January 2015

Pre-service and In-service Teachers' Rubric Assessments of Mathematical Problem Solving

Grant Kuehl
Emmanuel College - Boston, kuehlg@emmanuel.edu

Kimberly Sofronas
Emmanuel College - Boston, sofronki@emmanuel.edu

Abigail Lau
Emmanuel College - Boston, laua@emmanuel.edu

Pre-service and in-service teachers' rubric assessments of mathematical problem solving

Grant Kuehl, Kimberly Sofronas, and Abigail Lau

Emmanuel College

400 The Fenway

Boston, MA 02115

kuehlg@emmanuel.edu

sofronki@emmanuel.edu

laua@emmanuel.edu

Pre-service and in-service teachers' rubric assessments of mathematical problem solving

In mathematics education, there is often a conception that math skills are easily and objectively assessed. However, mathematical problem solving is a performance assessment that requires judging a student's overall performance on a problem, making it more complex than simply marking an answer right or wrong. Rubrics can reduce subjectivity in scoring mathematics problems (Stemler, 2004). Research has shown that repeated rubric use increases teacher confidence with rubrics and results in reliable rubric scores (Jonsson & Svigby, 2007; Silvestri & Oescher 2006). One might expect in-service teachers to have more confidence using rubrics and produce more reliable rubric scores than pre-service teachers; however more research is needed to explore those hypotheses. This study compares pre-service and in-service teachers use of a rubric designed to assess 4th-grade student mathematical problem solving. Additionally, the study adds to the mathematics education literature by examining pre-service and in-service elementary school teachers' attitudes toward assessment and confidence using rubrics.

Background Literature

Today, both local and national standards and accountability initiatives have increased demands for K-12 schools and teachers in the United States to use a variety of formative and summative assessment practices to document student learning and student work samples. Recent national initiatives such as No Child Left Behind, Race to the Top, and the adoption of Common Core State Standards have left many teachers overwhelmed by the increased emphasis on using various assessments to inform instructional design and evaluation at all levels (Abrams et. al 2003; Bryant & Driscoll, 1998; Mertler, 2011; Noddings 2007; Stecher, 2002; Stiggins, 2002; Vogler, 2002).

As U.S. schools and teachers draw increasingly upon assessment data to inform their instructional practices, the use of rubrics - a common tool for formative and summative assessment (Schafer, et. al, 2001) – will become increasingly important. A rubric is a "…document that articulates the expectations for an assignment by listing the criteria, or what counts, and describing levels of quality from excellent to poor" (Andrade, 2000). A Rubric can help measure the process and progress of learning while simultaneously serving as communication tool to help students understand their performance on a given task (Cooper & Gargan, 2009; Montgomery 2000; Moskal 2000). Rubrics also have the potential to facilitate greater student understanding of concepts and skills which, in turn, can lead to improved performance on subsequent tasks through effective feedback (Black et. al, 2002; Butler, 1988; Hattie & Timperley, 2005).

Research has shown that novice and experienced teachers have significant differences in their knowledge bases for teaching (Kleickmann et al., 2012; Quinn, 1997; Schempp et al., 1998) and in their skills and attitudes related to educational measurement (Alkharusi et al., 2011; Green, 1992). Teachers without rubric training often lack the knowledgebase to use or construct rubrics, to evaluate student work, and-or to interpret the results of rubric assessments (Davis, 1993; Lovorn & Rezaei, 2011; Plake et. al 1993; Reynolds-Keefer, 2010; Turley & Gallagher, 2008). With effective training, teachers can more reliably score student work across disciplines (Knoch, Read & von Randow, 2007; Schafer, et. al 2001). However, little is known about differences in pre-service and in-service teachers' use of rubrics to assess student understanding in mathematics. To that end, this study was designed to address gaps that exist in the mathematics education literature around pre-service and in-service teachers' use of rubrics and

their attitudes toward and knowledge of assessment. Specifically, this study examined the following research questions:

1. Are there any differences between pre-service and in-service teachers with respect to their attitudes toward assessment and-or rubrics?
2. How do the rubric scores produced by in-service teachers differ from the scores given by pre-service teachers?
3. Do pre-service teachers and in-service teachers differ in the use of rubrics for diagnosing students' strengths and weaknesses?

The researchers hypothesized (a) in-service teachers would report greater confidence levels using rubrics and exhibit better attitudes toward assessment than pre-service teachers, (b) in-service teachers would be more reliable in their rubric scores than pre-service teachers, and (c) in-service teachers would identify greater differences than pre-service teachers in their evaluations of the three student work samples used in this study.

Methodology

To answer the research questions posed in this study, the researchers developed a three-part survey instrument, including a mathematical problem solving rubric (see Appendix A).

Part One of the survey contained demographic questions.

Part Two contained 23 objective item stems on participants' (a) beliefs about the purpose and usefulness of rubrics, including ideas about providing feedback through rubrics; (b) perceived confidence and self-efficacy around the use and development of rubrics, including ideas around communicating and interpreting the results of a rubric assessment; and (c) general attitudes toward assessment, including ideas about the reliability of assessment data and its impact on teaching practices. There was no existing survey instrument that comprehensively

captured the desired constructs. However, a review of the literature on assessment along with existing survey instruments designed to assess teacher attitudes toward assessment and rubric usage (Alkharusi 2011; Green 1992) provided a foundation for the development of the new items stems in Part Two of the survey (see Appendix A). A 5-point likert scale ranging from Strongly Agree to Strongly Disagree was used.

