IUMPST: The Journal. Vol 2 (Pedagogy), April 2020 [www.k-12prep.math.ttu.edu]
ISSN 2165-7874

Exploring Preservice Teachers' Assessment of Students' Understanding of Mathematics

Robert F. Cunningham
Ewing, New Jersey, USA
The College of New Jersey
bobc@tcnj.edu

Kimberly N. Cook
Ewing, New Jersey, USA
The College of New Jersey
cookk3@tcnj.edu

Abstract

This study used a pretest-posttest design to explore the scoring by female preservice teachers (n=25) of student responses to a pattern problem. The treatment consisted of a cooperative activity that required the preservice teachers to assign scores to a rubric intended to enhance careful attention to student reasoning expressed in written explanations. On the posttest, the majority of preservice teachers elevated their scores for written explanations that demonstrated understanding. However, this did not prove sufficient to dissuade some from giving higher scores to correct answers paired with explanation that did not demonstrate understanding.

Introduction

Teachers must look beyond correct and incorrect answers and recognize the importance of students' mathematical understanding evidenced in written explanations. There is considerable research that points to the benefits of assessments that provide a window into student thinking. Requiring students to explain their reasoning can reveal their progress and allow teachers to shape instruction to meet the needs of their students (Choy, 2016; Grainger & Adie, 2014; Jacobs & Spangler, 2018; Morrow-Leong, 2016). Further, Wilson, Mojica, and Confrey (2013) found that teacher knowledge of students' mathematical thinking is not only a key component in changing instructional practice, but can also improve student achievement.

How teachers assess student work can influence what students consider important. The National Council of Teachers of Mathematics (NCTM, 2000) states "The tasks teachers select for assessment convey a message to students about what kinds of mathematical knowledge and performance are valued." In addition, the types of assessments should include opportunities for students to show mathematical proficiency. The Common Core State Standards for Mathematics (CCSSI, 2010) indicates that mathematically proficient students are able to justify their answers in a way that is appropriate to the students' mathematical maturity, construct viable arguments, justify their conclusions, and communicate their thinking to others. If teachers want to promote such values and skills then their assessments must mirror them.

While teachers must develop and grade authentic assessments that reveal student understanding in mathematics, their personal experiences of being assessed can influence their beliefs about assessment (Frykholm, 1999; Grainger & Adie, 2014). Too often, their own experiences in mathematics emphasized instrumental instruction focused on procedures with correct answers, defined as procedural fluency, rather than for relational understanding, which emphasizes reasoning and deep understanding. Because relational understanding is so critical to learning mathematics, it should be valued as much if not more than instrumental understanding.

Nevertheless, some teachers rely on procedural fluency as evidence of conceptual understanding (Morris, 2006; Spitzer, et al., 2011) and many grade by only using the judgement of correct or incorrect answers while providing few opportunities for students to explain their reasoning (Bieda, 2010). Although personal experiences and beliefs may interfere, teachers must include opportunities for students to show their mathematical proficiency by explaining their reasoning in written explanations and assign value to those explanations to promote relational understanding.

Research has confirmed the need for teacher education programs to include effective methods of assessment (Allen & Lambating, 2001; Eisenkraft & Eisenkraft, 2011; Frykholm, 1999). Yet few studies have addressed assessment practices of undergraduate preservice teachers, or how such practices might be improved (Grainger & Adie, 2014). Preservice teachers spend considerable time learning how to teach and programs include some instruction on assessment, but issues related to grading need more attention (Allen & Lambating, 2001; Randall & Engelhard, 2010). There is a need to examine how preservice teachers evaluate student responses when written explanations do, or do not, demonstrate understanding. Walkoe (2015) found that teaching for relational or conceptual understanding is a challenge, and this could apply to evaluating written explanations that reflect such understanding. Jacobs, Lamb, & Philipp (2010) suggest that teachers must learn to pay careful attention to different ways of student thinking. In this study, we examine how preservice teachers score student answers paired with written explanations for pattern problems and undertake an effort to improve their assessment practices.

