ASSESSING FUTURE TEACHERS' MATHEMATICAL WRITING: AN ASSESSMENT TOOL AND INSIGHTS

Erin Smith
Univ. of Nevada Las Vegas
erin.smith@unlv.edu

R. Alex Smith
Univ. of Nevada Las Vegas
alex.smith@unlv.edu

Madeline Price
Univ. of Nevada Las Vegas
pricem10@unlv.nevada.edu

Sarah Powell
Univ. of Texas Austin
srpowell@austin.utexas.edu

Mathematical writing (MW) plays an important part in students' mathematical learning and is one way students are expected to communicate their thinking to others. However, for teachers to effectively support students in this work, they must have an understanding of and be able to generate high quality MW responses themselves. Yet, little is known about future teachers' (FT) MW competencies and ways to assess FTs' MW. In this study, we used a K12 MW assessment tool in undergraduate elementary mathematics methods courses to understand the utility of the tool and FTs' MW competencies.

Mathematical writing (MW) plays a critical role in mathematical learning because it promotes reflection and clarification of ideas via explanations, descriptions, definitions, and critiques and can promote students' mathematical identities (Boaler, 2002; Freeman et al., 2016; Ivanič, 1998; Marks & Mousley, 1990; National Council of Teachers of Mathematics, 2000). Yet, writing is often overlooked in mathematics classrooms and teacher preparation programs (Powell et al., 2021). If teachers are to effectively foster and scaffold students' MW, teachers must have an understanding of and be able to generate high quality MW responses themselves. However, little is known about future teachers' (FTs') MW competencies and ways to assess FTs' MW.

Mathematical Writing

MW is "a writing activity in which students write about mathematics concepts or procedures" (Powell et al., 2021, p. 418) that can vary based on purpose, formality, audience, structure, and required language (Chval et al., 2021). Within MW, multiple genres exist (e.g., explanatory, argumentative; Casa et al., 2016), however, explanatory writing is the primary focus of teachers, curriculum, and state assessments (Casa et al., 2019; Gillespie et al., 2014). Explanatory writing occurs when writers "reason about concepts" (Casa et al., 2016, p. 9) as they explain (e.g., strategies, connections) or describe (e.g., observations, representations). Given the prominence of explanatory writing in teachers' work with students, it is important to understand the extent of FTs' explanatory writing competencies in an effort to better prepare FTs to support students' explanatory MW.

Future Teachers' Mathematical Writing

Scholars (e.g., Magiera & Zambak, 2020) have called for more research focused on FTs' MW that goes beyond proof-related work. In the small body of work that focuses on FTs' MW outside of proof, much of this examines how FTs' reflective writing impacts metacognition, beliefs about mathematics, and pedagogical content knowledge (Kilic & Dogan, 2020; Kuzle, 2013; Namakshi et al., 2022). Findings from the few studies that have examined FTs' explanatory MW indicate FTs' conceptual understanding is strongly related to their own MW and their ability to critique students' MW (Magiera & Zambak, 2020). The majority of research that examined FTs' MW in general has focused upon FTs' mathematical content knowledge with little examination of their MW mechanics or organization.

Assessing Future Teachers' Mathematical Writing

We are aware of no broadly available assessment tools designed to evaluate FTs' MW competencies. As a result, scholars have developed their own scoring frameworks for their respective studies (e.g., Magiera & Zambak, 2020). Although these assessment tools have served their purposes in the context of a specific study, they have not had their technical adequacy (e.g., reliability and validity) examined nor been applied beyond the context of a specific study. Furthermore, these assessment tools often omit writing and writing mechanics assessment (e.g., organization, grammar), which can impact the perceived quality of mathematical understanding and inform future instruction.

Within the context of K12 mathematics classrooms, one group of scholars (i.e., Namkung et al., 2020) have sought to identify MW assessment tools that evaluate both mathematics and writing mechanics as well as demonstrate technical adequacy. Namkung et al. examined four potential assessment tools (i.e., holistic, analytic rubric, elements scoring, and MW sequences) for upper-elementary students. The authors determined that all four scoring methods were moderately correlated with criterion measures of mathematics and writing while confirmatory factor analyses indicated satisfactory construct validity. Furthermore, all tools demonstrated adequate reliability with the exception of increased variability for the grammar dimension of the analytic strategy. Although Namkung et al. recommended use of a holistic scoring approach, holistic scoring has minimal instructional utility as it does not indicate specific areas of need, has little use as formative assessment to measure growth across time, and is primarily useful only for summative evaluations. For formative assessment purposes and instructional utility, analytic scoring strategies may be the most beneficial. The analytic rubric employed by Namkung et al. demonstrated significant correlations with criterion measures of math fluency (*rs* = .33), word problem solving (*rs* = .66), and essay composition (*rs* = .36). The rubric was also scored efficiently with acceptable inter-rater reliability (IRR) and provided targeted information that could identify student strengths/needs in either mathematics content, writing, or both.

