

Assessing Student Learning in Number Sense Using Model-Strategy-Application with Reasoning Approach

Shuhua An

California State University, Long Beach, USA

Zhonghe Wu

National University, USA

This paper addressed the challenges in urban learning by using a structured form of assessment with multiple aspects of the Model-Strategy-Application with Reasoning in assessing student learning in number sense standards among linguistically and culturally diverse classrooms. The participants were 29 third graders in an elementary school in Southern California. The results showed that there were significant differences in student learning in model, strategy, application, and reasoning. The students had the highest scores in modeling, and the lowest scores in application. The higher scores in modeling and lower scores in application and reasoning were also shown in diverse groups of gender and language. The study suggested using the MSAR approach to assess and support ELL student learning on application and reasoning.

Keyword: Assessment, model, strategy, application, reasoning

One of the key instructional components for effective elementary mathematics instruction is assessment. The National Research Council (1993) specifies mathematics assessment as “a way of measuring what students know and of expressing what students should learn” (p.1). Assessment is not only “a fact-finding activity” that describes conditions that exist at a particular time (Best & Kahn, 2006), but also should support the learning of important mathematics and furnish useful information to both teachers and students (National Council of Teachers of Mathematics [NCTM], 1995). The purpose of assessment addressed by NCTM (1995) - gathering evidence and making inferences from such evidence for a variety of purposes has played a vital role in mathematics assessment in the last two decades. Recent developments in the new Common Core State Standards for Mathematics (CCSSM) in the United States have further heightened the clear purpose of assessment in supporting and enhancing student learning in the *Mathematics Framework for California Public Schools* (California Department of Education [CED], 2016).

According to the NCTM Assessment Standards (1995), an equitable assessment should provide multiple forms of assessment that not only addresses the breadth of important mathematical content, but also provides all

students with an opportunity to demonstrate their mathematical power. Teachers are encouraged to use a variety of forms of assessment (Brookhart, 2003; Moss, 2003) because learning is multidimensional and cannot be adequately measured by a single instrument (Brookhart, 2003; Suurtamm et al., 2010). One of the important recent shifts is that assessment in mathematics should go beyond focusing on assessing algorithm or procedure; it must also address assessing students' mathematical understandings (NCTM, 1995; Suurtamm et al., 2010) with various forms such as answers and solutions, arguments and explanations, diagrams and mathematical models to meet needs of diverse students (CDE, 2015).

The purpose of this study was to address the challenges in urban learning by using a structured form of assessment with multiple aspects of the Model-Strategy-Application with Reasoning (MSAR) in assessing student learning in number sense standards among linguistically and culturally diverse classrooms. The MSAR is defined as developing students' conceptual understanding using various visual Models, building procedural fluency with different computational Strategies, and developing problem solving skills with real world Applications (An & Wu, 2014), while focusing on fostering reasoning skills throughout the three components of the MSAR. This project was aimed at investigating the following research questions:

1. Are there any significant differences in student learning between model, strategy, application, and reasoning reflected in student MSAR worksheets?
2. Are there any significant differences in student learning between model, strategy, application, and reasoning reflected in student MSAR worksheets between genders and between language groups?
3. What models and strategies for computations and applications in word problems are used frequently by children to demonstrate conceptual understanding, procedural fluency, and problem solving in number sense?
4. What models and strategies for computations and applications in word problems are used frequently by diverse groups to demonstrate conceptual understanding, procedural fluency, and problem solving in number sense?

Theoretical Framework

Assessing Understanding of Arithmetic and Number Systems Based on CCSSMPs

The new CCSS for Mathematical Practice (CCSSMP) standards describe a variety of expertise that mathematics educators at all levels should seek to develop in their students (National Governors Association Center for Best Practices & Council of Chief State School Officers [NGACBP & CCSSO], 2010). These CCSSMPs require development of a deep

understanding and reasoning in mathematics with high-level mathematics thinking, reasoning, arguments, communication, and modeling skills. However, these CCSSMPs have posed challenges to classroom teachers on how to foster students' deep understanding, sound reasoning, and argument skills in mathematics classrooms. In the recently released (2015) California Assessment of Student Performance and Progress results, more than 40% of elementary students in grades 3-5 scored below standards in the areas of Concepts and Procedures, followed by more than 38% below standards in Problem Solving/Modeling and Data Analysis. The area of Communicating Reasoning had more than 39% of students who did not meet standards at grades 4 and 5 (CED, 2015). Such data show an urgent need to develop an effective approach to support classroom teachers in improving elementary children's learning in arithmetic and number systems, as they are major standards in CCSSM.