Background

In this study, Skemp's (1976) two perspectives on understanding mathematics frame the act of evaluating answers paired with written explanations; one he describes as *instrumental* and the other he describes as *relational*. *Instrumental* understanding relates to those reliable and typically efficient procedures students might use to produce correct answers. *Relational* understanding refers to a deeper type of understanding, one that students might reveal in written explanations that reflects why and how mathematics works, and is applied. While it is common for teachers to provide students with opportunities to reveal instrumental understanding (Bieda, 2010), those that reveal relational understanding are less common. If students are to value the ability to explain their understanding of mathematics (NCTM, 2000), then teachers must learn to provide such opportunities on assessments.

Another aspect of the framework for the act of evaluation is *Teacher Noticing* or *Professional Noticing. Teacher Noticing* involves careful attention to making sense of what students say, do, or write (Grainger & Adie, 2014; Spitzer, et al., 2011; Walkoe, 2015). *Noticing* often refers to the teachers' response in the moment in the classroom (Jacobs & Spangler, 2018) but this study confines itself to *Teacher Noticing* of student understanding with careful attention to making sense of, and the evaluation of, written explanations**.** Callejo & Zapatera (2016) state that T*eacher's Noticing* of students' mathematical thinking is an important part of teacher competency. Many times teachers do not see something because they are not looking for it, or because they do not understand what it is they are asked to look for (Walkoe, 2015). Still recent studies of *Teacher Noticing* have suggested that it is a learnable practice for both preservice teachers and practicing teachers.

Teacher education programs need to prepare preservice teachers to develop effective methods to assess student understanding and develop grading practices and assessments that clearly and accurately communicate their goals (Allen & Lambating, 2001). This process is not

ISSN 2165-7874

as straight forward as one might imagine. Frykholm (1999) found that preservice teachers considered the "act" of grading as objective, sound, and reliable, but ironically, not when it came to their personal experiences. Many preservice teachers reported dissatisfaction with exams that they felt were unfair and did not reflect what they knew about the material. Both the types of test items chosen and how they are scored can influence student satisfaction, and their expectations about the kind of mathematical understanding that is valued.

Despite written explanations offering a window into student understanding, the scoring of them can present a challenge. Rubrics can improve assessment by ensuring a higher degree of inter- and intra-grader reliability, and they have the potential to improve learning by making expectations and criteria explicit (Eisenkraft & Eisenkraft, 2011; Jonsson & Svingby, 2007). Eisenkraft and Eisenkraft (2011) studied the scoring by 202 educators including teachers from all grade levels and recommend that if students are going to be evaluated using tests then rubrics should be used. The examples used were trivial fraction problems with definitive answers. Using one of their examples and looking beyond their study suggests that focusing only on grading answers can devalue reasoning and send the wrong message to students. The instruction for Examples A, B, and C shown in Figure 1 is to "Simplify the fractions and show all work."

Figure 1.
Sample solutions

Example A: $\frac{16}{64}=\frac{1}{4}$ Example B: $\frac{16}{64}=\frac{1\cancel{6}}{\cancel{6}4}=\frac{1}{4}$ Example C: $\frac{16}{64}=\frac{1\cdot\cancel{16}}{4\cdot\cancel{16}}=\frac{1}{4}$

All three examples have the correct answer but the associated scoring can be problematic. If the same score is assigned to Example A (*correct answers without work*) as to Example C (*correct answers with work shown that demonstrated understanding*) then showing work is not necessary even when requested. Also, assume that for Example A and B that the same incorrect method (*cancelling the 6's*) is used, then assigning a lower score to B than A suggests that showing work can have a negative impact and might not be worth the effort. Finally, giving the same score to all three examples because they have correct answers ignores reasoning and sends the wrong message.

There is also a risk that students and others might embrace these faulty generalizations about showing work or providing written explanations. They might avoid or devalue written explanations, and develop limited beliefs about what types of understanding are important in mathematics (Frykholm, 1999; Grainger, & Adie, 2014). Boaler (2016) writes that often parents asked her: "What is the point of my child explaining their work if they get the right answer?" Her response is always the same: "Explaining your work is what, in mathematics, we call reasoning, and reasoning is central to the discipline of mathematics."

