

# Students With Specific Difficulties in Geometry: Exploring the TIMSS 2011 Data With Plausible Values and Latent Profile Analysis

Learning Disability Quarterly  
2021, Vol. 44(1) 11–22  
© Hammill Institute on Disabilities 2020  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/0731948720905099  
journals.sagepub.com/home/ldq  
SAGE

Jinsong Chen<sup>1</sup> , Liangfang Li<sup>1</sup>, and Dake Zhang<sup>2</sup>

## Abstract

This study analyzed the Trends in International Mathematics and Science Study 2011 data on fourth-grade U.S. students' mathematics performance to answer four research questions: (a) How did U.S. students' geometry performance compare with their performance on the other mathematics subscales? (b) What were the patterns of student achievement among the mathematics subscales? (c) Was there a group of students who demonstrated specific difficulties in geometry only? and (d) Which demographic variables contributed to students' classification in the group with geometry difficulties? We found that (a) U.S. students' performance was poorer on the Geometry subscale than on other mathematics subscales; (b) using latent profile modeling, we identified a group of students with the lowest scores across all three mathematics subscales who showed a significant discrepancy between Geometry and the other subscales that did not exist within the high-achieving and average-achieving groups; and (c) gender, age, home language, race, and preference for mathematics and science significantly influenced the probability of being classified in the group with the lowest performance and the largest gap between Geometry and other mathematics subscales. Implications for educational theory and practice are discussed.

## Keywords

geometry, learning difficulty, TIMSS, latent profile analysis, plausible value

There are many reasons for recommending that schools pay particular attention to children's achievement in geometry. As a subject in K-12 mathematics education, geometry instruction involves helping students to learn spatial relations and properties of shapes, which are crucial for students to succeed in science, technology, engineering, and mathematics (STEM) subjects at the college level (Hsi et al., 1997; Wai et al., 2010). However, geometry instruction is often overlooked in current education research and practice in the United States (Clements & Sarama, 2011). A historical review of the geometry curriculum in U.S. schools (Sinclair, 2008) suggested a long history of neglecting geometry instruction. Not until 1844, when geometry began to be required for college entrance, did U.S. high school administrators and teachers realize the necessity to add this subject to the mathematics curriculum. Therefore, the geometry curriculum was initiated with an elite college-bound status possibly because of Euclidean geometry's link to ancient Greek scholars, whereas other mathematics domains are typically rooted in everyday problem solving and applications.

U.S. students' performance in geometry can be considered less than satisfactory. Although fourth-grade U.S. students showed progress in numerical subscales in the Trends in International Mathematics and Science Study (TIMSS)

2015 report, there was a significant drop of 9 standardized points from the 2011 testing performance in the subscale of Geometric Shapes and Measures (Provasnik et al., 2016). An earlier study showed that although TIMSS reported significant improvement in U.S. eighth-graders' algebra performance between 1999 and 2003, significant improvement was not found in their geometry performance during that period (Gonzales et al., 2004).

## Theoretical Framework

The categorization of subtypes of math learning disabilities or mathematics difficulties has been in discussion.

<sup>1</sup>Sun Yat-sen University, Guangzhou, China

<sup>2</sup>Rutgers University–New Brunswick, NJ, USA

### Corresponding Authors:

Dr. Jinsong Chen, Lab of Psychometrics & Artificial Intelligence, Department of Psychology, Sun Yat-sen University, Guangzhou 510006, China.

Email: jinsong.chen@live.com

Dr. Dake Zhang, Graduate School of Education, Rutgers University, 10 Seminary Place, Room 312, New Brunswick, NJ 08901, USA.

Email: dake.zhang@gse.rutgers.edu

Some researchers have argued that students' particular difficulties in geometry may represent a unique and specific mathematics learning disability subtype, different from other subtypes of mathematics difficulties. For example, Geary and Hoard (2005) described three subtypes of students with mathematics difficulties, including procedural (i.e., children present a delay in acquiring simple arithmetic strategies), semantic memory (i.e., children show deficits in retrieval of facts because of a long-term memory deficit), and visual-spatial difficulties (i.e., children show deficits in the spatial representation). Karagiannakis et al. (2014) categorized mathematics learning disabilities into four subcategories: core number, memory retrieval and processing, reasoning, and visual-spatial. Similar conceptual categorizations were also advocated by the National Center for Learning Disabilities (2006) who listed visuo-spatial impairments as a distinct area of weakness exhibited by many students with mathematics difficulties. Recently, Bartelet et al. (2014) assessed more than 200 elementary school children with mathematics learning disabilities with a battery of cognitive instruments. With a data-driven approach, they reported six distinct groups of mathematics learning disabilities, including a spatial difficulties group, who were particularly weak in visual working memory. Evidence from neuropsychology has demonstrated that spatial deficits may be associated with dysfunction in posterior regions of the right hemisphere and may be related to the parietal cortex of the left hemisphere (Malhotra et al., 2009).

Spatial abilities refer to the ability to understand, reason, retain, retrieve, and transform visual images, and are highly predictive to students' geometry performance (Clements et al., 1997; Giofrè et al., 2013; Kytälä & Lehto, 2008; Spelke et al., 2010), and interestingly this correlation is even greater in students with poorer geometry performance (Battista, 1990). Deficits in spatial abilities have been documented to explain students' difficulties with geometry (Clements et al., 1997; Passolunghi & Mammarella, 2012). Given the strong relation between the geometry subject and spatial abilities, it sounds plausible to hypothesize that geometry difficulties may represent a unique and specific learning difficulty due to the high reliance on spatial abilities.

Contrarily, the construct of "geometry difficulty" as a specific mathematics learning difficulty can be questioned for a few reasons. First, many mathematical tasks require spatial thinking (van Garderen, 2006), and a deficit in spatial abilities may affect not only geometry achievement but also a broad range of mathematics domains. Second, the reasons why students encounter geometry difficulties may be multifaceted: A recent meta-analysis (Peng et al., 2016) found that the literature (Giofrè et al., 2013, 2014; Passolunghi et al., 2008) reported a weak relation between geometry achievement and all categories of working

memory, including visual-spatial working memory. On top of spatial abilities, students' geometry achievement is also influenced by many other cognitive factors, such as verbal working memory (Bizarro et al., 2018), fluid intelligence, and reasoning (Giofrè et al., 2014). In particular, because geometry includes considerable proof-oriented problems, it is typically considered highly related to the deductive thinking (Dawkins, 2015) and verbal logical reasoning (Battista, 1990). Geometry learning difficulties can also be related to non-cognitive factors, including motivation and persistence (Nichols, 1996), emotions (Bailey et al., 2014), meta-cognition abilities (Aydın & Ubuz, 2010), knowledge (Bokosmaty et al., 2015), and how they use knowledge (Lawson & Chinnappan, 1994). And these non-cognitive problems universally exist in all mathematical domains for struggling students, which also challenges the hypothesis that geometry difficulties should be viewed as a specific and unique subtype of math learning difficulty.