Part Three was comprised of three 4th-grade student mathematics work samples. Participants were asked to evaluate each using the rubric tool developed for this study (see Appendix A) and then answer open-ended questions to elicit participants' ideas about additional criteria for evaluating the problem, the students' mathematical strengths and weaknesses, and providing parental feedback.

The rubric in the survey featured evaluation criteria aligned with the key mathematical concepts, skills and processes involved in solving the mathematical problem developed for the purpose of this study. To validate the rubric tool, five Ph.D. mathematicians were asked to participate in think-aloud sessions using an adaptation of van Someren, Barnard, and Sandberg's (1994) protocol. A think-aloud is a protocol for understanding a person's thought processes as s/he engages in a task. An observer encourages the person to voice his or her thinking aloud and illuminate methods used to complete the task along with any difficulties encountered (van Someren et al., 1994). In this study, two researchers were present at think-aloud sessions to record observations and communicate with the Ph.D. mathematicians. The rubric tool was refined in several ways, including, but not limited to: (a) the introduction of a continuous rating scale, (b) the elimination of the *Exceeds Expectations* scale and (c) labeling only the low and high ends of the scale.

The final robust rubric tool contains a continuous rating scale with 1 = Does Not Meet Expectations and 4 = Meets Expectations to evaluate student work samples on five dimensions: (i) Mathematical Knowledge of Multiplication, Division and Number Sense, (ii) Understanding of the Problem, (iii) Accuracy, (iv) Process and (v) Mathematical Reasoning Skills. In addition to providing ratings on each dimension, raters also give a holistic rubric score.

Ten fourth-graders from an urban Catholic school in the Northeast United States provided work samples for possible inclusion in the survey instrument. Students were asked to solve two mathematical problems (i.e., The Field Trip Problem and The Monroe Shirt Store Problem) each aligned with Common Core State Standards for Mathematics. Student work samples from *The Field Trip Problem* were selected for use in this study: *The entire fourth-grade at Smith Elementary School is going on a field trip. There are 6 fourth-grade classes at the school. Each class has 20 students and 1 teacher. Each bus holds 30 people. What is the fewest number of buses needed for the field trip? Assume every fourth-grade student and teacher will attend.* Responses to the problem were selected from Sam, Lauren and Jeff [1] for inclusion in the survey instrument: Sam's work (see Appendix B) was selected because he decided to "round down" to four buses instead of "rounding up" to five buses to accommodate the extra six individuals who would not fit on four buses by rationalizing that the six teachers could each squeeze into a seat with children instead of ordering an additional bus. The researchers were interested in differences in pre-service and in-service teachers' evaluation of his interesting interpretation of the remainder. Lauren's work (see Appendix B) was selected because it contains both symbolic and graphical representations of the mathematical concepts. In addition, all of the Ph.D. mathematicians who participated in the think-alouds gave her work sample a holistic rubric score of 4. The researchers were interested in determining whether or not pre-service and in-service

[1] Pseudonyms have been used in place of students' actual names.

teachers would score this work sample similarly. Jeff's work (see Appendix B) was selected for inclusion in the survey to explore differences in pre-service and in-service teachers' scores on a difficult-to-navigate solution.

Participants and procedures. The teachers who participated in the study were 33 pre-service teachers and 43 in-service teachers. Pre-service teachers were recruited primarily from colleges and universities in the Northeast region of the United States via flyers and Facebook. All pre-service teachers participating in the study were enrolled in an initial licensure teacher preparation program for elementary education. In-service teachers were recruited from online networks for teachers, teacher professional development workshops, and via flyers and Facebook. All in-service teachers participating in the study were full-time classroom teachers teaching in grades 1 - 6. Of the in-service teachers who participated, 42% held a Bachelor's degree, 47% held a Master's degree, 2% held an Ed.D., and 9% held another form of higher degree (e.g., CAGS). In-service teachers' years of experience teaching in elementary school classrooms ranged from less than five years to over 30 years in the classroom. The average number of years of teaching experience among in-service teachers was between 10 and 20 years. Collectively, the in-service teachers in this study reported teaching experience that spanned all grade-levels from Pre-K to Grade 6.

Data Collection. Qualtrics is an HIPPA-compliant online-survey tool that was used as a platform to develop and administer the instrument. Data collection for this survey began in January 2014 and closed in August 2014. Responses were gathered over time as participants were recruited. Consenting participants were given a link to the survey and were informed of their right to withdraw from the study or skip any questions on the survey for any reason. Those

who completed the survey had the option of entering a raffle for one of two $25.00 online gift-cards.

Analysis and Results

Missing Data. Not all participants in the study completed Part Three of the survey, which involved rating student work samples using the developed math rubric. Specifically, only 18 out of 33 (54%) of pre-service teachers provided responses in Part Three, and 29 out of 43 (67%) of in-service teachers participated in this part of the study. The attrition at this point in the survey is likely due to survey fatigue and low motivation to do the portion of the survey that required more thought and concentration.

An additional 8 of the in-service teachers who submitted rubric scores for Sam's work and Lauren's work did not complete the rubric for Jeff's work. Thus, only 21 in-service teachers submitted scores for Jeff's work. Only 1 pre-service teacher who began Part Three of the survey did not complete scores for all three students. The differential rate of attrition across pre-service and in-service teachers for Jeff's work suggests that survey fatigue is not the only cause of missing data.