Since short, structured items typically dominate mathematics tests in schools and often fail to assess reasoning (Jones & Inglis, 2015), then the augmentation of written explanations to some items might seems fitting. Yet, if teachers include problems that require a written explanation and an answer, they may score them as a bundle failing to *Notice* student understanding with the answer becoming the predominant consideration (Bieda, 2010; Spitzer, et al., 2011). One way to circumvent this might be to have them develop a rubric that requires assigning scores to different combinations of correct and incorrect answers paired with written explanations. This strategy might prevent answers from being the predominant consideration while promoting increased attention to written explanations and student understanding. Despite the importance of correct answer, written explanations are just as important, worthy of consideration and the careful attention of *Teacher Noticing.*

Besides using a rubric to improve assessment and clarify expectations, allowing preservice teachers to assign scores to the rubric as a group might enhance their *Noticing* of student understanding. Studies have shown that participating in a group setting helped teachers make sense of classroom interactions and increased teachers' ability to attend to and reason about student thinking (Borko, 2016; Gamoran Sherin & van Es 2009; Jacobs, et al., 2010). If small groups of preservice teachers were provided with an opportunity to discuss and reach consensus on assigning scores to a rubric, they might avoid grading using only the judgement of correct or incorrect answers (Spitzer, et al., 2011) and *Notice* student understanding expressed in written explanations.

The generalization of pattern problems is a suitable topic for this study, since it allows for variety of student explanations and often appears in research on *Noticing* (Callejo & Zapatero, 2016; Rusdiana, et al., 2017; Yilmas , Durmus, & Yaman, 2018). In addition, the topic can assist students transitioning from arithmetic thinking to algebraic thinking, since it requires them to produce rules that can be used to determine any term of the pattern (Rusdiana, et al., 2017). Another reason reported is that when *Noticing* algebraic thinking some teachers are reticent to make sense of student thinking (Walkoe,2015). Algebra teachers often focus on performing procedures (Stephens, 2008; Walkoe, 2015), and pay attention to this aspect above other types of algebraic thinking, such as finding patterns or generalizing. This topic will provide a yardstick for preservice teachers' assessment of understanding exemplified by sample student responses.

For this study, sample responses to a pattern problem will be used in a pretest–posttest design with a treatment that integrates a rubric as part of a cooperative learning activity and requires preservice teachers to assign grades to combinations of answers and written explanations. The sample responses will include an *incorrect answer paired with an explanation that does demonstrate understanding,* and a c*orrect answer paired with an explanation that does not demonstrate understanding*. This is to provide insight to preservice teachers' valuations of answers and written explanations. Prior research suggests that correct answers might dominate scoring but the use of a rubric and a cooperative group activity might reduce such dominance. This might improve both assessment and *Teacher Noticing* of student understanding that is so critically important in mathematics. Ultimately, this might advance the goal of increased relational understanding and achievement (Wilson, Mojica, & Confrey, 2013).

Research Question. Will elementary preservice teachers' rescoring of a pattern problem on a posttest differ from the pretest scoring after participating in a cooperative group activity that requires them to reach consensus on the assignment of scores to a rubric designed to enhance *Teacher Noticing* of written explanations?

Methods

Participants. Twenty-five undergraduate preservice teachers enrolled in a selective state college in the northeastern United States participated in the study. The college has earned national recognition for its commitment to excellence and has become an example of the best in public higher education. The participants were female elementary education majors with the following academic concentrations: English (8), Psychology (4), I-stem (4), Gender (2),

ISSN 2165-7874

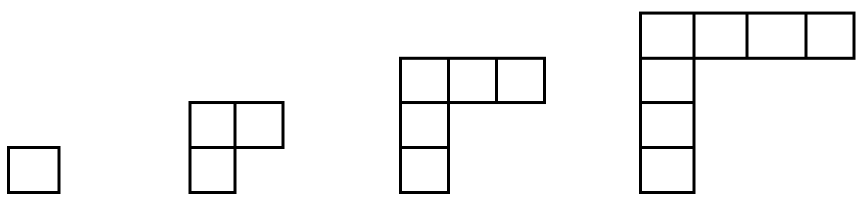
Mathematics (2), Sociology (2), History (1), Music (1), and Spanish (1). There were 21 freshmen, 1 sophomore and 3 juniors.