That said, there has been sparse research that provides empirical evidence to support the above theoretical conjectures that geometry learning difficulties are specific and unique among mathematics learning difficulties. There is also little research on the relation between performance on geometry and other mathematics content domains. There is a need to empirically explore the existence of "specific geometry difficulties" by investigating whether there is a special group of students who demonstrate a discrepancy between geometry achievement and achievement in other mathematics areas. Specifically, we will examine whether (a) there is a group with difficulties in geometry who show average or above-average performance in other mathematics areas, or (b) there is a group showing difficulties with performance on multiple mathematics subscales and more severe difficulties on the Geometry subscale than on other mathematics subscales. The availability of existing large-scale achievement assessment data, such as TIMSS, makes it possible to examine the existence of such groups.

## Research Questions

In the present study, we aimed to answer four research questions based on an analysis of TIMSS-2011 fourth-grade mathematics data:

**Research Question 1:** How did U.S. students' geometry performance compare with their performance on the other mathematics subscales?

**Research Question 2:** What were the patterns of student achievement among the mathematics subscales?

**Research Question 3:** Was there a group of students who demonstrated specific difficulties in geometry only?

**Research Question 4:** Which demographic variables contributed to students' classification in the group with geometry difficulties?

## Method

### *TIMSS in Mathematics*

The TIMSS provides reliable and timely data on school mathematics and science achievement in the United States and other countries. TIMSS data have been collected every 4 years from students in fourth and eighth grades since 1995. In this research, we analyzed TIMSS (2011) fourth-grade data to examine U.S. students' geometry performance. It is worth noting that the analyses are nationally generalizable with the available sampling weights (Foy et al., 2013). The 2011 mathematics assessment at the fourth-grade level included three content subscales, namely, Data Display (Display), Geometric Shapes and Measures (Geometry), and Number. TIMSS also provided demographic and related information data about the participating students. We chose the fourth-grade rather than the eighth-grade data because fourth-grade geometry is closer to so-called "intuitive geometry," which is independent from instruction, experiences, and culture (Giofrè et al., 2013). In contrast, eighth-grade geometry involves a greater proportion of geometry academic content that is closely dependent on learning and education and closely related to other mathematics subjects such as algebra (Giofrè et al., 2014). Specifically, for fourth graders, the TIMSS Geometry scale includes items such as "Length of string pulled straight," "Position of shape after a 1/2 turn," "Pieces of cardboard to make shape," "Rotate the shape 1/4 clockwise," and "Relating net with its 3-D figure."

TIMSS provides five plausible values (von Davier et al., 2009) per content subscale. Within the context of large-scale assessments, plausible values are random numbers that are drawn from the distribution of scores of individual students—that is, from the marginal posterior distribution. Multiple plausible values are more accurate and can better capture the expected values and variance in subgroups, especially when the true distribution of mastery or accurate estimates of individual performance are difficult to obtain on short tests. The main challenge of analyzing plausible values is based on the method of multiple imputation, which is widely used to analyze missing data (Schafer & Graham, 2002). With latent variables, multiple imputation models are more difficult to converge, with fewer fit statistics available for model assessment. For this research, we chose Mplus (L. K. Muthén & Muthén, 1998–2017) to analyze the five sets of plausible values with multiple imputation and accordingly used all the fit and descriptive statistics that were available in the software for analysis. For missing values, missing at random was assumed. Finally, appropriate sampling weights have been applied based on TIMSS User Guide (Foy et al., 2013). Note that Mplus was the only software we found satisfying our requirements, namely, latent profile analysis (LPA) with covariate, multiple imputation, and sampling weights.

### *Data Analysis Plan*

We followed a three-step analysis. First, we calculated the mean scores of the three mathematics subscales of TIMSS and compared the difference between any of the two subscales using *Z* tests to answer the first research question. We also provided correlations between the three subscales. Second, to answer Research Questions 2 and 3, we used LPA to classify students into different groups. We hypothesized a group of students would perform poorly on the Geometry subscale only and would perform at average or above-average levels on the other mathematics subscales. Similar to latent class analysis (LCA), LPA is a form of mixture modeling used to identify different latent classes of individuals with similar response patterns based on a set of observed variables (Lubke & Muthén, 2005). The difference is that the LPA is for continuous observed data, whereas the LCA is for categorical observed data. To determine the number of classes, we compared a range of latent class models with different number of classes in terms of their relative model fit. For LCA or LPA with multiple imputation in Mplus, three fit indices are available, that is, Akaike's (1974) information criterion (AIC), Schwarz's (1978) Bayesian information criterion (BIC), and entropy. Entropy is an index to measure the variability, or chaos, in a stochastic system, with higher values indicating greater precision of classification. Entropy with values approaching 1 indicates clear delineation of classes, while those above .8 are generally considered acceptable in assigning individual cases into appropriate classes (Celeux & Soromenho, 1996).

Third, to answer Research Question 4, we examined the effects of demographic variables on class membership using the LPA with covariates. LPA or LCA with covariates is analogous to a multinomial logistic regression approach with latent class membership serving as a categorical dependent variable and the covariates as independent variables (B. O. Muthén & Satorra, 1995). Among different approaches to incorporate covariates in latent classes (Vermunt, 2010), the one-step approach was adopted due to the exploratory nature of the present research. The one-step approach reduces the methodological difficulties in addressing latent classes with plausible values. In a more confirmatory or high-stake context, however, the more sophisticated three-step approach (see Vermunt, 2010) could be more appropriate.

We tested nine demographic covariates in the latent class models of interest: (a) gender, (b) age, (c) English (English spoken at home, 1 = *always* to 4 = *never*), (d) Spanish versus other non-English language spoken at home, (e) how much the student likes mathematics (likemath; 1 = *like* to 3 = *not like*), (f) how much the student likes science (likesci; 1 = *like* to 3 = *not like*), (g) White or Black (WvsB), (h) White or Hispanic (WvsH), and (i) White or other non-Black or Hispanic races (WvsO). The covariates were

**Table 1.** Means and Correlations Among Three Subscales.

Subscales	<i>M</i>	<i>SD</i>	Display	Geometry	Number
Display	544.65	75.55	1.0	9.953***	1.605
Geometry	534.56	84.91	0.903**	1.0	8.327***
Number	543.10	77.53	0.897***	0.903***	1.0

Note. Below the diagonal are the correlation coefficients among the three subscales. Above the diagonal are the *Z* values comparing the mean scores of the three subscales.

\* $p < .1$ . \*\* $p < .05$ . \*\*\* $p < .001$ .