Assessing Understanding of Arithmetic and Number Systems Using Mathematics Tasks

The effective assessment task should be an open-ended task, and it should include the three components recommended by the National Research Council (1993): 1) Involving significant mathematics; 2) Enhancing mathematics learning and supporting good instructional practice; 3) Supporting every student's opportunity to learn important mathematics. Mathematical tasks are important vehicles for classroom instruction, which aim to enhance students' learning. Current research indicates that if students are given opportunities to work on problem solving skills in rich mathematical tasks, they enjoy learning mathematics more and become more confident in studying mathematics (Hewson, 2011). Rich mathematical tasks can also develop students' capacity in reasoning and argument (Mok & Kaur, 2006; Smith, Bill, & Hughes, 2008). In order to achieve quality mathematics instruction, the role of mathematical tasks to stimulate students' cognitive processes is crucial (Hiebert & Wearne, 1993). One example of rich mathematical tasks in arithmetic and number systems demonstrated in Wu and An's study (2016) is using the Model-Strategy-Application with Reasoning (MSAR) in teaching mathematics and assessing the effects of children's mathematics learning in diverse classrooms. The results show that applying the MSAR approach in teaching and assessing student learning helped diverse students at grades 2, 3, and 5 understand arithmetic and number systems from drawing visual models, develop procedural fluency using various computational strategies, and build competence in problem solving in real world applications (Wu & An, 2016).

According to Wu and An (2016), the four components of the MSAR are aligned with four categories of CCSSMP standards in modeling and using tools, seeing the structure and generalizing, connecting to real word application, and reasoning and explaining as addressed in California

Mathematics Framework (CED, 2016). The MSAR also includes the essential components of five indicators of student mathematics proficiency by RAND (2003) and the National Research Council (2001). It is supported by the NCTM Process Standards (2000) and aligned with the Guiding Principles of California Mathematics Conceptual Framework that calls for a balance within mathematics by focusing on the three key components -- conceptual understanding, computational skills, and problem solving, for more effective math programs (CED, 2006). The new California CCSSM also calls to have students experience a balanced approach to instruction and learning that supports the three key components of the California Mathematics Framework (CED, 2016). The eight CCSSMP Standards rest on these important “processes and proficiencies” which have longstanding importance in math education (NGACBP & CCSSO, 2010). Therefore, the MSAR model is a fundamental framework for teachers to teach CCSSM and for students to learn CCSSM effectively and proficiently. Importantly, it is an appropriate assessment tool to gauge student learning in multiple aspects (Wu & An, 2016).

Method

Site and Subjects

The study was conducted at an elementary school in Southern California in 2013 and 2014. The school is located in a low socioeconomic area that serves primarily neighborhood families. On the annual standardized tests in mathematics, the school had 18% of students who were advanced proficient in mathematics, while 37% were simply proficient, and 45% not proficient. The entire school has 846 enrolled students, and 88.2% of these students are socioeconomically disadvantaged; 6.5% of the students have a learning disability, and 65.8% of the students are English Learners.

The participants were 29 students from a third-grade classroom. About 80% of the students were Hispanic, 6% were African American, 6% were White, and 8% were other; 17 of the students were English Language Learners (ELL) and 12 spoke English only (EO); there were 16 girls and 13 boys. On the annual standardized test in mathematics, approximately 30% students were Advanced Proficient, 50% were proficient, and 20% were partially proficient (see Table 1). The classroom teacher has been teaching the third grade for the past 5 years. However, before teaching the third grade, she worked as a second-grade teacher for 7 years, kindergarten for 5 years, and 5 years with a third-grade literacy class.

Data Collection

The participants were given five MSAR assessments at the beginning of fall 2013 to assess student prior knowledge on conceptual understanding, procedural fluency, and problem solving in forms of model, strategy, and

application with an explanation for each area. Each MSAR included one content skill in five content areas of Geometry, Number Sense Base Ten, Number Sense Fractions, Operations and Critical Thinking, and Measurement for a larger study of the MSAR project. This study examined student knowledge and skills in a Number Sense Base Ten standard in a pre-MSAR assessment.

Table 1
Demographic Information of Participants

	Gender		Language			Achievement	
	Girl	Boy	ELL	EO	A	P	PP
Demographic	55	45	59	41	30	50	20
%	Ethnicity						
	Hispanic		African American		White	Other	
			6		6	8	

Note: A = Advanced Proficient; P = Proficient; PP = Partially Proficient

Instrument

The five MSAR's covered the five domains: Geometry, Number Sense Base Ten, Number Sense Fractions, Operations and Critical Thinking, and Measurement and Data - one for each content area. Each MSAR had four parts: Modeling, Strategy, and Application, along with an explanation for each part. The current study only used the MSAR on a Number Sense Base Ten standard (see Figure 1) and it was created by a group of classroom teachers for a larger study of the MSAR.

The student MSAR worksheets were evaluated by the author, who also developed the MSAR rubrics that have five levels (see Table 2). Since all MSAR tasks were created based on the CCSSM at the third grade level, it ensured the content validity of the MSAR tasks. The MSAR tasks were tested for reliability in this study. The Cronbach's Alpha for the four areas of the MSAR instrument was .852, which indicates a high level of internal consistency (Cronbach, 1951) for the four areas of the MSAR in this study.

Data Analysis

Data were analyzed both quantitatively and qualitatively. The One-Way Within Subject Analysis of Variance (ANOVA) was used to find the differences in student learning in the four areas of the MSAR for research question 1. A dependent samples t-test was performed to identify significant pairs among four areas if a significant difference was found. The One-Between-One-Within Subject Analysis of Variance (ANOVA) was used to identify the differences in four areas between gender and language groups for research question 2. Research questions 3 and 4 were analyzed qualitatively

by examining student MSAR worksheets, identifying patterns, and comparing the differences in student demonstrations in each area of the MSAR.