The preservice teachers were enrolled members of one section of a required elementary school mathematics content course, MAT105 Math Structures & Algorithms for Education I. This is the first of a two-course sequence taken by preservice elementary teachers. The text used in the course is *Mathematics for Elementary School Teachers* (Bassarear & Moss, 2016). The course covers the first 6 Chapters: Foundations of Mathematics, Sets & Numeration, The Four Fundamental Operations of Arithmetic, Extending the Number System, Ratio and Proportion, Algebraic Thinking. The topic of pattern generalization falls under Algebraic Thinking. One of the authors served as the instructor for the course.

Instruments. The scoring instrument in Figure 2. was used in the pretest and the posttest to determine the scores assigned by preservice teachers to fictitious responses from Student A and Student B. Each response contained a written explanations and an answer to a pattern problem.

Figure 2
Scoring Instrument

Question: Determine the number of boxes in the nth term and explain.



Student A: "There is a row and a column for each figure. Each figure has N boxes in a row and N boxes in a column but one box is counted twice." Answer: 2 N + 1

Student B: "In the second figure it looks like one box has been subtracted from a 2 x 2 square." Answer: 2 N – 1

The correct answer is 2 N -1. Score each student response to the given question.

How many points (0 to 5) would you give to the response given by **Student A**?
Explain your reasoning:

How many points (0 to 5) would you give to the response given by **Student B**?
Explain your reasoning:

Prior to the Posttest, students worked in groups and reached consensus on the scores assigned to each rectangle in the Rubric Instrument given in Figure 3. The instructions asked students to determine how many points will be awarded for the four types of responses by placing a number from 0 to 5 in each box.

Figure 3
Rubric Instrument

Rubric for Pattern Problems		
	Explanation demonstrates understanding	Explanation does not demonstrate understanding
Correct answer		
Incorrect answer		

Procedure. A Pretest-Posttest design was used to measure the change in scoring of two sample student responses. At no time was a reference made by the instructor to valuing answers or explanations prior to the pretest up until the conclusion of the posttest.

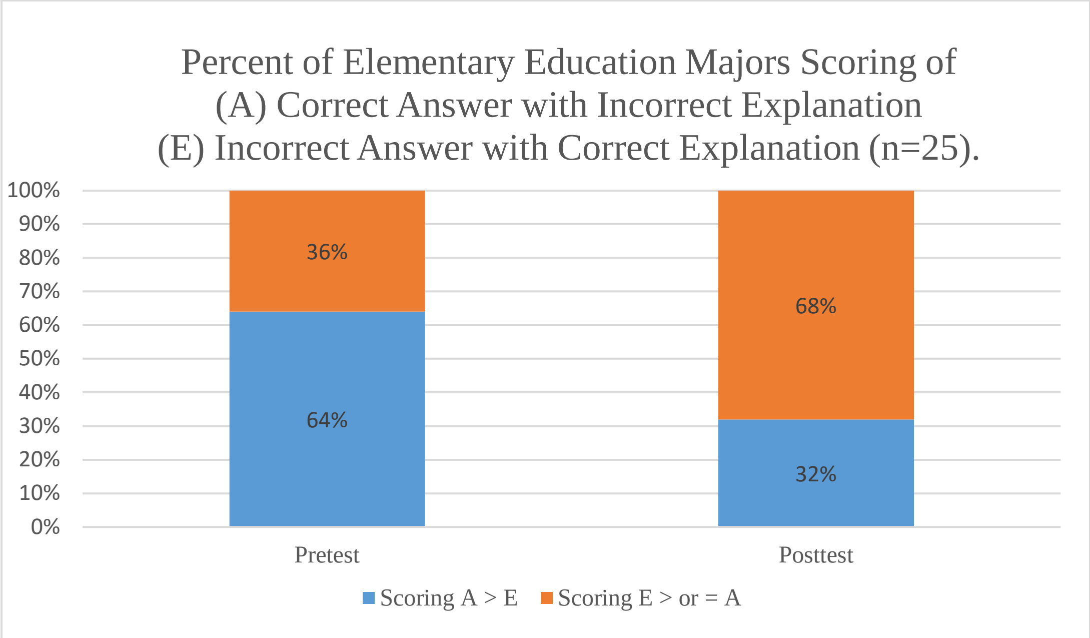
Prior to collecting scoring data, two 80-minute class period were devoted to the exploration of pattern problems that required finding the next term, an intermediate term, and generalization to the nth term. These class periods were to familiarize the preservice teachers with pattern problems and to give them the necessary understanding prior to scoring the sample student responses on the scoring instrument. Class time included lecture, small group explorations with individual student presentations of several problems contained in Chapter 6 of *Mathematics for Elementary School Teachers* (Bassarear & Moss, 2016), along with 6 homework problems from the exercises in the text. Following the two class periods, the pretest consisting of the scoring instrument was administered allowing for 15 minutes during class time.