**Table 2.** Model Fitting Indices Under Each Classification.

No. of classes	AIC	<i>SD</i>	BIC	<i>SD</i>	Entropy
2	418,337.558	491.761	418,411.948	491.761	0.827
3	407,617.199	630.861	407,721.35	630.861	0.864
4	401,460.164	547.765	401,594.066	547.765	0.871
5	397,456.498	462.982	397,620.155	462.982	0.873
6	395,239.118	420.827	395,432.532	420.827	0.863
7	393,781.306	467.284	394,004.475	467.284	0.861
8	392,811.421	393.191	393,064.347	393.191	0.849

Note. AIC = Akaike information criterion; BIC = Bayesian information criterion.

included in the model one by one to avoid the possible impact of multicollinearity among the covariates.

## Results

### Descriptive Statistics

Descriptive statistics among the three mathematics subscales of TIMSS (i.e., Display, Geometry, and Number) are displayed in Table 1. Students' performances on the three subscales were highly correlated, with correlation coefficients of .903, .903, and .897 for the Geometry–Display, Geometry–Number, and Display–Number relationships, respectively. The mean score of Geometry was substantially lower than the mean score of Display ( $Z = 9.953, p < .001$ ) and the mean score of Number ( $Z = 8.327, p < .001$ ), while the mean scores of Display and Number were close to each other ( $Z = 1.605, p > .05$ ).

### LPA

To determine the appropriate number of classes, we compared models from two to eight classes using LPA, and the goodness of model fit can be found in Table 2, with smaller values of AIC and BIC for a better model. However, AIC and BIC should not be used alone for model comparisons with plausible values because of a concern about their uncertainty (i.e., *SD*) due to multiple imputation (Chaurasia & Harel, 2012). Instead, we combined them with the entropy index. In Table 2, one can see that both AIC and BIC values decrease with models of more classifications, but the trend of decreasing is much alleviated starting from

the five-class model. Considering the highest entropy (0.873) among all models, the five-class model seemed to be the best choice. To account for the uncertainty of AIC or BIC, two adjacent models with the second and third highest entropies, that is, the four-class and six-class models, were also selected for analysis simultaneously.

**Four-class model.** There were descending trends of all mean scores from the first to the last class. Latent Class 1 (Class 1; 16.9% of the entire sample,  $N = 2,122$ ) exhibited the highest mean scores of all three subscales (Display,  $M = 648.16$ , Geometry,  $M = 651.40$ , and Number,  $M = 649.94$ ). Thus, this class was the group with the highest scores across all three mathematics subjects, and the group members achieved higher than the Advanced International Benchmark (625). In contrast, Class 4 demonstrated the lowest mean scores across all three subscales (Display,  $M = 416.26$ , Geometry,  $M = 389.78$ , and Number,  $M = 412.34$ ) and is the only group whose members performed below the Low International Benchmark (400) in geometry, while performing above the Low International Benchmark on both the Number and Display subscales; 1,411 students were categorized into this group. Among the two middle-level classes, Class 2 members performed above the High Benchmark (550) on all three mathematics domains, and the Class 3 members performed above the Intermediate Benchmark (475) on all three mathematics subjects.

To understand the discrepancy between Geometry and the other two subscales in each of the latent classes, we further examined the differences in the mean level of subscales across different classes of students. Table 3 shows that only in Class 3 and Class 4, the two lowest performing groups,

**Table 3.** Characteristics of Subscales From the Four- to Six-Class Models (Weighted).

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
<b>Four-class model</b>						
Display	648.16 (6.51)	572.37 (6.22)	501.50* (7.18)	416.26** (7.35)		
Geometry	651.40 (6.57)	565.88 (6.95)	485.46 (7.81)	389.78 (8.06)		
Number	649.94 (6.50)	571.27 (6.72)	497.19* (6.69)	412.34** (7.31)		
<i>n</i>	2,122	4,982	4,054	1,411		
Proportions	0.169	0.396	0.323	0.112		
<b>Five-class model</b>						
Display	664.01 (4.72)	596.25 (4.36)	537.07 (5.03)	473.81* (7.169)	396.00** (7.74)	
Geometry	669.45 (4.56)	593.39 (4.42)	525.37 (5.23)	454.31 (7.28)	366.67 (9.68)	
Number	666.34 (5.33)	596.30 (5.01)	534.36 (5.27)	468.58* (6.37)	392.82** (8.50)	
<i>n</i>	1,303	3,599	4,215	2,622	831	
Proportions	0.104	0.286	0.335	0.209	0.066	
<b>Six-class model</b>						
Display	679.78 (8.70)	617.62 (8.26)	564.89 (8.18)	513.17 (8.55)	455.38* (8.61)	382.44** (9.57)
Geometry	687.08 (10.22)	617.55 (9.62)	557.43 (9.40)	498.41 (8.98)	433.57 (9.33)	351.14 (12.24)
Number	682.66 (7.89)	618.46 (8.64)	563.56 (8.34)	509.34 (8.50)	449.75* (8.21)	379.72** (10.74)
<i>n</i>	757	2,487	3,681	3,238	1,852	553
Proportions	0.060	0.198	0.293	0.258	0.147	0.044

Note. Standard errors in parentheses. Means and stand errors are based on weighted data. H0: The subscale and Geometry means are not significantly different.

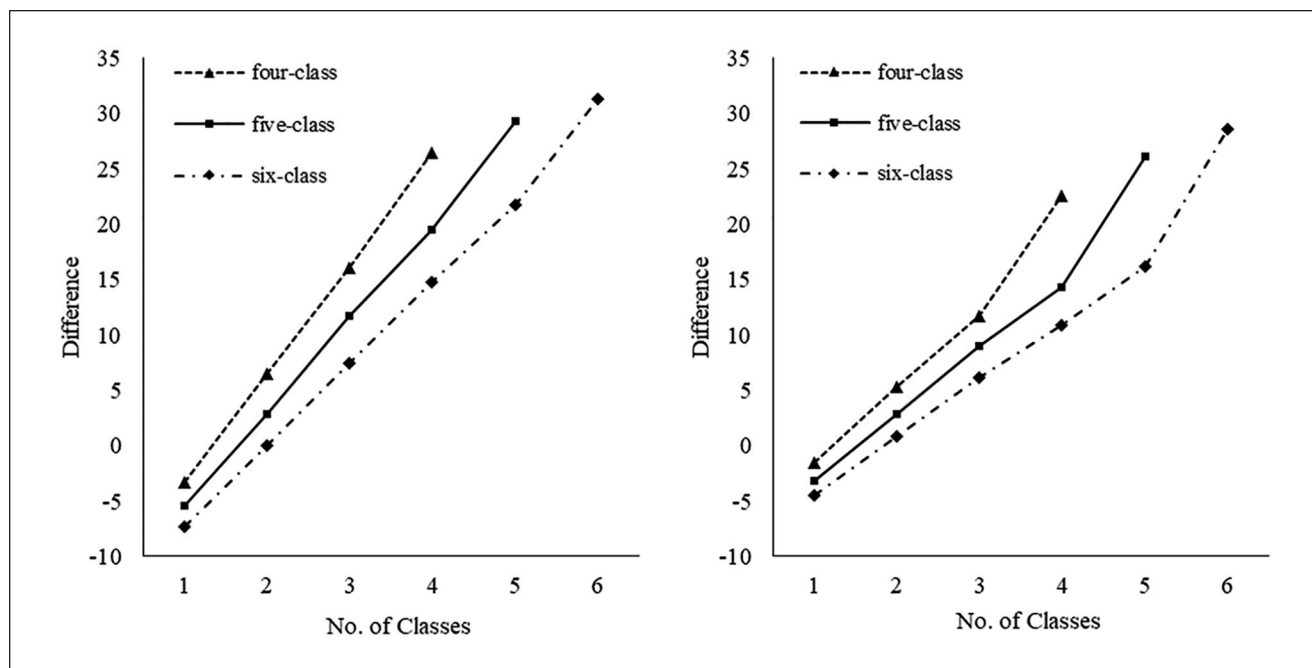
\* $p < .05$ . \*\* $p < .01$ .