MSAR

Jill made 5 rows of blocks, with each row containing 7 blocks. How many blocks did Jill have altogether?

Modeling	Strategies of Computation	Creating and solving a similar word problem
Explain why:	Explain why:	Explain why:

Figure 1. MSAR example.

Table 2
The MSAR Rubrics

Level	Modeling	Strategy	Application	Reasoning
0	No model	No computation	No word problem written	No explanation
1	Model completely inappropriate	Many computational errors	Word problem impossible to follow	Inappropriate explanation
2	Appropriate model used, but either not fully demonstrated, or possibly based on the operation only, limited process of conceptual development	Few computational errors, but followed rules and formulas on computations (routine way), or only by trial and error	Word problem attempted, but difficult to understand	Incomplete explanation
3	Appropriate model used, and the process of modeling demonstrated	No computational errors, but solved problem by routine way or only by trial and error	Student's word problem fairly clear, but not appropriate or not connected to real life application	Appropriate explanation
4	Model used highly efficient and meaningful, revealing and comprehensive understanding	No computational errors and used a flexible or creative strategy in computation, revealing complete understanding of solving	Word problem very clear, appropriate, and connected to real life application	The process of MSA explained with academic language and deep understanding

Results

Difference in Four Areas of the MSAR in Number Sense

The results of the One-Way Within Subject Analysis of Variance (ANOVA) in Table 3 show that there were significant differences in mean scores between model, strategy, application, and reasoning, Sphericity Assumed $F(3, 75) = 3.072, p = 0.033 < 0.05$, partial $\eta^2 = .109$.

Table 3
One-Way Within Subject Analysis of Variance Results

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
MSAR	Sphericity Assumed	5.308	3	1.769	3.072	.033	.109
Error)	Sphericity Assumed	43.192	75	.576			

To assess which of the four areas differed from one another, dependent samples t-tests were conducted. The results from Tables 4 and 5 indicate that three pairs were significantly different from each other: the model scores ($M = 2.033, SD = 1.3767$) were significantly higher than the strategy scores ($M = 1.633, SD = 1.3515$), $t(29) = 2.449, p < .05$, the application scores ($M = 1.533, SD = 1.3578$), $t(29) = 2.236, p < .05$, and the reasoning scores ($M = 1.846, SD = .9249$), $t(29) = 3.934, p < .05$.

Table 4
Paired Samples Test for Four Areas of the MSAR

		Paired Differences		95% Confidence Interval		t	Sig. (2-tailed)
		Mean	SD	Lower	Upper		
M NS - S NS		.40	.894	.066	.733	2.449	.021
M NS - A NS		.50	1.224	.042	.957	2.236	.033
M NS - R NS		.50	.648	.238	.761	3.934	.001

Table 5
Paired Samples Statistics for Four Areas of the MSAR

		Mean	N	SD	Std. Error Mean
Pair 1	Model	2.03	30	1.376	.2514
	Strategy	1.63		1.351	.2467
Pair 2	Model	2.03	30	1.376	.2514
	Application	1.53		1.357	.2479
Pair 3	Model	2.34	26	1.198	.2350
	Reason	1.84		.924	.1814

Figure 2 confirms these significant differences between the four areas of the MSAR: Students had the highest scores in the model, and the lowest scores in the application.

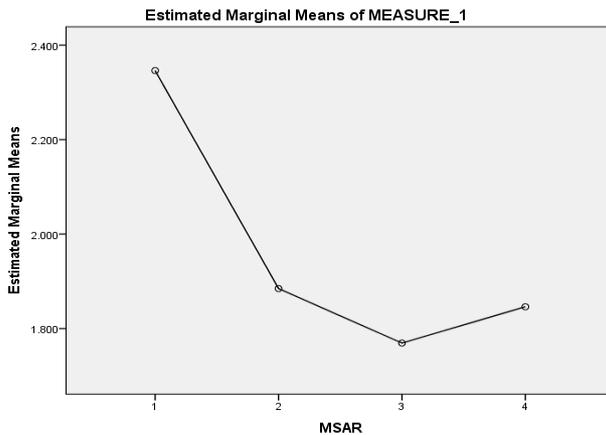


Figure 2. Difference between the four areas of the MSAR.

Difference in the Four Areas of the MSAR between Gender Groups and between Language Groups

The results of the One-Between-One-Within Subject Analysis of Variance (ANOVA) for answering question 2 were summarized in Tables 6 and 7. There was no significant difference in the four areas of the MSAR between gender groups, $F(1, 24) = 1.448$, $p = .241 > 0.05$, partial $\eta^2 = .057$, and between language groups, $F(1, 24) = .948$, $p = .340 > 0.05$, partial $\eta^2 = .038$.

Table 6
Tests of Between-Subjects Effects in Gender Groups

Source	Type III Sum of Squares	Mean Square	F	Sig.	Partial Eta Squared
Intercept	400.154	400.154	104.607	.000	.813
Gender	5.538	5.538	1.448	.241	.057
Error	91.808	3.825			

Table 7
Tests of Between-Subjects Effects between Language Groups

Source	Type III Sum of Squares	Mean Square	F	Sig.	Partial Eta Squared
Intercept	402.470	402.470	103.147	.000	.811
Language	3.701	3.701	.948	.340	.038
Error	93.645	3.902			

Although the differences in the four areas of the MSAR were not at a significant level between diverse groups, Figures 3 and 4 identify some patterns of differences in the four areas between diverse groups. For the gender group, boys had higher scores in all four areas than girls; both girls and boys had the highest scores in the model, but girls had the lowest scores in the application, and boys had the lowest scores in the reasoning.