One of the authors conducted a cooperative group activity two weeks after the pretest. The class was divided into 7 groups of 3 or 4 students. The groups worked independently. They discussed and reached consensus on filling in the scoring rubric. The responses were constructed so that they included the following 4 pairings: *correct answer with explanation that demonstrates understanding*, *correct answer with explanation that* ***does not*** *demonstrate understanding*, *incorrect answer with an explanation that demonstrates understanding*, and *incorrect answer with explanation that* ***does not*** *demonstrate understanding*. When each group had reached consensus, each individual used the rubric to re-score the same sample student responses from the pretest. As before about 15 minutes was allowed for scoring.

Results

On the pretest, of the 25 preservice teachers 16 (64%) scored the *correct answer – incorrect explanation (one that did not demonstrate understanding*) higher than *incorrect answer – correct explanation (one that did demonstrate understanding*). On the posttest, this number was 8 (32%). In addition, on the pretest 9 (36%) preservice teachers scored (*incorrect answer – correct explanation*) higher than or equal to (*correct answer – incorrect explanation).* On the posttest, this number was 17 (68%). Figure 4 displays these results.

Figure 4
Pretest and Posttest Scoring



After the cooperative group activity that required each group to reach consensus on assigning scores to the rubric, all but two of the seven groups allocated scores to E (I*ncorrect answer - explanation that demonstrated understanding*) higher than or equal to A (*Correct answer - explanation that did not demonstrate understanding*).

Discussion

Examining the pretest scoring of responses to the pattern problem revealed that the majority of preservice teachers scored correct answers paired with explanations that did not demonstrate understanding higher than incorrect answers paired with explanations that demonstrated understanding. This suggests that for some preservice teachers the answer became the predominant consideration when scoring (Bieda, 2010; Spitzer, et al., 2011). However, on the posttest fewer preservice teachers showed such bias. This suggests that for this sample of preservice teachers, the cooperative group activity involving the reaching of consensus on a scoring rubric and use of the resulting rubric elevated scoring of the response that contained the explanation that demonstrated understanding. The preservice teachers *Noticed*, or paid more careful attention to student thinking on the posttest. While correct answer are important, written explanations are just as important and should be part of *Teacher Noticing.*

Looking at some of the individual responses (using fictitious names) also suggests a change in scoring from pretest to posttest. On the pretest, Elizabeth gave the Student A response (2 points) because *"the answer is incorrect and the explanation explains the figures."* She assigned the Student B response (5 points) because *"the answer is correct but the explanation is only for one figure."* Elizabeth scored the response with the correct answer higher than the response with the correct explanation.