group members' mean score in geometry was significantly lower than in the other two subscales. No significant gaps between Geometry and the other two subscales were found in Class 1 and Class 2. When we further calculated the differences between the mean scores of Geometry and the other two subscales (see Figure 1), we found a clear pattern of increasing differences from the first to the last class.

**Five-class model.** Latent Class 1 (Class 1; 10.36% of the entire sample,  $N = 1,303$ ) exhibited the highest mean score (Display,  $M = 664.01$ , Geometry,  $M = 669.45$ , and Number,  $M = 666.34$ ) of all three subscales, which surpassed the Advanced International Benchmark (625). Conversely, Class 5 exhibited the lowest mean scores in all three subscales (Display,  $M = 396.00$ , Geometry,  $M = 366.67$ , and Number,  $M = 392.82$ ). All three subscale means of Class 5 were below the Low International Benchmark (400), with the Geometry significantly lower than the other two subscales. The second lowest class, Class 4, performed between the Low International (400) and the Intermediate Benchmark (475), with the Geometry score (454.31) significantly inferior to the other two measures (Display,  $M = 473.81$

and Number,  $M = 468.58$ ). Class 3 members performed between the High Benchmark and the Intermediate Benchmark, and Class 2 members performed between the High Benchmark (550) and the Advanced Benchmark across all three mathematics subjects. No significant gaps between Geometry and the other two subscales were found in Class 1, Class 2, and Class 3. Similar to the four-class model, we calculated the differences between the mean score of the Geometry and the two subscales. A similar pattern of increasing differences from the first to the last class was identified, as displayed in Figure 1.

**Six-class model.** With this model, we divided students into six groups (see Table 2). The mean score of each class decreased gradually from Class 1 to Class 6. Class 1 was composed of students with the highest achievement and above the Advanced Benchmark (625) across all the three subscales, whereas the lowest achieving students were categorized into Class 6, whose members scored below the Low Benchmark (400). In the second lowest performing class, Class 5, participants scored between the Low International (400) and the Intermediate (475) Benchmarks. A



**Figure 1.** The differences between the Geometry and other subscales under the four-class, five-class, and six-class models, respectively.

Note. Left panel for Display-Geometry; right panel for Number-Geometry.

significant discrepancy between Geometry and the other two subscales was only found in the two lowest performing latent classes, Class 6 and Class 5. Similar to above, one can find increasing differences between the mean scores of Geometry and the other two subscales from the first to the last class (see Figure 1).

**Summary.** The four-, five-, and six-class models demonstrated very similar patterns, with descending trends of all mean scores from the first to the last class, together with a clear pattern of increasing differences between Geometry and other subscales' mean scores along the same line. The differences also tended to be increasingly significant from the first to the last class. The results implied that the gap between geometry and other domains tended to be larger when students' overall mathematical ability decreased. Moreover, the greater the number of latent classes, the more likely that the lowest performing group members showed (a) the poorest performance on all three subscales and (b) a lower performance in Geometry than in the other content domains. In answer to Research Question 3, we did find a lowest performing group whose members performed extremely low in geometry (i.e., did not meet the Low Benchmark of 400 points) and showed a discrepancy between geometry and other areas; however, we cannot make the argument that this group represented a population with learning difficulties in geometry only, because their other subscale scores could be below the Intermediate

Benchmark (475) but be significantly higher than their geometry scores.

### LPA With Covariates

Last, we tested whether any demographic factors would affect the membership as classified in the above LPA. We incorporated the demographic variables as covariates in all latent profile models and found very similar results across all three models. To save space, here we only presented results from the five-class model, which appeared to be more representative. Results from other models were similar and available upon request. Class 5, or the last class with the lowest mean scores and the largest discrepancy between Geometry and the other subscales, was set as the reference group. Table 4 shows the estimates of the intercepts ( $\beta_0$ ), regression coefficients ( $\beta_1$ ), and related odds and odds ratios (ORs), with further explanation below.

**Gender.** Girls were used as the baseline group (i.e.,  $X = 0$ ). Based on the intercepts and related odds, girls were significantly more likely to be in Classes 2, 3, and 4 ( $\beta_0 = 1.451, 1.638, 1.203$ ; odds = 4.267, 5.145, 3.330, corresponding to Classes 2, 3, and 4 vs. Class 5, respectively) rather than Class 5, the reference class. Based on the coefficients and related ORs, boys were slightly more likely to be included in Class 5 than Class 2, 3, or 4 in comparison with girls. Moreover, Table 4 shows that the odds for boys of being in

**Table 4.** Five-Class Model With Covariates.

Covariate	Latent class							
	Class 1 vs. 5		Class 2 vs. 5		Class 3 vs. 5		Class 4 vs. 5	
	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$
Gender	0.253	0.384**	1.451**	0.024	1.638**	-0.043	1.203**	-0.110
Age	0.474†	-0.025	1.515**	-0.117	1.683**	-0.162	1.173**	-0.060
English	0.727**	-0.600**	1.703**	-0.507**	1.794**	-0.338**	1.248**	-0.155†
Spanish	-0.286	1.024**	1.245**	0.105	1.633**	-0.268	1.245**	-0.260
Likemath	0.893**	-0.637**	1.734**	-0.368**	1.787**	-0.218**	1.275**	-0.149*
Likesci	0.644*	-0.335**	1.617**	-0.264**	1.755**	-0.218**	1.289**	-0.200**
WvsB	1.277**	-4.116**	2.228**	-2.914**	2.216**	-1.866**	1.500**	-0.916**
WvsH	1.390**	-2.364**	2.290**	-1.637**	2.255**	-0.998**	1.544**	-0.555**
WvsO	0.581	-0.271	1.637**	-0.460*	1.778**	-0.416†	1.244**	-0.262
	$\beta_{0s}$ odds	$\beta_{1s}$ OR	$\beta_{0s}$ odds	$\beta_{1s}$ OR	$\beta_{0s}$ odds	$\beta_{1s}$ OR	$\beta_{0s}$ odds	$\beta_{1s}$ OR
Gender	1.288	1.468	4.267	1.024	5.145	0.958	3.330	0.896
Age	1.606	0.975	4.549	0.890	5.382	0.850	3.232	0.942
English	2.069	0.549	5.490	0.602	6.013	0.713	3.483	0.856
Spanish	0.751	2.784	3.473	1.111	5.119	0.765	3.473	0.771
Likemath	2.442	0.529	5.663	0.692	5.972	0.804	3.579	0.862
Likesci	1.904	0.715	5.038	0.768	5.783	0.804	3.629	0.819
WvsB	3.586	0.016	9.281	0.054	9.171	0.155	4.482	0.400
WvsH	4.015	0.094	9.875	0.195	9.535	0.369	4.683	0.574
WvsO	1.788	0.763	5.140	0.631	5.918	0.660	3.469	0.770