Research Question 1. *Were there any differences between pre-service and in-service teachers with respect to their attitudes toward assessment and-or rubrics?* T-tests were used to identify differences between pre-service and in-service teachers on item stems 1 - 22 of the survey instrument. Analysis of the survey data showed differences between pre-service and in-service teachers on the items stems depicted in Table 1. The results in Table 1 indicate that practicing teachers are more confident than teachers in training with respect to the use of rubrics

to evaluate students' strengths and weaknesses in mathematics. They also feel more prepared to develop and use rubrics to assess student work in mathematics (as well as other disciplines). Not

Table 1
Mean Differences in Pre-Service and In-Service Teachers Attitude Items

Item	Pre-Service Teachers	In-Service Teachers	t	df
Rubrics can be used to provide the teacher with feedback about student understanding.	3.81 (0.786)	4.22 (0.652)	-2.306*	66
Rubrics can be used to provide the student with feedback about his or her own understanding.	3.93 (0.730)	4.34 (0.693)	-2.369*	66
I would feel confident using a rubric to evaluate a student's strengths and weaknesses in mathematics.	3.38 (0.852)	4.00 (0.816)	-2.941**	64
I feel prepared to develop my own rubrics to assess student work in mathematics.	2.85 (0.967)	3.68 (1.118)	-3.099**	64
I feel prepared to develop my own rubrics to assess student work in other disciplines.	3.31 (1.011)	3.90 (0.955)	-2.406*	64
I feel prepared to use rubrics to assess student work in mathematics.	3.31 (1.011)	3.88 (0.911)	-2.367*	64
Standardized test results can be used to improve student learning.	2.88 (0.993)	3.40 (0.982)	-2.074*	64

Note. * = $p < .05$, ** = $p < .01$, *** = $p < .001$. Standard Deviations appear in parentheses below the means.

surprisingly, in-service teachers reported more frequent use of rubrics ($\bar{x}$= 3.44, sd = 0.590, N = 43) than pre-service teachers ($\bar{x}$= 2.87, sd = 0.922, N = 31), which might explain their self-confidence and perceived level of preparedness to use and develop mathematics rubrics. Interestingly, pre-service and in-service teacher responses did not differ significantly with

respect to their agreement around (a) rubrics resulting in more consistent and accurate assessments of student work, (b) confidence sharing the results of a rubric assessment with a student or parent in mathematics or other disciplines, (c) preparation to use rubrics to assess student work in other disciplines, (d) inclusion of more instruction on rubric design and implementation in pre-service and in-service teacher training, and (e) standardized tests as measures of teacher effectiveness and student achievement.

Research Question 2. *How did the rubric scores produced by in-service teachers differ from the scores given by pre-service teachers?* The descriptive statistics for the rubric scores, including scores for the 7 dimensions and the holistic score, are provided in Table 2 across pre-service and in-service teachers. T-tests were conducted to compare the mean rubric ratings across pre-service and in-service teachers for each student.

Compared to pre-service teacher ratings, in-service teachers rated Sam's work higher. These higher means were unlikely to occur by chance ($p < .05$) for understanding of the problem, accuracy, and math reasoning skills. Differences in mean rubric scores across pre-service and in-service teachers for Lauren's work were likely to occur by chance ($p > .05$). The pre-service teachers provided higher ratings than the in-service teachers for Jeff's work and the observed differences in holistic and problem-solving scores were unlikely to occur by chance ($p < .05$). In-service teachers gave more credit to Sam than pre-service teachers, but were harsher on the Jeff.

Table 2
Descriptive statistics for rubric scores across in-service and pre-service teachers

	In-service Teachers		Pre-service Teachers				
Rubric Criteria	*N*	*Mean*	*N*	*Mean*	*df*	*t*	*p*
Partially Correct Student							
Knowledge of Multiplication	35	3.61	21	3.61	54	0.03	0.40
Knowledge of Division	32	3.16	21	2.92	51	0.92	0.26
Knowledge of Number Sense	31	3.32	19	3.04	48	1.20	0.19

Understanding of the Problem	33	3.13	20	2.49	51	2.69	0.01
Accuracy	34	3.26	21	2.71	53	2.66	0.01
Problem Solving Process	34	3.23	21	3.10	53	0.74	0.30
Mathematical Reasoning Skills	35	3.27	21	2.67	54	2.45	0.02
Holistic Score	35	3.36	21	3.06	54	2.13	0.04
Correct Student							
Knowledge of Multiplication	30	3.92	19	3.87	47	0.65	0.32
Knowledge of Division	30	3.86	19	3.87	47	-0.09	0.40
Knowledge of Number Sense	30	3.89	17	3.92	45	-0.36	0.37
Understanding of the Problem	30	3.92	19	3.92	47	0.04	0.40
Accuracy	30	3.90	19	4.00	47	-1.80	0.08
Problem Solving Process	30	3.88	19	3.92	47	-0.49	0.35
Mathematical Reasoning Skills	30	3.93	19	3.90	47	0.41	0.36
Holistic Score	29	3.89	19	3.92	46	-0.31	0.38
Incorrect Student							
Knowledge of Multiplication	28	1.68	18	1.74	44	-0.26	0.38
Knowledge of Division	27	0.64	17	0.76	42	-0.70	0.31
Knowledge of Number Sense	28	0.98	18	1.42	44	-1.92	0.06
Understanding of the Problem	26	1.33	18	1.48	42	-0.70	0.31
Accuracy	24	1.13	18	1.44	40	-1.42	0.14
Problem Solving Process	25	0.87	18	1.40	41	-2.51	0.02
Mathematical Reasoning Skills	27	0.99	18	1.27	43	-1.50	0.13
Holistic Score	27	1.00	18	1.39	43	-2.25	0.03