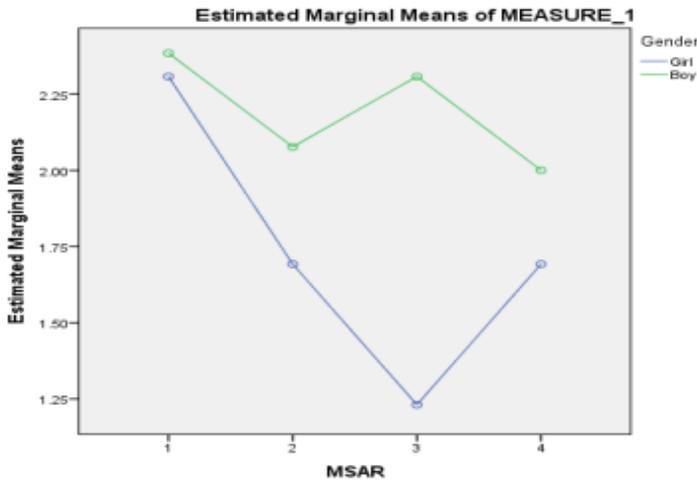


Figure 3. Difference in MSAR between gender groups.

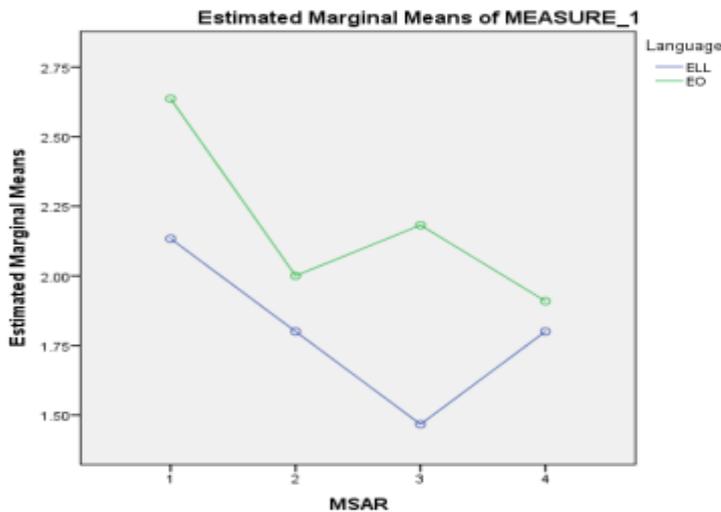


Figure 4. Difference in MSAR between language groups.

For the language groups, EO students had higher scores in all four areas than ELL learners; both groups had their highest scores in the model, but ELLs had the lowest scores in the application, and EO students had the lowest scores in the reasoning.

Common Models, Strategies, and Applications Used by Students

Tables 8 and 9 show the results of analyzing student worksheets in the four areas of the MSAR. Table 8 shows that multiple visual representations were used by the third graders to demonstrate their conceptual understanding in number sense: 14 students used Block Arrays with 10 being correct, and 5 used Set Models to show their understanding with 3 correct. However, 3 students used additional models with 5 blocks adding 7 blocks. Of all models shown, about 66% were correct.

For the computational strategies, about 12% of students used Repeated Addition, 15% used Multiplication, and 7% used both Repeated Addition and Multiplication, however, more than 63% of students solved it using incorrect computational strategies, such as using two numbers related to 5 and 7: $5+7=11$, $5+2=7$, $5-7=2$, $5-2=3$; some combined repeated addition and regular addition with different numbers, such as $5+5+7=17$, $10+10+10+10+4=42$, $5+5+5+5+4+7=31$, $5+5+5+5+5+5+4=34$. Of all strategies used, only 35% were correct.

For the application, most students used the block and row ideas to create their Real Life word problems; some used different life situations. Of all word problems developed, 50% of word problems were correct. About 23% of students could not come up with a word problem. They either draw different Blocks and Rows (2 students) or direct put math sentences (4 students), such as $10 + 10 + 10 + 10 + 4 = 34$. Of all applications, only 38% were able to create correct word problems.

Table 8

Common Models, Strategies, and Application Used by Students

Model		Strategy		Application	
Type	Total (Correct)	Type	Total (Correct)	Type	Total (Correct)
Set	5 (3)	Repeated Addition	3 (3)	Real Life App	20 (10)
Array	5 (2)	Multiplication	4 (3)	Draw Block & Row	2 (1)
Grid	2 (2)	2 Strategies: ++ and x	2 (2)	Math App	4 (1)
Block	14 (10)	Repeated add w/ Diff	4		
Aarray					
2 Models	2	$5 + 7 = 12$	5		
		Other: $5-7=2$, $5-2, ..$	2		
		Add with + or	3		
		Explain			
		Multiply with x or	3		
		Explain)			

For the reasoning part, about 38% students were able to explain how they used visual representations to demonstrate their conceptual understanding with correct explanations; only 27% explained correctly the strategies they used, and 31% explained correctly their word problems in applications.