Note.  $\beta_0$  = estimates of the intercepts;  $\beta_1$  = regression coefficients; OR = odds ratio.

† $p < .1$ . \* $p < .05$ . \*\* $p < .01$ .

Class 1 rather than Class 5 was 1.468 times as the corresponding odds for girls (OR = 1.468), the coefficient of which was significant ( $\beta_1 = 0.384$ ,  $p = .003$ ). The results show that girls were more likely to be found in the middle classes, and boys in the highest or lowest classes.

**Age.** The oldest students were used as the baseline group (i.e.,  $X = 0$ ) in the analysis. The significant intercepts and related odds suggested that the oldest students were significantly more likely to be in Classes 2, 3, and 4 rather than Class 5 ( $\beta_0 = 1.515$ , 1.683, 1.173; odds = 4.549, 5.382, 3.232, respectively). In contrast, the ORs of regression ( $\beta_1$ ) were slightly smaller than 1 for each latent class (ORs = 0.975, 0.890, 0.850, 0.942, respectively), which indicated that younger students had a slightly higher probability than older students to be included in Class 5 than in any other classes.

**English spoken at home.** Students reported the frequency of speaking English in the home environment on a Likert-type scale from 1 (*always*) to 4 (*never*), with the category of *always* as the intercept group in the analysis. The significant intercepts and related odds suggested that students always speaking English at home were significantly less likely to be in Class 5 than in any other classes ( $\beta_0 = 0.727$ , 1.703, 1.794, 1.248; odds = 2.069, 5.490, 6.013, 3.483, respectively). All

negative and significant regression coefficients suggested that speaking less English at home significantly increased the chance to be included in Class 5 than in any other classes except for Class 4 ( $\beta_1 = -0.600$ ,  $-0.507$ ,  $-0.338$ , corresponding to Classes 1, 2, and 3 vs. Class 5, respectively). That is, the students who always speak English at home had a greater chance to be in Classes 1, 2, and 3 than in the lowest performing class, Class 5. The change of chance from Class 1 to Class 5 was especially salient from being a native English speaker to non-native speakers, as suggested by the ORs of regression in Table 4.

**Spanish spoken at home.** Students reported what language was spoken at home other than English, and speaking Spanish was set to be the baseline group. The intercepts and related odds suggested that students speaking Spanish at home were significantly more likely to be in Classes 2, 3, and 4 than Class 5 ( $\beta_0 = 1.245$ , 1.633, 1.245; odds = 3.473, 5.119, 3.473, respectively). However, speaking Spanish was less likely to be in Class 1 than Class 5. Moreover, the only significant regression coefficient ( $\beta_1 = 1.024$ ,  $p < .001$ ) and related OR (= 2.784) suggested that speaking a language other than Spanish at home significantly increased the chance to be included in Class 1 rather than Class 5 by 3 times.

*How much students liked mathematics.* Students reported how much they liked mathematics on a Likert-type scale from 1 (*like*) to 3 (*not like*), with the category of *like* as the intercept group. The significant intercepts and related odds suggested that students who liked mathematics were significantly less likely to be in Class 5 than in any other classes ( $\beta_0 = 0.893, 1.734, 1.787, 1.275$ ; odds = 2.442, 5.663, 5.972, 3.579, corresponding to Classes 1, 2, 3, and 4 vs. Class 5, respectively). In contrast, negative regression coefficients were significant, and suggested that students who did not like mathematics had a higher chance to be included in Class 5 than in any of the other classes ( $\beta_1 = -0.637, -0.368, -0.218, -0.149$ ; ORs = 0.529, 0.692, 0.804, 0.862, corresponding to Classes 1, 2, 3, and 4 vs. Class 5, respectively). For example, if the preference for mathematics decreased by one unit, the odds of being in Class 1 relative to Class 5 dropped by about half (OR = 0.529).

*How much students liked science.* Students reported how much they like learning sciences on a Likert-type scale from 1 (*like*) to 3 (*not like*), with the category of “like” as the baseline group in the analysis. The intercepts and related odds suggested that students who liked science were significantly more likely to be in Classes 1, 2, 3, and 4 than Class 5 ( $\beta_0 = 0.644, 1.617, 1.755, 1.289$ ; odds = 1.904, 5.038, 5.783, 3.629, respectively). Similar to above, all regression coefficients were significant and negative, suggesting that the students who did not like learning science were more likely to be included in Class 5 than in any of the other classes ( $\beta_1 = -0.335, -0.264, -0.218, -0.200$ , respectively).

*White/Black.* Students’ race was coded into a set of dummy variables. White students were set as the reference in the dummy variables. The significant intercepts and related odds suggested that White students were significantly less likely to be in Class 5 than in any other classes ( $\beta_0 = 1.277, 2.228, 2.216, 1.500$ ; odds = 3.586, 9.281, 9.171, 4.482, respectively). All regression coefficients were negative and significant, showing that Black students were more likely to be included in Class 5 than in any other classes in comparison with White students ( $\beta_1 = -4.116, -2.914, -1.866, -0.916$ , respectively). It is noteworthy that the regression’s ORs were 0.016, 0.054, and 0.155 for Classes 1, 2, and 3 vs. Class 5, respectively, suggesting that Black students were about 70, 20, and 6 times more likely than White students to be included in Class 5 than in Classes 1, 2, and 3, respectively.

*White/Hispanic.* Similar to above, White students had a higher probability of being in each of the latent classes other than the reference latent class, Class 5. All regression coefficients were significant and negative, suggesting that

being a Hispanic student significantly increased the chance to be in Class 5, the lowest performing class, than in any other classes in comparison with White students ( $\beta_1 = -2.364, -1.637, -0.998, -0.555$ , respectively). Further analysis with the ORs of the regression coefficients suggests that Hispanic students were approximately 11, 5, 3, and 2 times more likely than White students to be in Class 5 than in Classes 1, 2, 3, and 4 (ORs = 0.094, 0.195, 0.369, 0.574, respectively).