This study fills the aforementioned research gaps by (1) providing an assessment tool for MW and (2) shares insights into FTs' MW competencies.

Methods

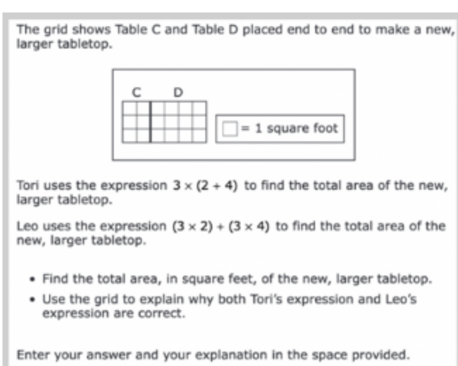
This study is part of a larger project examining the impact of a MW module on FTs' MW and their ability to assess elementary students' MW. Data was collected from 119 FTs at three U.S. universities. All FTs in this study were junior or senior undergraduates enrolled in a mathematics methods course that was designed to prepare elementary FTs seeking initial licensure in elementary education, special education, or elementary and special education.

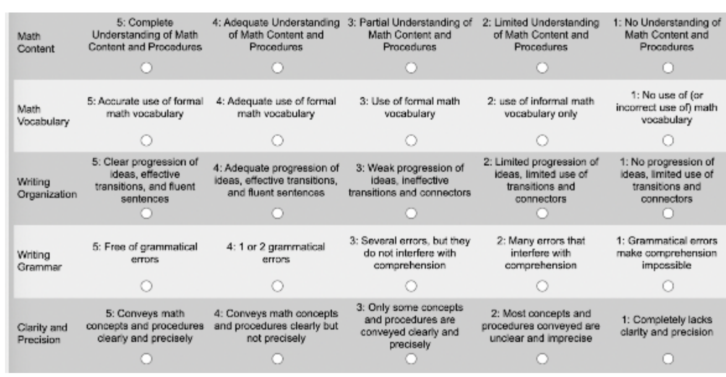
Mathematical Writing Task

Data for this study was composed of FTs' responses to an elementary grade level task completed prior to engaging with an online module on MW. The task was a third-grade released assessment item from New Jersey (see Figure 1). The audience for these responses were the instructors of the FTs' methods course.

Figure 1

Third Grade Released Assessment Item Given to Future Teachers and Scoring Rubric





Math Content	5: Complete Understanding of Math Content and Procedures	4: Adequate Understanding of Math Content and Procedures	3: Partial Understanding of Math Content and Procedures	2: Limited Understanding of Math Content and Procedures	1: No Understanding of Math Content and Procedures
Math Vocabulary	5: Accurate use of formal math vocabulary	4: Adequate use of formal math vocabulary	3: Use of formal math vocabulary	2: use of informal math vocabulary only	1: No use of (or incorrect use of) math vocabulary
Writing Organization	5: Clear progression of ideas, effective transitions, and fluent sentences	4: Adequate progression of ideas, effective transitions, and fluent sentences	3: Weak progression of ideas, ineffective transitions and connectors	2: Limited progression of ideas, limited use of transitions and connectors	1: No progression of ideas, limited use of transitions and connectors
Writing Grammar	5: Free of grammatical errors	4: 1 or 2 grammatical errors	3: Several errors, but they do not interfere with comprehension	2: Many errors that interfere with comprehension	1: Grammatical errors make comprehension impossible
Clarity and Precision	5: Conveys math concepts and procedures clearly and precisely	4: Conveys math concepts and procedures clearly but not precisely	3: Only some concepts and procedures are conveyed clearly and precisely	2: Most concepts and procedures conveyed are unclear and imprecise	1: Completely lacks clarity and precision

Analysis

FTs' MW was scored using an analytic rubric adapted from Namkung et al. (2020) that assesses MW across five dimensions (math content, math vocabulary, writing organization, writing grammar, clarity and precision) with a rating from 5 (ceiling) to 1 (floor) (see Figure 1). Clarity and precision were added to Namkung et al.'s rubric based on expert feedback. Two researchers with over 20 hours experience with the MW rubric independently scored all FT responses. Consensus IRR and correlations between rubric dimensions are reported. FT mean, standard deviation, and frequency of rubric performance within each dimension was analyzed for patterns of strengths and weaknesses.