Table 9

Reasoning on Models, Strategies, and Application Used by Students									
	Model			Strategy			Application		
	Total	Correct	%	Total	Correct	%	Total	Correct	%
Reasoning	26	10	38	26	7	27	26	8	31

Types of Models, Strategies, and Applications Used by Diverse Groups

1. Type of Models Used by Diverse Groups

a. Models used by diverse students

Table 10 and Figure 5 show that more EO students (4) used the regular Array than ELL students (3); 10 (about 70%) of ELL learners used Block Array, compared to only 4 (less than 40%) of EO students using Block Array. For the gender groups, 5 girls used Set Model, while 0 boys used it.

Table 10

Common Models Used by Diverse Groups of Students

Model Type	Total	Girl	Boy	ELL	EO
Set	5	5	0	3	2
Array	5	2	3	1	4
Grid	2	1	1	1	1
BlockAarry	14	7	7	10	4
2 Models			1	1	

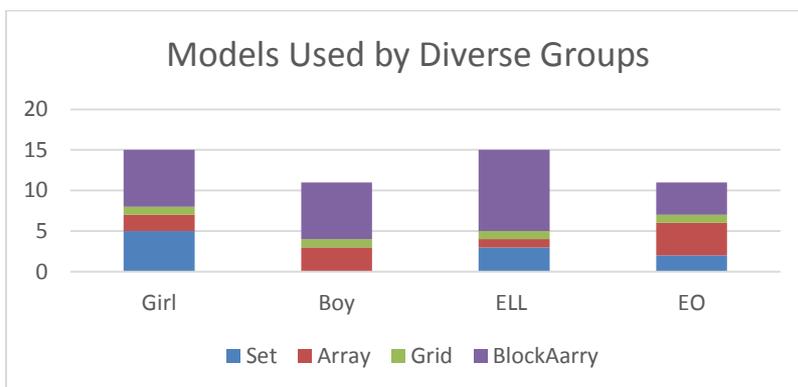
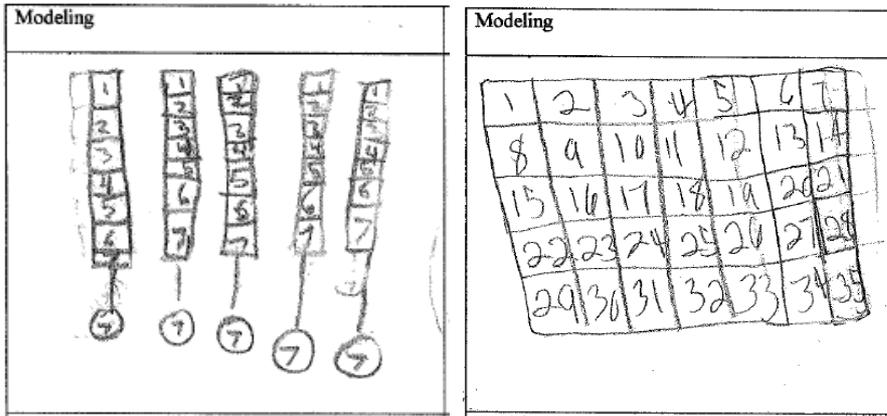


Figure 5. Models used by gender and language groups.

b. Example of models used by diverse students

For example, Alberto, a male ELL learner, used a Block Array method below in part a in Figure 6. Scott, a male EO learner, used a regular Array in part b in Figure 6.

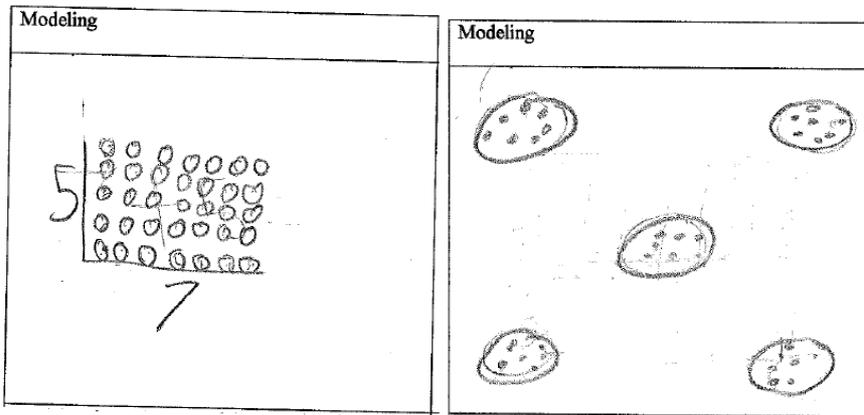


a. Block Array by Alberto

b. Array by Scott

Figure 6. Example of models used by language groups.

Robort, a male EO learner used a Grid Array Model (see part a in Figure 7); Cynthia, a female ELL learner used a Set Model (see part b in Figure 7).



a. Grid model by Robert

b. Set model by Cynthia

Figure 7. Example of models used by Gender groups.