*White/Other races.* This dummy variable compares White students’ chances of being in Class 5 versus students of Other races (i.e., non-Hispanic/Black). Similar to above, White students had a lower probability of being in Class 5 than in any other classes. In contrast, being a student of Other races increased the chance to be in Class 5 than in any other classes in comparison with White students, and the increment was significant or marginally significant for the two middle-achieving classes, Classes 2 and 3 ( $\beta_1 = -0.460, -0.416$ , respectively).

*Summary.* As the results showed in the above five-class model analysis, all covariates were significantly associated with class membership in specific ways. In summary, the following conclusions can be reached: (a) Female students were more likely to be classified in the middle-achieving groups, and male students were more likely to be classified into the advanced or lowest groups; (b) older students were more likely to be classified in a higher achieving group than in the lowest performing group; (c) students who spoke English at home had a greater chance to be classified in a high-achieving group than in Class 5; (d) students’ preference for mathematics and science significantly decreased their likelihood to be in the lowest achieving group; and (e) Black and Hispanic students who scored lower on the Geometry subscale were more likely than White students to be classified in the lowest achieving group.

## Discussion

This study aimed to portray a profile of students who performed poorly on a geometry standardized assessment by investigating the TIMSS 2011 mathematics data of fourth graders in the United States. We aimed to answer four major questions: (a) How did U.S. students’ geometry performance compare with their performance on the other mathematics subscales? (b) What were the patterns of student achievement among the mathematics subscales? (c) Was there a group of students who demonstrated specific difficulties in geometry only? and (d) Which demographic variables contributed to students’ classification in the group with geometry difficulties?

First, results suggested that, in general, geometry was the weakest mathematics content domain for fourth graders



in the United States as reflected in the comparison of mean scores on the mathematics subscales in the TIMSS 2011 data. This finding echoes a previous study analyzing the TIMSS 2011 data that reported that among the four mathematics content domains (i.e., algebra, number, data and chance, and geometry), geometry was the weakest area of proficiency for U.S. eighth graders (Provasnik et al., 2012). In contrast, Shanghai students scored higher on the Geometry subscale compared with Display and Number (Organisation for Economic Co-Operation and Development, 2014). A recent cross-culture comparison study (Kern & Henrick, 2016) suggested that, although Chinese teachers outperformed U.S. teachers on many dimensions of mathematics teaching quality, the higher achievement of Chinese students was primarily driven by geometry instruction: In geometry instruction, the study showed that Chinese students had more opportunities than U.S. students to develop their conceptual understanding and to make explicit connections between representations and methods for solving problems.

In answer to our second and third research questions, “What were the patterns of student achievement among the mathematics subscales?” and “Was there a group of students who demonstrated specific difficulties in geometry only?” we employed LPA to classify the whole student population into different groups according to the patterns of performance across the three mathematics domains. We found that students with the lowest Geometry subscale scores typically also had the lowest achievement in other mathematics domains. When we analyzed the four-, five-, and six-class models, we found a pattern that showed the more specifically and the greater number of groups into which we categorized the students, the more likely we were to find a group of students who showed the lowest scores on all three subscales, with a broader gap between geometry and the numerical mathematics domains. That is, students’ geometry performance was consistent with their performance in the non-geometry domains, and students with the lowest geometry performance was most likely to be found in the group of students with the poorest overall mathematics performance. That is, the more problems the students displayed with mathematics in general, the greater difficulties the students were likely to manifest in geometry performance.

In sum, using the TIMSS fourth-grade mathematics data, we did identify a group of students (i.e., the lowest performing class in each of the three models) who showed poorer performance in geometry than in other mathematics areas; however, these students neither reached the average or above-average benchmarks in other mathematics domains. Generally speaking, students with the lowest mean geometry scores also had very low mean scores in other mathematics areas, whereas students with high mean scores in other mathematics subscales also had high mean scores in

geometry. The students who performed poorly in the numerical areas tended to have even poorer performance in geometry, but such a discrepancy did not exist with the high-performing and average-performing students.

Results contributed to the literature by showing that geometry performance as measured by a standardized assessment is not independent of performance in other mathematics areas, and likewise, students’ poor performance in geometry does not stand alone from their difficulties in other mathematics areas. Results supported the theoretical proposition that performance in different mathematics domains is intercorrelated and reflects mutual underlying cognitive abilities (Peng et al., 2018). A student with poor spatial abilities may manifest difficulties in many mathematics areas (e.g., story problems; van Garderen, 2006) and other STEM areas (Wai et al., 2009) rather than in geometry only. Although it sounds plausible to categorize a group of students who demonstrate difficulties only in geometry due to poor spatial abilities, our results from the TIMSS 2011 data did not find evidence to support this hypothesis.

To answer the fourth research question, “Which demographic variables contributed to students’ classification in the group with geometry difficulties?” we examined the influences of various covariates on students’ probability to be classified into the geometry difficulty group, and found that being White, English-speaking at home, and liking mathematics and science significantly decreased the students’ probability to be classified into the geometry difficulty group. In contrast, being non-White and disliking mathematics and science significantly increased the chance to be classified into the lowest achieving group with particular geometry difficulties. These sociocultural-dependent predictors again indicated that geometry achievement is significantly influenced by many non-cognitive factors, such as ethnicity, culture, language, and interest.

It has been well established that culture and education play an important role in students’ mathematics achievement (Agirdag et al., 2011; Martin, 2000; Starkey & Klein, 2008). Although intuitive geometry is considered to be culture-free and to rely on cognitive abilities (Giofrè et al., 2013), the results of the present study indicated that as early as fourth grade, students’ geometry achievement has been significantly influenced by many educational and cultural factors. Fourth-grade geometry in the TIMSS was primarily about basic spatial manipulation tasks; however, even these so-called “intuitive geometry” tasks that seem to be cognitive-based were significantly affected by students’ cultural and educational backgrounds. Echoing previous research, a student’s social and cultural background can influence his or her academic achievement because of many factors, including parental involvement (Lee & Bowen, 2006; Mau, 1997), self-concept (Chiu & Klassen, 2010), school resources (Roscigno & Ainsworth-Darnell,

1999), social history context (Martin, 2000), and community forces (Martin, 2000).

Previous literature reported a small effect size of male superiority in geometry performance (Hyde et al., 1990). In this study, being a girl significantly increased students' probability to be included in an average-achieving class, whereas being a boy significantly increased the chance to be classified into either the highest achieving class or the lowest achieving class. Previous literature repeatedly demonstrated that spatial ability is the only cognitive domain that shows the most robust gender differences favoring males, especially in the area of spatial rotations (Feng et al., 2007; Terlecki & Newcombe, 2005); however, the results of this study suggested that although males were more likely to show the highest achievement scores in geometry, they also had a greater probability of showing the lowest achievement scores in geometry, suggesting that geometry achievement is not completely dependent on spatial abilities but is a function of multiple cognitive and non-cognitive factors.