To better understand how pre-service and in-service teachers varied in their rubric ratings, the total variability in the rubric scores was analyzed to see what factors in the assessment process seem to be causing the variance (Shavelson & Webb, 1991). The variance components for the rubric scores are presented in Table 3 for pre-service and in-service teacher samples. Differences in scores are due to differences in student abilities, differences in aspects of math ability (rubric criterion), differences in teachers, and the interactions between teachers, students, and the rubric criteria. Variance due to student abilities is considered true variance and is not measurement error. Variability due to other factors is considered a type of unreliability or measurement error. The proportion of variance attributable to students is one way of describing the reliability of the rubric scores. The proportion of variance attributable to students is 88% for

pre-service teachers and 76% for in-service teachers. Thus, it appears that pre-service teachers may have produced more reliable scores. Differences across in-service teachers accounted for 4% of the variability in rubric scores as compared to only 1% for pre-service teachers. Again, this suggests that pre-service teachers were less variable in their rubrics scores than in-service teachers.

Table 3

Analysis of variability in rubric scores across in-service and pre-service teachers

Variance Component	In-service	Teachers	Pre-service	Teachers
	σ	%	σ	%
Teacher	.071	4%	.017	1%
Rubric Criteria	.012	1%	.009	0%
Student	.079	4%	.023	1%
Teacher*Item	.000	0%	.008	0%
Teacher*Student	.054	3%	.104	4%
Teacher*Student*Item	.238	12%	.155	6%
Student	1.459	76%	2.274	88%

Research Question 3. *Did pre-service teachers and in-service teachers differ in the use of rubrics for diagnosing students' strengths and weaknesses?* To answer this research question, responses to the open-ended questions about student work in Part Three of the survey were examined.

Identification of other rubric criteria. Teachers were asked if there are any other criteria they would use to evaluate the students' work that were not included in the rubric. In-service teachers were inclined to also include "background knowledge / experience" in the evaluation of the student's mathematical work. For example, they wanted to know if Sam takes the bus to

school and-or whether his school district allows additional students to ride on a bus when it has reached its capacity. In addition, in-service teachers perceived a lack of clarity around the expectations for the solution to the mathematics problem posed for the purpose of this study, noting that no directions were given to the student on how to show his/her work. Finally in-service teachers identified "modeling" as an additional evaluation criterion. Pre-service teachers only added "student effort" as an additional criterion.

Identifying the student's strengths and areas for improvement. After rating the student's work using the rubric, teachers were asked to describe in their own words the student's strengths and needs. More than one-third of the in-service teachers who provided written comments commended Sam's resourceful and cost-efficient solution to a real-world problem. Says in-service teacher R_ePX[2], "This student was able to demonstrate understanding of multiplication and division strategies. He even identified the meaning of the remainder and problem-solved in a way that most teachers would do in real life. We are asking students to assume that a new bus would be ordered for six people when in fact [the student's] answer reflects a more realistic solution!" Some in-service teachers raised questions about the importance of "accuracy" versus "reasoning" when evaluating Sam's work on the mathematics problem, with a few expressing concern over the lack of accuracy in the Sam's final solution to the problem. In-service teacher R_6FK suggested Sam needs to work on being accurate and answer the question without making up his own rules.

Pre-service teachers noted Sam's solid understanding of the necessary multiplication and division strategies. Although one quarter of the pre-service teachers who provided written

[2] In this study, personal identifiers were not collected from survey participants. Qualtrics randomly assigned each participant a 17-character "ID" linked to their responses for purposes of data organization and analysis. The first four characters from those Qualtrics-assigned IDs are used here when sharing participants' written comments.

comments highlighted Sam's resourceful and cost-efficient solution as a strength of his work, half perceived it as a weakness. Their focus was on a need for accuracy and following directions, which they ***did not*** identify as unclear. For example, pre-service teacher R_0oa stated, "The student correctly used multiplication and division. However, the student did not fully understand the problem and needs to work on reading the problem carefully."

In-service and pre-service teachers saw no areas for improvement in Lauren's mathematical work. In-service teachers believed Lauren's use of modeling demonstrated higher order thinking and strong reasoning skills. Says in-service teacher R_e3A "This student has a good grasp of relationships between the numbers as well as how to use them. The child was able to model the problem well [and] explain what the remainder of the problem meant". Pre-service teachers identified a need to offer Lauren greater challenge but were not specific in suggesting methods for delivering that challenge.

In their evaluation of Jeff's work, in-service teachers identified several areas for improvement including fluency with multiplication facts, representing mathematical problems, applying multiplication and division strategies to work problems, and keyword identification. They recommended re-teaching concepts, practicing number facts, and possible evaluation for special education services.

Pre-service teachers noted Jeff's apparent lack of understanding of the problem, weak mathematical reasoning, and deficits in carrying out multiplication computations; however, pre-service teachers did make mention of Jeff's ability to recognize that multiplication is needed to solve the problem. Unlike the in-service teachers, pre-service teachers offered no specific strategies for supporting Jeff and, rather, expressed in a general way that additional *practice* may help develop his skills.

Feedback to Parents. Teachers were also asked what feedback they would give the parents of the student. Table 4 summarizes the different responses from in-service and pre-service teachers. Both groups of teachers focused on similar things in their feedback, but used different language.