On the posttest, Elizabeth reversed the scoring and gave explanations that are more detailed. She gave the Student A response (4 points) because "*the answer is incorrect and the explanation works for all the figures and shows the student understands.*" She assigned the Student B response (2 points) because *"While they had the right answer, they used the wrong method to get there. It seemed like they only used the second figure. The third figure looks like 4 boxes have been taken out of a 3x3 square but the student didn't mention that. I gave the student 2 points because they had the right answer but the wrong reasoning."*

Yet for others like Isabella, even with the rubric, they continued to give correct answers higher scores and focus more on answers rather than on explanations. On the posttest, Isabella assigned Student A response (2 points) because "*for their effort and recognizing the 2n portion of the equation but their reasoning is incorrect. It would be illogical to count the same box twice.*" She gave the Student B response (5 points) because "*they found the correct equation as well as identified their own pattern and visual image of the equation.*" Isabella justified her scoring by mentioning "*effort*" and individually generated "*pattern and visual image.*" These justifications relied on assumptions outside the realm of the rubric. Randall & Engelhard (2010) found that teachers generally graded achievement, but sometimes considered variable such as behavior, ability, or effort. When teachers include these variables, they confuse the meaning of grades since grades should reflect only achievement.

It is possible that the rubric and associated discussion failed to influence the impact of Isabella's own mathematical assessment experiences; experiences that might have provided limited opportunities to explain her reasoning. In addition, her predilection to overvalued correct answers confirms a finding by Morris (2006) who used videotaped mathematics lessons to show that preservice teachers tend to over-attribute understanding to correct answers. Preservice teachers need to be aware of the meaning of grades, and realize that correct answers do not provide as clear a picture of student understanding and achievement as written explanations can provide.

Conclusion

For the majority of this small sample of preservice teachers, the cooperative activity associated with the rubric resulted in the elevation of scores for the Student A response (I*ncorrect answer - explanation that demonstrated understanding*) which showed a more careful *Noticing* of student understanding. However, this did not prove sufficient to dissuade some of the preservice teachers from giving high scores to correct answers regardless of explanation. When teachers score answers as only correct or incorrect without attention to written explanations or justification, it can send the wrong message to students about what is of value in mathematics.

Assessment should support the learning of important mathematics and written explanations can furnish useful information to both teachers and students. For teachers it can reveal student thinking and help shape instruction. For students the opportunity to give written explanations allows them to communicate mathematically and justify their reasoning. Such justification is a prelude to proof that is seldom part of mathematics instruction in schools, yet essential for student learning in mathematics.

Results of this study suggest that preservice educators must have their students develop and grade authentic assessments that reveal student understanding prior to them entering their own classrooms. Both relational and instrumental understanding are critical for student success. Consequently, student assessment should examine both types of understanding. Since Callejo, M.L. & Zapatero, A. (2016) and others have stated that *Teacher Noticing* is a learnable practice; educators need to engage preservice teachers in opportunities for *Teacher Noticing* of student thinking. Preservice teachers need to be aware of the importance of relational understanding and that being able to explain ones work is central to the discipline of mathematics (Boaler, 2016). How preservice teachers develop and grade assessments will have a profound influence on their students' achievement in mathematics.

References

Allen, J. D., & Lambating, J. (2001). Validity and Reliability in Assessment and Grading: Perspectives of Preservice and Inservice Teachers and Teacher Education Professors. Paper presented at the Annual Meeting of the American Educational Research Association (Seattle, WA, April 10-14, 2001). ED 453 167 SP 039 964.

Bassarear, T., & Moss, M. (2016). *Mathematics for Elementary School Teachers*, 6th edition, Boston, MA: Cengage Learning.

Bieda, K. (2010). Enacting proof-related tasks in middle school mathematics: Challenges and opportunities . *Journal of Research in Mathematics Education*, 41(4), 351-382.

Boaler, J. (2016). *Mathematical Mindsets: unleashing students' potential through creative math, inspiring messages and innovative teaching.* San Francisco, CA: Jossey-Bass

Borko, H. (2016). Methodological contributions to video-based studies of classroom teaching and learning; a commentary. *ZDM*, 48(1), 213-218.