### *Limitations and Future Research*

The present study has two major limitations. First, we analyzed fourth graders' TIMSS 2011 Geometry subscale scores to identify students with difficulties in geometry. The students who were classified into the low-achieving group in this study performed poorly on the TIMSS 2011 assessment and did not meet the Low International Benchmark. However, we should be cautious when reaching any conclusions about labeling these students with "Learning Disabilities (LD)" because these low-achieving students were not necessarily students with learning disabilities. Some students do well in class activities but are not good at taking standardized tests (Kearns, 2011), and it is widely recommended (Nitko & Brookhart, 2014) to adopt multiple measures (i.e., teachers' reports, classroom observations, standardized tests, curriculum-based assessment) to ensure the validity to refer a child to Tier 2 interventions. While many of the lowest performing group in this study may have geometry learning difficulties or learning disabilities, TIMSS does not provide data regarding students' special education status. Future research is warranted to examine patterns of students' abilities in different mathematics areas among students who have been officially diagnosed with mathematics learning disabilities.

Second, we need to be cautious regarding conclusions about whether there is a group of students who achieve low scores only in geometry but achieve average or above-average scores in other mathematics domains. Typically, LCA or LPA is carried out in an exploratory manner in which a strong a priori hypothesis does not exist regarding the number or nature of the latent classes underlying the data (Hojtink, 2001). For a future study, when there is a well-developed theory to hypothesize the classes of students, a

more confirmatory approach that allows for testing specific hypotheses will be used to verify the existence of hypothesized classes.

### *Implications for Educational Practice*

Although this study does not provide direct implications for educators as to how to provide effective interventions to improve geometry performance of students with difficulties solving geometry problems, results of this study portrayed a larger picture of the geometry performance of U.S. fourth graders and the patterns of U.S. students' achievement in three mathematics domains assessed by TIMSS 2011. First, the significantly lower performance in geometry than in the other two mathematics domains among U.S. students overall signifies a strong need for greater attention from policy makers and educational practitioners. In particular, special attention needs to be paid to students who are experiencing mathematics learning difficulties, as the results of this study showed that students with poor achievement in numerical domains tended to have even poorer performance in geometry. Given that geometry topics comprise only a relatively small part of the mathematics curriculum that is required to be taught to elementary school students, teachers and parents often do not address the importance of providing early geometry intervention to students who have mathematics learning disabilities or who are at risk. In fact, the results of this study suggested that students with overall difficulties in mathematics tend to experience greater difficulties in geometry than in other mathematics areas; thus, providing intensive prevention and intervention focusing on geometry subjects is warranted.

This study also sheds light on the importance of teachers and policy makers understanding students' geometry performance from a sociocultural perspective. Many STEM areas, including mathematics and geometry, are not solely determined by students' cognitive abilities or personal efforts. Even for tasks that are considered "intuitive geometry," students' performance can be influenced by many sociocultural factors. In comparison with the rich literature and resources of mathematics remediation programs that help students with mathematics difficulties to learn using cognitive approaches, less research and practice exist in terms of how to address students' mathematics learning problems from a sociocultural perspective (Waite, 2017).

### **Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### **Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article:

This work was supported by the Open Fund of National Higher Education Quality Monitoring Data Center (Higher Education Research Institute), Sun Yat-sen University (Grant Number: M1903), and also supported by the Key Nurturing Program for Young Teachers, Sun Yat-sen University (Grant Number: 18wkzd14).

## ORCID iD

Jinsong Chen  <https://orcid.org/0000-0002-0157-5469>

## References

- Agirdag, O., Van Houtte, M., & Van Avermaet, P. (2011). Why does the ethnic and socio-economic composition of schools influence math achievement? The role of sense of futility and futility culture. *European Sociological Review*, 28(3), 366–378.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716–723.
- Aydin, U., & UbuZ, B. (2010). Structural model of metacognition and knowledge of geometry. *Learning and Individual Differences*, 20(5), 436–445.
- Bailey, M., Taasooobshirazi, G., & Carr, M. (2014). A multivariate model of achievement in geometry. *The Journal of Educational Research*, 107(6), 440–461.
- Bartelet, D., Ansari, D., Vaessen, A., & Blomert, L. (2014). Cognitive subtypes of mathematics learning difficulties in primary education. *Research in Developmental Disabilities*, 35(3), 657–670.
- Battista, M. T. (1990). Spatial visualization and gender differences in high school geometry. *Journal for Research in Mathematics Education*, 21(1), 47–60.
- Bizarro, M., Giofrè, D., Girelli, L., & Cornoldi, C. (2018). Arithmetic, working memory, and visuospatial imagery abilities in children with poor geometric learning. *Learning and Individual Differences*, 62, 79–88.
- Bokosmaty, S., Sweller, J., & Kalyuga, S. (2015). Learning geometry problem solving by studying worked examples: Effects of learner guidance and expertise. *American Educational Research Journal*, 52(2), 307–333.
- Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, 13(2), 195–212.
- Chaurasia, A., & Harel, O. (2012). Using AIC in multiple linear regression framework with multiply imputed data. *Health Services and Outcomes Research Methodology*, 12(2–3), 219–233.
- Chiu, M. M., & Klassen, R. M. (2010). Relations of mathematics self-concept and its calibration with mathematics achievement: Cultural differences among fifteen-year-olds in 34 countries. *Learning and Instruction*, 20(1), 2–17.
- Clements, D. H., Battista, M. T., Sarama, J., & Swaminathan, S. (1997). Development of students' spatial thinking in a unit on geometric motions and area. *The Elementary School Journal*, 98(2), 171–186.
- Clements, D. H., & Sarama, J. (2011). Early childhood teacher education: The case of geometry. *Journal of Mathematics Teacher Education*, 14(2), 133–148.
- Dawkins, P. C. (2015). Students' strategies for assessing mathematical disjunctions. In *North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 339–346). Michigan State University. <https://files.eric.ed.gov/fulltext/ED584215.pdf>
- Feng, J., Spence, I., & Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychological Science*, 18(10), 850–855.
- Foy, P., Arora, A., & Stanco, G. M. (2013). *TIMSS 2011 user guide for the international database*. TIMSS & PIRLS International Study Center.
- Geary, D. C., & Hoard, M. K. (2005). Learning disabilities in arithmetic and mathematics: Theoretical and empirical perspectives. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 253–268). Psychology Press.
- Giofrè, D., Mammarella, I. C., & Cornoldi, C. (2014). The relationship among geometry, working memory, and intelligence in children. *Journal of Experimental Child Psychology*, 123, 112–128.
- Giofrè, D., Mammarella, I. C., Ronconi, L., & Cornoldi, C. (2013). Visuospatial working memory in intuitive geometry, and in academic achievement in geometry. *Learning and Individual Differences*, 23, 114–122.
- Gonzales, P., Guzmán, J. C., Partelow, L., Pahlke, E., Jocelyn, L., Kastberg, D., & Williams, T. (2004). *Highlights from the Trends in International Mathematics and Science Study (TIMSS) 2003* (NCES 2005-005). U.S. Department of Education.
- Hojtink, H. (2001). Confirmatory latent class analysis: Model selection using Bayes factors and (pseudo) likelihood ratio statistics. *Multivariate Behavioral Research*, 36(4), 563–588.
- Hsi, S., Linn, M. C., & Bell, J. E. (1997). The role of spatial reasoning in engineering and the design of spatial instruction. *Journal of Engineering Education*, 86(2), 151–158.
- Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, 107(2), 139–155.
- Karagiannakis, G., Baccaglini-Frank, A., & Papadatos, Y. (2014). Mathematical learning difficulties subtypes classification. *Frontiers in Human Neuroscience*, 8, Article 57.
- Kearns, L. L. (2011). High-stakes standardized testing & marginalized youth: An examination of the impact on those who fail. *Canadian Journal of Education*, 34(2), 112–130.
- Kern, E., & Henrick, E. (2016, July). *Investigating middle grades mathematics teaching in the United States and China: A cross-national comparison of instructional quality in urban districts* [Paper presentation]. 2016 International Congress of Mathematics Education Conference (ICME), Hamburg, Germany.
- Kyttälä, M., & Lehto, J. E. (2008). Some factors underlying mathematical performance: The role of visuospatial working memory and non-verbal intelligence. *European Journal of Psychology of Education*, 23(1), 77–94.
- Lawson, M. J., & Chinnappan, M. (1994). Generative activity during geometry problem solving: Comparison of the performance of high-achieving and low-achieving high school students. *Cognition and Instruction*, 12(1), 61–93.