Table 4
Summary of Pre-Service and In-Service Teachers' Feedback to Parents

Student	In-Service Teachers	Pre-Service Teachers
Sam	● Good math skills ● Strong critical thinking / reasoning ● Creative thinking ● Can apply knowledge to problem-solving contexts ● Needs to focus on what problem is asking	● Strong computational skills ● Creative thinking ● Understands which operations to use ● Needs to understand remainders ● Needs to check work / read directions carefully
Lauren	● Strong problem-solving skills ● Strong math skills ● Has achieved mastery	● Strong problem-solving skills ● Student understands the problem ● Student is ready for deeper / more complex understanding of division
Jeff	● Student does not understand the Problem ● Weak number sense ● Difficulty with multi-step problem-solving ● Student needs additional help; may benefit from re-teaching the concept(s), writing out each step, practicing multiplication facts, illustrating the problem	● Student demonstrated good effort ● Strengths in multiplication and math Facts ● Student needs extra practice work to develop problem-solving skills, multiplication skills, and division Skills

Discussion

This study was designed to address gaps that exist in the mathematics education literature around pre-service and in-service teachers' use of rubrics and their attitudes toward and knowledge of assessment. Findings of this research have shown that in-service teachers did report greater confidence using rubrics and exhibited more positive attitudes toward assessment than pre-service teachers. More specifically, in-service elementary school teachers felt more confident using rubrics in mathematics, more prepared to develop their own rubrics, and reported more experience using rubrics. In-service teachers also agreed more strongly than pre-service teachers that standardized test results can be used to improve student learning. In-service teachers' classroom teaching experience may explain these positive outcomes.

By contrast, the findings of this research did not support the hypothesis that in-service teachers would produce more reliable rubric scores than pre-service teachers when assessing 4th-grade student work in mathematics. Analysis of rubric scores revealed more measurement error variance in the ratings provided by in-service teachers. Pre-service teachers differed less in their rubric scores. One possible explanation for the less consistent ratings of in-service teachers might be their varied experiences versus the relatively similar experience of pre-service teachers.

In-service teachers did, in fact, identify greater differences in the student work samples than pre-service teachers. Neither teacher group was consistently more stringent than the other. Rather, the in-service teachers gave more partial credit (suggesting leniency) for Sam's work sample, yet they gave lower scores for Jeff's work sample. This finding may also be related to in-service teachers' years of experience teaching and assessing students' mathematical understanding. In-service teachers' more developed knowledge of the "landscape of learning" in elementary school mathematics may mean that they are more informed about what 4th-grade students should know and be able to do mathematically. In other words, in-service teachers'

ratings of Sam's work and Jeff's work might be further apart than pre-service teachers ratings of those same work samples because of their "big-picture" understanding of elementary school mathematics. In their interpretation and use of the rubric, in-service teachers were able to provide more specific recommendations for how a student could improve.

The inferences that can be drawn from this study are limited because of a relatively small sample of the teacher populations and attrition from the survey. The incomplete data for many participants may be due to survey fatigue. Future research will consider ways to retain study participants, especially the in-service teachers. Another limitation of the study was the use of only three student work examples. The number of student work pieces examined was limited in order to keep the survey length reasonable. However, a larger sample of student work would allow for stronger comparisons of pre-service and in-service score reliability. Also, the survey instrument to assess teacher attitudes toward assessment and the use of rubrics was developed for this study and has not been externally validated. If a larger sample of teachers was obtained, factor analytic techniques could have been used to explore possibly summing the survey items to create a more reliable measure of teacher attitudes.

Pajares (1992) discusses the importance of thinking about how teachers' attitudes and beliefs impact their instructional decisions and classroom practices. The results from this study show that teacher's attitudes may change as they gain experience in the classroom. In particular, their attitudes towards assessment and confidence using rubrics appears to develop through classroom teaching experience. Teachers gain confidence using rubrics through classroom experience; however, that does not necessarily translate into more reliability in rubric scores. Although, teachers with more experience in the classroom may be able to give more specific diagnostic feedback after using a rubric to evaluate student learning. Teacher education should

address teachers' attitudes toward assessment, confidence using rubrics, and ability to reliably evaluate student work using rubrics.

References

Abrams, L., Pedulla, J., & Madaus, G. (2003).Views from the classroom: Teachers' opinions of statewide testing programs. *Theory into Practice, 42(1)*, 18 – 29.

Alkharusi, H., Kazem, A., & Al-Musawai, A. (2011). Knowledge, skills and attitudes of preservice and inservice teachers in educational measurement. *Asia-Pacific Journal of Teacher Education, 39(2)*, 113 – 123.

Andrade, H. (2000). Using rubrics to promote thinking and learning. *Educational Leadership, 57(5),* 13 – 18.

Black P. J., Harrison C., Lee C., Marshall B., Wiliam D. (2002). *Working inside the black box: Assessment for learning in the classroom.*. London, UK: King's College London School of Education.

Bryant, D., & Driscoll, M. (1998). *Exploring classroom assessment in mathematics: A guide for professional development*. National Council of Teachers of Mathematics: Reston.

Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task□involving and ego□involving evaluation on interest and performance. *British Journal of Educational Psychology*, *58*(1), 1-14.

Cooper, B., & Gargan, A. (2009). Rubrics in education: Old term, new meanings. *The Phi Delta Kappan, 91(1)*, 54 – 55.

Davis, B.G., (1993). *Tools for Teaching*. San Francisco: Jossey-Bass.

Green, K.E. (1992). Differing opinions on testing between preservice and inservice teachers. *Journal of Educational Research, 86(1)*, 37 – 42.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, *77*(1), 81-112.

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

Kleickmann, T., Richter, D., Kunter, M., Eisner, J., Besser, M., Krauss, S., & Baumert, J. (2012). Teachers' content knowledge and pedagogical content knowledge: The role of structural differences in teacher education. *Journal of Teacher Education, 64*, 90 – 106.