2. Type of Strategies Used by Diverse Groups

a. Strategies used by diverse students

Table 11 shows that 8 computational strategies were used by the diverse groups (gender and language). Computational strategies 1-3 are correct strategies; strategy 4 is not an effective strategy; strategies 5 and 6 are

incorrect; strategy 7 is either explained or indicated by students with no actual computational procedures shown.

Table 11
Common Strategies Used by Diverse Groups of Students

Stragtey	#	Gender		Language	
		Girl	Bo	ELL	EO
1 Repeated Addition	3	1	2	1	2
2 Multiplication (5x7 or 7x5)	4	2	2	2	2
3 2ways: ++, x(5+, 7+, 5x 7), 14+14	2	1	1	2	
4 Repeated Addition with Diff #	4	3	1	1	3
5 5 + 7 = 12	5	3	2	3	2
6 Other: 5-7=2, 5-2, ..	2	2		1	1
7 Addition (Explain only)	3	1	2	3	
8 Multiplication (Explain only)	3	1	2	2	1
Total	26	14	12	15	11

Figure 8 shows that the same number of boys used Repeated Addition and Multiplication strategies, but a lesser number of girls used Repeated Addition. More girls used incorrect strategies 4, 5 and 6 than those in boy groups. But, more boys only indicated or explained their strategies of addition or multiplication, without showing their actual computations.

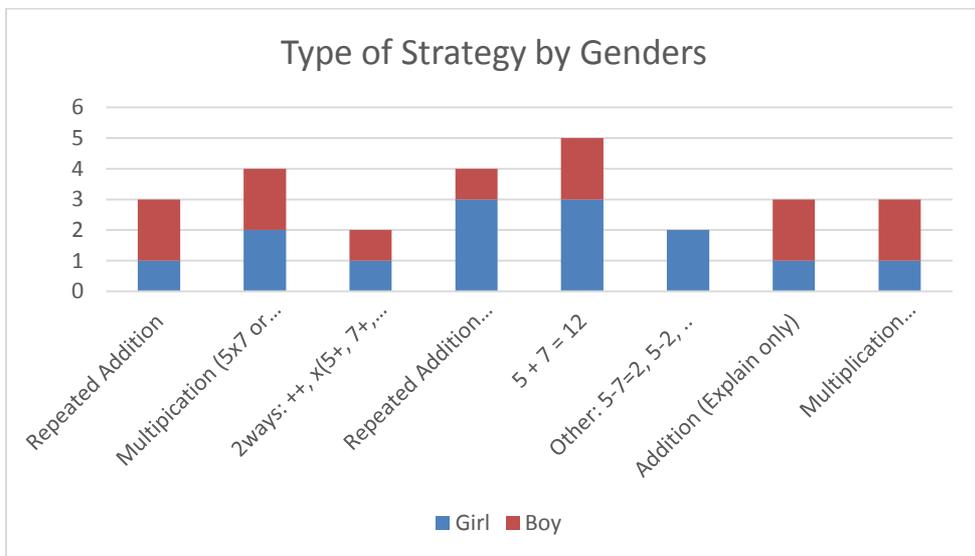


Figure 8. Type of strategies used by gender groups.

Results from Figure 9 show that the same number of students in both ELL and EO groups used multiplication, 5 x 7 or 7 x 5; more EOs used repeated addition, but their strategies were not effective. Only ELLs used

more than one type of strategy. Also, more ELLs only used x or + to indicate their computational strategies, not showing the actual computational procedures.

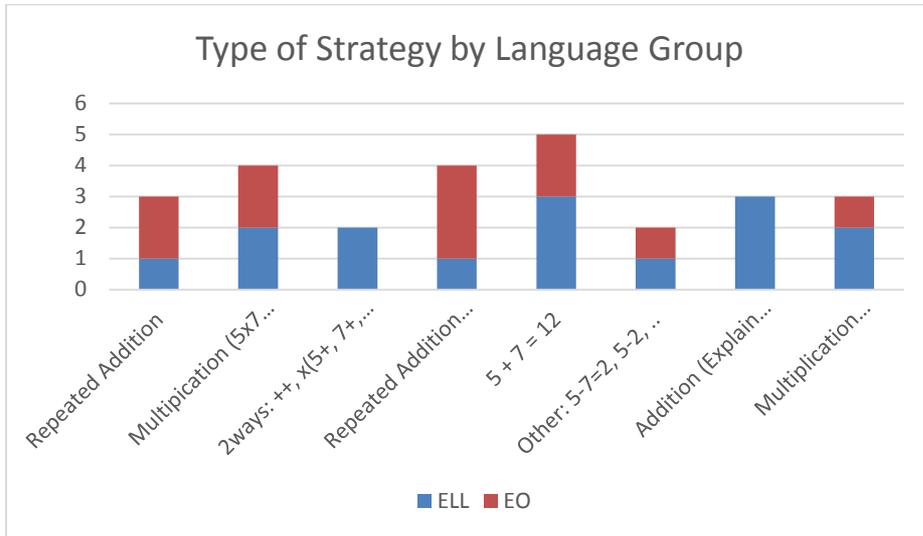
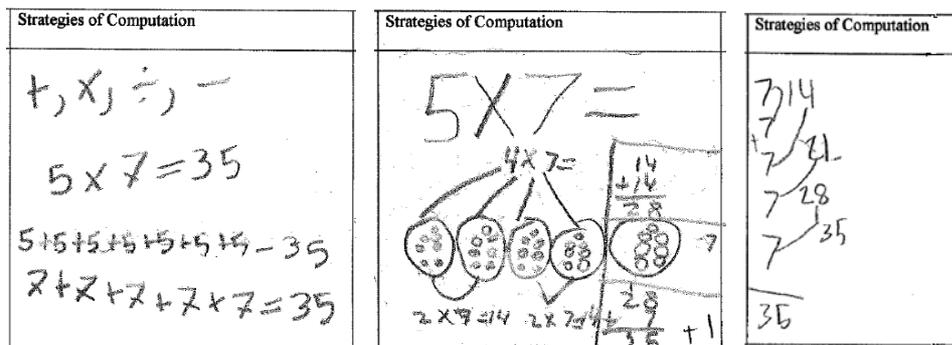