- Lee, J. S., & Bowen, N. K. (2006). Parent involvement, cultural capital, and the achievement gap among elementary school children. *American Educational Research Journal*, 43(2), 193–218.
- Lubke, G. H., & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*, 10(1), 21–39.
- Malhotra, P., Coulthard, E. J., & Husain, M. (2009). Role of right posterior parietal cortex in maintaining attention to spatial locations over time. *Brain*, 132(3), 645–660.
- Martin, D. B. (2000). *Mathematics success and failure among African-American youth: The roles of sociohistorical context, community forces, school influence, and individual agency*. Routledge.
- Mau, W. C. (1997). Parental influences on the high school students' academic achievement: A comparison of Asian immigrants, Asian Americans, and White Americans. *Psychology in the Schools*, 34(3), 267–277.
- Muthén, B. O., & Satorra, A. (1995). Technical aspects of Muthén's LISCOMP approach to estimation of latent variable relations with a comprehensive measurement model. *Psychometrika*, 60(4), 489–503.
- Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (8th ed.).
- National Center for Learning Disabilities. (2006). *Dyscalculia*. LD Online. <http://www.ldonline.org>
- Nichols, J. D. (1996). Cooperative learning: A motivational tool to enhance student persistence, self-regulation, and efforts to please teachers and parents. *Educational Research and Evaluation*, 2(3), 246–260.
- Nitko, A. J., & Brookhart, S. M. (2014). *Educational assessment of students* (7th ed.). Pearson Education.
- Organisation for Economic Co-Operation and Development. (2014). *Education at a glance 2014: OECD indicators*. <http://www.oecd.org/education/Education-at-a-Glance-2014.pdf>
- Passolunghi, M. C., & Mammarella, I. C. (2012). Selective spatial working memory impairment in a group of children with mathematics learning disabilities and poor problem-solving skills. *Journal of Learning Disabilities*, 45(4), 341–350.
- Passolunghi, M. C., Mammarella, I. C., & Altoè, G. (2008). Cognitive abilities as precursors of the early acquisition of mathematical skills during first through second grades. *Developmental Neuropsychology*, 33(3), 229–250.
- Peng, P., Namkung, J., Barnes, M., & Sun, C. (2016). A meta-analysis of mathematics and working memory: Moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *Journal of Educational Psychology*, 108(4), 455–473.
- Peng, P., Wang, C., & Namkung, J. (2018). Understanding the cognition related to mathematics difficulties: A meta-analysis on the cognitive deficit profiles and the bottleneck theory. *Review of Educational Research*, 88(3), 434–476.
- Provasnik, S., Kastberg, D., Ferraro, D., Lemanski, N., Roey, S., & Jenkins, F. (2012). *Highlights from TIMSS 2011: Mathematics and science achievement of U.S. fourth- and eighth-grade students in an international context* (NCES 2013-009). National Center for Education Statistics.
- Provasnik, S., Malley, L., Stephens, M., Landeros, K., Perkins, R., & Tang, J. H. (2016). *Highlights from TIMSS and TIMSS advanced 2015: Mathematics and science achievement of U.S. students in Grades 4 and 8 and in advanced courses at the end of high school in an international context* (NCES 2017-002). National Center for Education Statistics, U.S. Department of Education.
- Roscigno, V. J., & Ainsworth-Darnell, J. W. (1999). Race, cultural capital, and educational resources: Persistent inequalities and achievement returns. *Sociology of Education*, 72, 158–178.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7(2), 147–177.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6, 461–464.
- Sinclair, N. (2008). *Research in mathematics education: The history of the geometry curriculum in the United States* (B. J. Dougherty, Ed.). Information Age.
- Spelke, E., Lee, S. A., & Izard, V. (2010). Beyond core knowledge: Natural geometry. *Cognitive Science*, 34(5), 863–884.
- Starkey, P., & Klein, A. (2008). Sociocultural influences on young children's mathematical knowledge. In O. N. Saracho & B. Spodek (Eds.), *Contemporary perspectives on mathematics in early childhood education* (pp. 253–276). Information Age.
- Terlecki, M. S., & Newcombe, N. S. (2005). How important is the digital divide? The relation of computer and videogame usage to gender differences in mental rotation ability. *Sex Roles*, 53(5–6), 433–441.
- van Garderen, D. (2006). Spatial visualization, visual imagery, and mathematical problem solving of students with varying abilities. *Journal of Learning Disabilities*, 39(6), 496–506.
- Vermunt, J. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, 18(4), 450–469.
- von Davier, M., Gonzalez, E., & Mislevy, R. J. (2009). What are plausible values and why are they useful? In M. von Davier & D. Hastedt (Eds.), *IERI Monograph series: Issues and methodologies in large-scale assessment* (Vol. 2, pp. 9–36). Educational Testing Service.
- Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, 101(4), 817–835.
- Wai, J., Lubinski, D., Benbow, C. P., & Steiger, J. H. (2010). Accomplishment in science, technology, engineering, and mathematics (STEM) and its relation to STEM educational dose: A 25-year longitudinal study. *Journal of Educational Psychology*, 102(4), 860–871.
- Waite, S. (Ed.). (2017). *Children learning outside the classroom: From birth to eleven*. Sage.