Knoch, U., Read, J., & von Randow, J. (2007). Re-training writing raters online: How does it compare with face-to-face training? *Assessing Writing, 12*(1), 26-43.

Lovorn A. R., & Rezaei, M. (2011). Assessing the assessment: Rubrics training for pre-service and new in-service teachers. *Practical Assessment, Research and Evaluation 16(16)*, 1 – 18.

Mertler C. (2011). Teachers' perceptions of the influence of No Child Left Behind on classroom practices. *Current Issues in Education, 13(3)*, 1 – 35.

Montgomery, K. (2000). Classroom rubrics: Systematizing what teachers do naturally. *The Clearing House, 73(6)*, 324 – 328.

Moskal, B. (2000). Scoring rubrics: What, when & how? Practical Assessment, *Research and Evaluation, 7(3)*, 1-7.

Noddings, N. (2007). *When School Reform Goes Wrong*. New York, NY: Teachers College Press.

Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice, 12(4)*, 10–12.

Quinn, R. (1997). Effects of mathematics methods courses on the mathematical attitudes and content knowledge of preservice teachers. *The Journal of Educational Research, 91(2)*, 108 – 113.

Reynolds-Keefer, L. (2010). Rubric-referenced assessment in teacher preparation: An opportunity to learn by using. *Practical Assessment, Research & Evaluation, 15(8)*, Retrieved May 21, 2014: http://pareonline.net/getvn.asp?v=15&n=8

Schafer, W., Swanson, G., Bene, N., & Newberry, G. (2001). Effects of teacher knowledge of rubrics on student achievement in four content areas. *Applied Measurement in Education, 14(2)*, 151 – 170.

Schempp, P. G., Tan, S., Manross, D., & Fincher, M. (1998). Differences in novice and competent teachers' knowledge. *Teachers and Teaching: Theory and Practice, 4*, 9 – 20.

Shavelson, R. J. and Webb, N. M. (1991) *Generalizability Theory: A Primer*. Sage Publications, Thousand Oaks, CA.

Silvestri, L., & Oescher, J (2006). Using rubrics to increase the reliability of assessment in health classes. *International Electronic Journal of Health Education, 9*, 25 – 30.

Stecher, B. (2002). Consequences of large-scale, high-stakes testing on school and classroom practice. In L.S. Hamilton, B.M Stecher, & S.P. Klein (Eds.) *Making sense of test-based accountability in education.* (pp.79-100). Santa Monica, CA: RAND.

Stemler, S. (2001). An overview of content analysis. *Practical Assessment, Research & Evaluation, 7(17)*. Retrieved April 22, 2014 from http://PAREonline.net/getvn.asp?v=7&n=17.

Stiggins, R. J., (2002) Assessment crisis! The absence of assessment FOR learning. *Phi Delta Kappan. 83(10),* 758-765.

Turley, E., & Gallagher, C. (2008). On the uses of rubrics: Reframing the great rubric debate. *English Journal, 97(4)*, 87 – 92.

van Someren, M.W., Barnard, Y.F., & Sandberg, J.A.C. (1994). The think aloud method: A practical guide to modelling cognitive processes (knowledge-based systems). Academic Press.

Vogler, K.E. (2002). The impact of high-stakes, state-mandated student performance assessment on teachers' instructional practices. *Education, 123(1)*, 39-55.

Appendix A
Survey Instrument

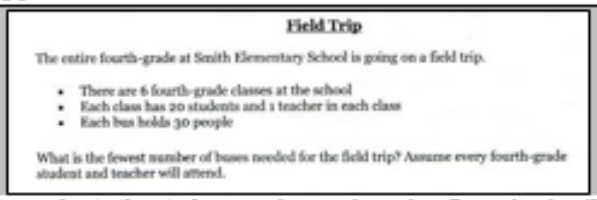
PART ONE: DEMOGRAPHIC QUESTIONS (IN-SERVICE TEACHERS ONLY)								
Which of the following best describes you?	Pre-service teacher currently enrolled in a teacher preparation program seeking elementary teacher licensure			In-service teacher currently teaching full-time in an			None of the Above	
What is the highest degree you have earned?	Bachelor's	Master's		Ed.D.		Ph.D.	Other	No Degree Earned
How many years have you been teaching at the elementary level?	Less than 5 years		Between 5-10 years		Between 10-20 years		Over 30+ years	
Which grade(s) have you taught? (Choose all that apply)	Pre-K	Kindergarten	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6
Please share the frequency with which you have used rubrics in the past.	Never		Rarely		Sometimes		Often	

PART TWO: OBJECTIVE QUESTIONS Please select the answer choice that best indicates your level of agreement with each statement[1].							
1. Rubrics help teachers to be consistent when assessing student work.							
2. Rubrics help teachers to be accurate when assessing student work.							
3. Rubrics can be used to provide the teacher with feedback about student understanding.							
4. Rubrics can be used to provide the student with feedback about his or her own understanding.							
5. Teacher preparation programs should include instruction on the use of rubrics to assess student learning.							
Please select the answer choice that best indicates your level of agreement with each statement[2]. **I would feel confident...**							
6. Using a rubric to evaluate a student's strengths and weaknesses in mathematics.							
7. Sharing the results of a rubric assessment with a student.							
8. Sharing the results of a mathematics rubric assessment with a parent.							
9. Sharing the results of a non-mathematics rubric assessment with a parent (e.g., language arts rubric, science rubric).							
10. I would feel confident using a rubric to evaluate a student's strengths and weaknesses in other disciplines including: (Check all that apply)	PE	English Language Arts	Art	History	Science	Foreign Language	Music
Please select the answer choice that best indicates your level of agreement with the statement[3]. **I feel prepared to...**							
11. Develop my own rubrics to assess student work in mathematics.							
12. Develop my own rubrics to assess student work in other disciplines.							
13. Use rubrics to assess student work in mathematics.							
14. Use rubrics to assess student work in other disciplines.							