Figure 9. Type of strategy by language group.

b. Example of strategies used by diverse students.

Cynthia, a female ELL learner used two strategies in her computations: $5 \times 7 = 35$, $5 + 5 + 5 + 5 + 5 + 5 + 5 = 35$, and $7 + 7 + 7 + 7 + 7 = 35$ (see part a in Figure 10). Alberto, a male ELL learner, used addition first by grouping two sets of 7 two times, and then added the results of two 14s to get 28. Finally, he added the last set of 7 to 28 to get to 35 (see part b in Figure 10). Paul, a male EO student, used the repeated addition for all 7s first, and then add two 7s get 14, then adding another 7 to get to 21, and repeated this process, reaching 35.



a. Cynthia's strategy.

b. Alberto's strategy.

c. Paul's strategy.

Figure 10. Examples of strategies used by diverse students.

3. Type of Applications Used by Diverse Groups

a. Word problem application by diverse students

Table 12 and Figure 11 show the application problems used by diverse groups. The results indicate that more boys (11) were able to create real life math problems than girls (9); more ELLs were able to create real life math problems than EO students.

Table 12
Type of Application by Diverse Groups

Type of Application	Total	Gender		Language	
		Girl	Boy	ELL	EO
Real Life Application	20	9	11	11	9
Draw Block & Row	2	2		1	1
Only Math App	4	2	2	3	1
Total	26	13	13	15	11

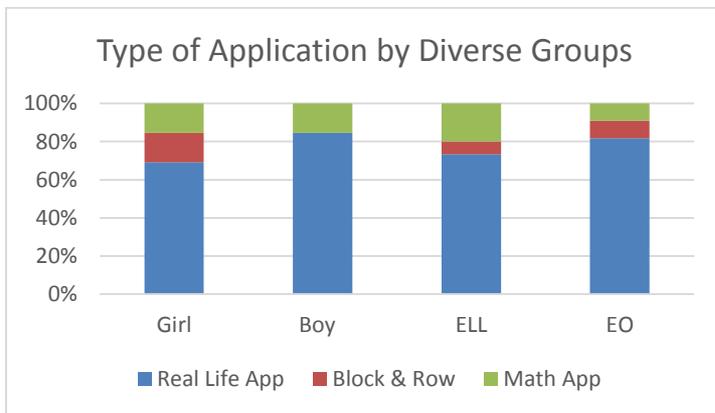


Figure 11. *Type of Application by Diverse Groups.*

b. Example of word problem application used by diverse students.

Christina, a female ELL student created a word problem, along with a set model in repeated addition to solve it: “Jack (had) 6 blocks in 7 rows. How many blocks did Jack have?” (See part a in Figure 12). Sandra, a female EO learner created a word problem which differed from blocks and rows: “Mr. Tan has 8 boxes of pizza. Each box has 5 slides. How many are in all?” (See part b in Figure 12).

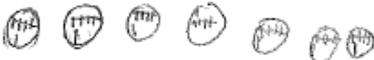
Creating and solving a similar word problem and solve it	Creating and solving a similar word problem and solve it
<p>Jack, 6 blocks in 7 rows. how many blocks did Jack have. $7 \times 6 = 42$ $\div 1$</p> 	<p>Mr. Tan has 8 boxes of pizza each box has 5 slices. how many is in all</p> <hr/> <p>$\begin{array}{r} \times 8 \\ 5 \\ \hline 40 \end{array}$ $\div 1$ 40 slices</p>

a. Christina's application.

b. Sandra's application.

Figure 12. Example of strategies used by language groups.

Tracy, a female EO learner developed a word problem as her error computational strategy $5+2=7$: "Jill need two more blocks to make seven" (see part a in Figure 13). Juan, a male ELL learner created a word problem as his error computational strategy $7+5=12$: "Arty has 5 games. He buys 11 more games. How many all together?" (see part b in Figure 13). These two students did not understand concepts of the original problem, as they mistakenly used addition instead of multiplication in their computational strategies and word problem applications.

Creating and solving a similar word problem and solve it	Creating and solving a similar word problem and solve it
<p>Jill need two more block to make Seven.</p> 	<p>Arty has 5 games then he buys 11 more games how many all together</p>

a. Tracy's word problem.

b. Juan's word problem.

Figure 13. Example of strategies used by gender groups.

