stating "…future math-focused interventions with females might be better targeted toward elementary grades…" (p. 296) and "…for reading…during elementary and middle school, we should focus more attention on the lowest-achieving males" (p.298). The only wording I would change in these comments is that we, the educators of our students, should focus attention on *all* of our students to ensure the success of *every* student. The authors finish their article with a very strong statement that I fully agree with, "[S]tudents of both genders deserve instruction that will optimize their reading and math learning" (p. 298).

References

No Child Left Behind Act. (2002). Retrieved March 16, 2012, from National Education Association: http://www.nea.org/home/NoChildLeftBehindAct.html

Rasch analysis. (2010). Retrieved March 16, 2012, from Rasch Analysis: http://www.rasch-analysis.com/rasch-model-specification.htm

Fast facts: English language learners. (2011). Retrieved March 16, 2012, from National Center for Education Statistics: http://nces.ed.gov/fastfacts/display.asp?id=96

Creswell, J. W. (2008). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research.* Upper Saddle River: Pearson Education.

Hamilton, J. (2009, July 24). *President Obama, U.S. Secretary of Education Duncan Announce National Competition to Advance School Reform*. Retrieved March 16, 2012, from U.S. Department of Education: http://www2.ed.gov/news/pressreleases/2009/07/07242009.html

Isaac, S., & Michael, W. B. (1995). *Handbook in research and evaluation.* San Diego: EdITS.

Robinson, J. P., & Theule Lubienski, S. (2011, April). The development of gender achievement gaps in mathematics and reading during elementary and middle school: Examining direct cognitive assessments and teacher rating. *American Educational Research Journal, 48*(2), 268-302.