[1] Items stems in this section of the survey used a 5-point Likert Scale (1 = Strongly Disagree to 5 = Strongly Agree)
[2] Items stems in this section of the survey used a 5-point Likert Scale (1 = Strongly Disagree to 5 = Strongly Agree)
[3] Items stems in this section of the survey used a 5-point Likert Scale (1 = Strongly Disagree to 5 = Strongly Agree)

Please select the answer choice that best indicates your level of agreement with the statement[4].										
15. I would benefit from more course work and-or professional development in rubric design and implementation.										
16. Standardized test results can be used to improve student learning.										
17. Students are tested too frequently in present-day elementary schools.										
18. Students' performance on a standardized test will reflect my effectiveness as an elementary school teacher.										
19. I feel pressure related to my students' achievements on standardized tests.										
20. Standardized test results accurately measure student achievement.										
21. Present-day elementary school teachers are encouraged to teach to the test.										
22. Schools should use students' performance on standardized tests to measure the effectiveness of a teacher.										
23. A student who scores __ or higher on an assessment demonstrates subject matter proficiency.	0 – 10%	10 – 20%	20 – 30%	30 – 40%	40 – 50%	50 – 60%	60-70%	70-80%	80-90%	90-100%

PART THREE: STUDENT WORK SAMPLES - In this section, you will be asked to evaluate 3 student work samples obtained from fourth-grade students at the end of the 2012 - 2013 academic year. Each student was asked to solve the following problem:



Field Trip

The entire fourth-grade at Smith Elementary School is going on a field trip.

- There are 6 fourth-grade classes at the school
- Each class has 20 students and 1 teacher in each class
- Each bus holds 30 people

What is the fewest number of buses needed for the field trip? Assume every fourth-grade student and teacher will attend.

A rubric will be provided to evaluate the student work samples using five criteria: (1) Mathematical Knowledge of Multiplication, Division and Number Sense, (2) Understanding of the Problem, (3) Accuracy, (4) Problem Solving Process, and (5) Mathematical Reasoning Skills.[5]

Student Work Sample A[6]

[4] Items stems in this section of the survey used a 5-point Likert Scale (1 = Strongly Disagree to 5 = Strongly Agree)

[5] Item stems in this section of the survey used a 4-point sliding scale rubric (1=Does Not Meet Expectations to 4=Meets Expectations

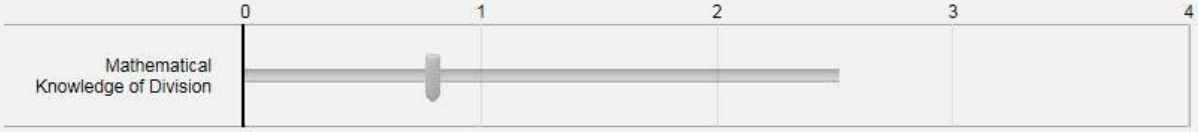
[6] Student Work Samples A, B and C were presented one-at-a-time and were each followed by the sliding scale evaluation criteria shown below.

How would you evaluate this student's "mathematical knowledge of multiplication" ?

1 Does not meet expectations	2	3	4 Meets expectations
The student did not identify the need to use a multiplication strategy to determine the total number of people attending the field trip.	The student used a multiplication strategy, but was unable to apply that knowledge correctly to determine the total number of people attending the field trip.	The student used a less efficient, but valid, multiplication strategy to correctly determine the total number of people attending the field trip.	The student used an efficient multiplication strategy to correctly determine the total number of people attending the field trip.

How would you evaluate this student's "mathematical knowledge of division"?

1 Does not meet expectations	2	3	4 Meets expectations
The student did not identify the need to use a division strategy to determine the total number of buses needed for the field trip.	The student used a division strategy, but was unable to apply that knowledge correctly to determine the total number of buses needed for the field trip.	The student used a less efficient, but valid, division strategy to correctly determine the total number of buses needed for the field trip.	The student used an efficient division strategy to correctly determine the total number of buses needed for the field trip.

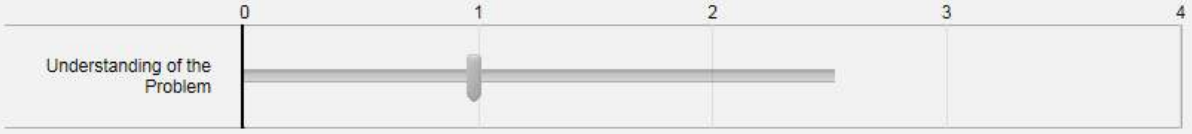


How would you evaluate this student's "mathematical knowledge of number sense"?

1 Does not meet expectations	2	3	4 Meets expectations
The student demonstrated little or no understanding of the operations, relationships between the operations, and the reasonableness of results.	The student demonstrated some understanding of the operations, relationships between the operations, and the reasonableness of results.	The student demonstrated a good understanding of the operations, relationships between the operations, and the reasonableness of results.	The student demonstrated strong understanding of the operations, relationships between the operations, and the reasonableness of results.

How would you evaluate the student's "understanding of the problem"

1 Does not meet expectations	2	3	4 Meets expectations
The student demonstrated little or no understanding of the purpose of the problem.	The student demonstrated some understanding of the purpose of the problem.	The student demonstrated good understanding of the purpose of the problem.	The student demonstrated strong understanding of the purpose of the problem.