Discussion

The results of this study show that there were significant differences in students learning between model, strategy, application, and reasoning. Overall, students did better work in the model and did poorly in the application and reasoning. The higher scores in the model compared to the other parts were also shown with diverse groups. For gender groups, both girl and boy groups had their highest scores in the model, but girls had the lowest scores in the application, and boys had the lowest scores in the reasoning. For language groups, both groups had their highest scores in the model, but ELLs had the lowest scores in the application, and EO students had the lowest scores in the reasoning.

The findings on the difficulty in the application and reasoning from this study corroborate the ideas of Bernardo (2005), who suggested that teachers should support students' learning in word problems because ELL learners' difficulties begin when they encounter word problems in a second language. One underlying objective of the MSAR was the application of mathematics to real life scenarios. The word problems used were selected to reflect student knowledge of real world experiences. Students were required to recreate a word problem that reflected the original word problem in the MSAR task or create a word problem based on a given mathematics computation problem. This allowed students to explore the word problems in a manner that encouraged more in-depth understanding of the word problem while applying mathematics to real life applications (Wu & An, 2016). In addition, it encouraged students to create word problems according to their experience (Wiest, 2000).

Another important finding was that both EO and ELL learners had lower scores in the reasoning. Mathematics reasoning is one of the main weaknesses of US students' learning as identified by international assessments such as 2012 PISA. Although CCSSMP2 and CCSSMP4 on reasoning posed challenges to classroom teachers on how to foster their students' sound reasoning and argumentation skills in mathematics classrooms, the MSAR model in this study provided diverse students an opportunity to show their reasoning behind their model, strategy, and application, which will enhance their ability in writing, reflecting, and justifying their mathematics reasoning.

The most interesting finding was that all diverse student groups had their highest scores in modeling, which confirms the MSAR as a powerful assessment tool that allows all students, especially ELL students, to draw visual representations or models to communicate their mathematics thinking and conceptual understanding (Wu, 2008). The present study confirmed previous findings and contributed additional evidence that added substantially to our understanding of the challenges and difficulties in diverse student learning in number sense, and suggested the need to assess and support student learning in multiple forms and approaches.

References

- An, S., & Wu, Z. (2014). Using the evidence-based MSA approach to enhance teacher knowledge in student mathematics learning and assessment. *Journal of Mathematics Education*, 7(2), 108-129.
- Bernardo, A. I. (2005). Language and modeling word problems in mathematics among bilinguals. *The Journal of Psychology*, 139(5), 413-425.
- Brookhart, S.M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, 22(4), 5-12.
- California Department of Education. (2006). *Mathematics framework for California public schools*. Sacramento, CA: California Department of Education.
- California Department of Education. (2015). *Mathematics framework for California public schools*. Sacramento, CA: California Department of Education.
- Hewson, S. (2011). What is a mathematically rich Task? *NRICH enriching mathematics*. Retrieved from <https://nrich.maths.org/6299>
- Hiebert, J., & Wearne, D. (1993). Instructional tasks, classroom discourse, and students' learning in second-grade arithmetic. *American Educational Research Journal*, 30, 393-425.
- Mok, I.A.C., & Kaur, B. (2006). Learning tasks. In D. Clarke, J. Emanuelsson, E. Jablonka, & I.A.C. Mok, (Eds.), *Making connections: Comparing mathematics classrooms around the world* (pp. 147-200). Rotterdam, The Netherlands: Sense.
- Moss, P. A. (2003). Reconceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practice*, 22(4), 13-25.
- National Council of Teachers of Mathematics. (1995). *Assessment standards*. Reston, VA: Author.
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston: VA: Author.
- National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards for mathematics*. Washington, DC: Authors.
- National Research Council. (1993). *Measuring what counts: A conceptual guide for mathematics assessment*. Washington, DC: National Academy Press.
- RAND Mathematics Study Panel. (2003). *Mathematical proficiency for all students toward a strategic research and development program in mathematics education*. Santa Monica, CA: RAND Education.
- Smith, M., Bill, V., & Hughes, E. (2008). Thinking through a lesson: Successfully implementing high-level tasks. *Mathematics Teaching in The Middle School*, 14(3), 132-138.

- Suurtaam, C., Koch, M., & Arden, A. (2010). Teachers' assessment practices in mathematics: Classrooms in the context of reform. *Assessment in Education: Principles, Policy & Practice*, 17(4), 399-417.
- Wiest, L. R. (2000). Mathematics that whets the appetite: Student-posed project problems. *Mathematics Teaching in the Middle School*, 5, 286-291.
- Wu, Z. (2008). A bridge for ELL student learning mathematics: Multiple representations. *National Association for Bilingual Education News*, July-October 7-12.
- Wu, Z., & An, S. (2015). *Teaching elementary and middle school mathematics using the MSA approach: Model, strategy, and application* (2nd ed.). Irvine, CA: Education for All Publisher.
- Wu, Z., & An. (2016). Addressing challenges in urban teaching and learning math using model- strategy-application with reasoning approach in linguistically and culturally diverse classrooms. *Journal of Urban Learning Teaching and Research*, 12, 47-60.

Authors:

Shuhua An
California State University, Long Beach, USA
Email: shuhua.an@csulb.edu

Zhonghe Wu
National University, USA
Email: zwu@nu.edu