Callejo, M.L., & Zapatero, A. (2016). Prospective primary teachers' noticing of students' understanding of pattern generalization. *Journal of Mathematics Teacher Education*, 20 (4), 309-333.

Choy, B. H. (2016). Snapshots of mathematics teacher noticing during task design. *Mathematics Education Research Journal*. DOI 10.1007/s13394-016-0173-3.

Common Core State Standards Initiative. (2010). *Common Core State Standards for Mathematics*. Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers. http://www.corestandards.org/wp-content /uploads/Math_Standards.pdf.

Eisenkraft, A., & Eisenkraft, N. (2011). When Wrong Answers Receive Top Grades. *Journal of College Science Teaching*, 41 (2), 28-31.

Ferrini-Mundy, J., Lappan, G., & Phillips, E. (1997). *Teaching Children Mathematics*, 3, 282-289.

Frykholm, J. (1999). Assessment in Mathematics Teacher Education: Introducing Preservice Teachers to Assessment Reform. *Teacher Educator*. 34(4), 244-258.

Gamoran Sherin, M., & van Es, E.A. (2009). Effects of video club participation on teachers' professional vision. *Journal of Teacher Education*, 60(1), 20-37.

Grainger, P.R., & Adie, L. (2014). How do preservice teacher education students move from novice to expert assessors? *Australian Journal of Teacher Education*, 39(7), 89-105.

Jacobs, V.R., Lamb, L. C. & Philipp, R. C, (2010). Professional Noticing of Children's Mathematical Thinking. *Journal for Research in Mathematics Education*, 41 (2), 169-202.

Jacobs, V.R. & Spangler D.A. (2018). In *Compendium for Research in Mathematics Education.* Edited by Jinfa Cai. Chapter 28 Research on Core Practices in K-12 Mathematics Teaching. Reston, VA: NCTM.

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability. Validity and educational consequences. *Educational Research Review*, 2, 130-144.

Jones, I., & Inglis, M. (2015). The problem of assessing problem solving: can comparative judgement help? *Educational Studies in Mathematics*, 89, 337–355.

Morris, A. (2006). Assessing pre-service teachers' skills for analyzing teaching. *Journal of Mathematics Teacher Education*, 9(5), 471-505.

Morrow –Leong, K. (2016). Evidence-Centered Assessment. *Teaching Children Mathematics*, 23(2), 82-89.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA:Author.

Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education* 26, 1372-1380.

Rusdiana, S., Sutadwidjaja, A., Bambang Irawan, E., & Sudirman, (2017). Pattern Generalization by Elementary Students. 5th South East Asia Development Research (SEA-DR) International Conference. *Advances in Social Science, Education and Humanities Research,* 100, 379-381 Atlantic Press.

Skemp, R. R. (1976). Relational understanding and instrumental understanding. *Mathematics Teaching*, 77, 20-26.

Spitzer, S., Phelps, C.M., Beyers, J.E.R., Johnson, D.Y., & Sieminski, E.E. (2011). Developing prospective elementary teachers' ability to identify evidence of student mathematical achievement. *Journal of Mathematics Teacher Education*, 14(1), 67-87.

Stephens, A. (2008). What "counts" as algebra in the eyes of preservice elementary teachers? *Journal of Mathematical Behavior*, 27(1), 33-47.

Walkoe, J. (2015). Exploring teacher noticing of student algebraic thinking in a video club. *Journal of Mathematics Teacher Education*, 18(6), 523-550.

Wilson, P.H., Mojica, G., & Confrey, J. (2013). Learning trajectories in teacher education: supporting teachers' understanding of students' mathematical thinking. *Journal of Mathematical Behavior*, 32, 103-121.

Yilmas , Y., Durmus, S., & Yaman, H. (2018). An investigation of pattern problems posed by middle school mathematics preservice teachers using multiple representations. *International Journal of Research in Education & Science*, 4(1), 148-164.