How would you evaluate the student's "accuracy"?

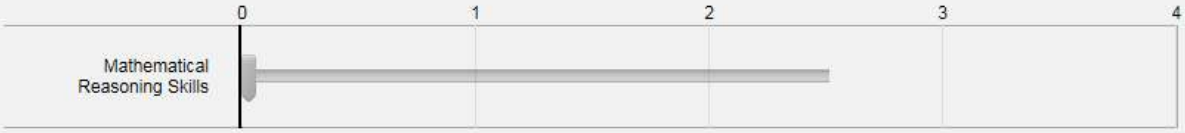
1 Does not meet expectations	2	3	4 Meets expectations
The work is completely inaccurate.	The work includes many significant computational errors. There is no final solution or the final solution is not accurate.	There are a few minor computational errors in the work. The final solution may or may not be accurate.	All computations in the work are correct. The final solution is accurate.

How would you evaluate the student's "problem solving process"?

1 Does not meet expectations	2	3	4 Meets expectations
The student was unable to solve any parts of the problem.	The student solved a few parts of the problem, but did not find a successful solution path.	The student solved most parts of the problem and found a nearly successful solution path.	The student solved all parts of the problems and found a successful solution path.

How would you evaluate the student's "mathematical reasoning skills"?

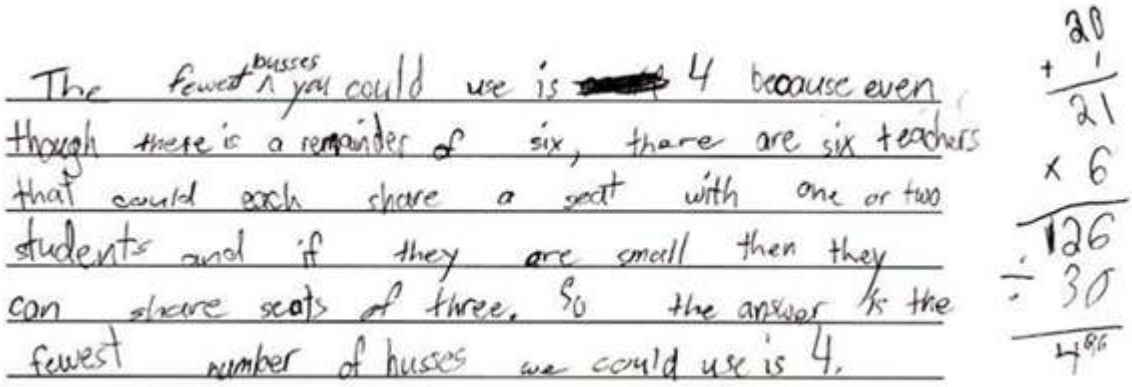
1 Does not meet expectations	2	3	4 Meets expectations
There was little or no logical organization to the ideas in the student work. Ideas were not supported or justified.	There was some organization to the ideas in the student work; however, ideas were generally not supported or justified.	The organization of the ideas in the student work was generally logical. Most ideas were supported and justified.	The organization of the ideas in the student work was very logical. All ideas were well supported and justified.



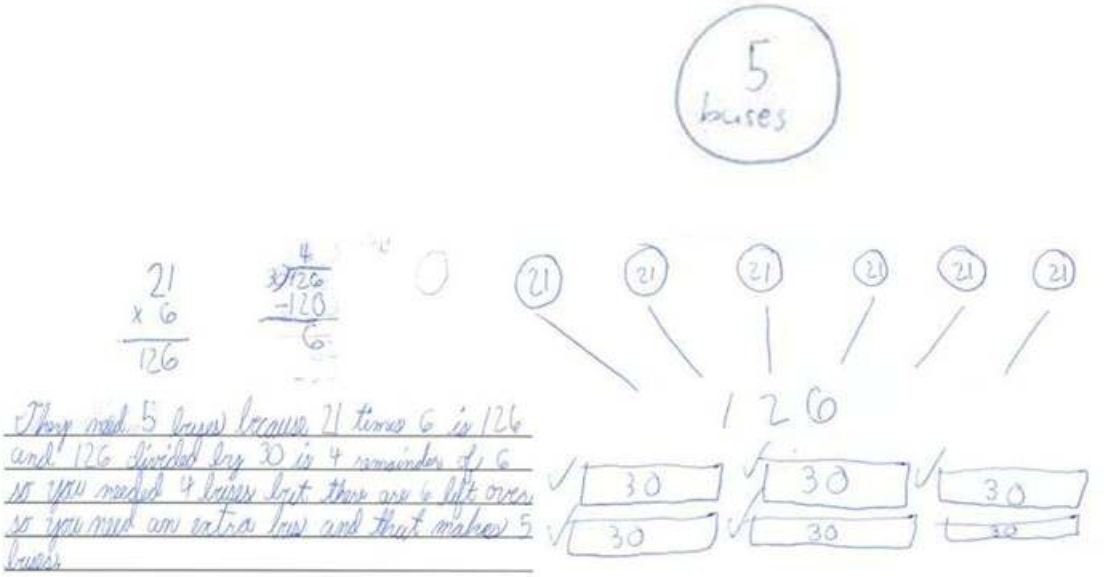
What is your overall score for the student's work on this problem?

Appendix B
Student Mathematics Work Samples

Sam's Work (Sample A)



Lauren's Work (Sample B)



Jeff's Work (Sample C)

I got an aunser of 87 bisses because 6×20= 180. 180 + 2 = 182. 182 ×30= 86.86+1=87.