Findings

MW Assessment Properties

We report the consensus approach (percent of exact agreement) and consensus within 1-point (indicates scorers assessed in the same range) for IRR, where 70% or better is considered acceptable (Namkung et al., 2020). IRR for math content was .59 (.97 within 1-point), math vocabulary was .47 (.83 within 1-point), writing organization was .50 (.93 within 1-point), writing grammar was .54 (.93 within 1-point), and writing clarity and precision was .50 (.89 within 1-point). The exact agreement IRR was consistently lower than Namkung et al. (2020) and well below 70%, but IRR within 1-point approached 90%. Although IRR rates here are not suitable for high-stakes decision-making, the within 1-point IRR rates were similar to those found with K12 students and allow for broad interpretations for formative assessment purposes. The analytic rubric was not time intensive and could be used in a typical classroom.

Correlations between the rubric dimensions and text segments (words, values, and expressions) are shown in Table 1. In general, correlations between dimensions were moderately strong to strong and suggest the rubric assesses highly related but distinct dimensions of MW, which could be useful for differentiating instruction. Mathematics vocabulary was correlated more strongly with writing clarity and precision (.71) than mathematics content (.59), while mathematics content correlated the strongest with writing clarity and precision (.85). The low correlations for writing grammar indicate this dimension needs revision. Although grammar is important, and perhaps even more so for younger students, it may be best assessed as an "acceptable/not-acceptable" checkbox on the rubric rather than a scaled score. One potential revision is to replace the writing grammar dimension with text segments.

Table 1.

Correlation Table of Mathematics Writing Assessment Dimensions

Dimension	1	2	3	4	5	6	7
1 Mathematics content	*	.59	.80	.11	.85	.89	.68

2 Mathematics vocabulary	.59	*	.61	.24	.71	.83	.43
3 Writing organization	.80	.61	*	.16	.82	.88	.73
4 Grammar	.11	.24	.16	*	.17	.35	.06
5 Clarity and precision	.85	.71	.82	.17	*	.92	.65

FTs' Performance

Generally speaking, FTs scored similarly across dimensions. Mean scores (see Table 2) across all dimensions indicate FTs' have a partial understanding of the mathematics content embedded within the task (i.e., operations, number) and are not yet able to construct a written response that demonstrates mastery in mathematics *and* writing. Consequently, there was not a ceiling effect on the FTs' MW scores even though the task was at a third-grade level.

Table 2

Mean, Standard Deviation, and Frequency of Assessment Scores for Future Teachers' Mathematics Writing

Dimension	Mean (SD)	Frequency				
		1	2	3	4	5
Mathematics content	3.03 (1.09)	4	37	27	31	20
Mathematics vocabulary	2.66 (1.00)	13	29	44	28	5
Writing organization	2.92 (1.13)	14	19	40	30	16
Writing grammar	3.73 (0.83)	3	3	27	53	33
Clarity and precision	2.54 (1.03)	15	38	35	25	6

Note: To calculate frequency we rounded the average score to the nearest whole number.

When looking across the students who scored a 5 on content, their responses followed a typical structure of first identifying the total square feet of the table top (first prompt) followed by the FT's reasoning of why Tori's and Leo's expressions matched the situation. These responses often included an explicit reference to the two tables represented in the grid and discussed the equivalence of the two expressions, similar to the following

> The total area of the tabletops is 18 sq. ft. Both students have written a correct expression for the problem. Tori's expression focuses on the tabletops as a whole. She said the width (3) times the length (2+4). This expression equals 18 sq. ft. Leo's expression focuses on the tabletops as individuals. He has expressions for the area of each tabletop, and then he adds those expressions together. The answer is 18 sq. ft. If a person counts the squares on the inside of the tabletops, they would find the answer 18 sq. ft.

In contrast, FTs who scored a 1 or 1.5 on mathematics content typically did not answer both prompts embedded in the question and their response lacked specificity (e.g., "Both of their answers are correct because ultimately they would both end up with 6x3 which is 18"). This appears to be somewhat related to scores received on mathematics vocabulary, since responses that received a 1 on mathematics vocabulary were brief and included little, if any, mathematics vocabulary (e.g., square feet, length, width). Differently, responses that were scored a 5 were lengthier and included frequent use of mathematics vocabulary and symbolic notation, such as the example above. Consequently, it seems that FTs who are able to produce more text generally performed better than peers who wrote less.

There was an average of 46 text segments and a range of 1 to 288. Of all respondents, only five FTs provided a text segment of 3 or less (e.g., "18"). When FT's overall score was rounded to the nearest whole number, FTs' who scored a 4 or more wrote an average of 83 text segments whereas those who scored a 2 or below averaged 18 text segments.

FTs' performed lowest on the clarity and precision dimension, with the majority of FTs earning between a 2 or 3. Responses in this range were typically of average length, complete, and used mathematics vocabulary. However, what was different about these prompts compared to those receiving higher scores was a lack of precision, like

> The answer is 18. Tori used the height of the grid to come up with 3, and she added the length of table C and table D together because they make up one side of the bigger tabletop. The width (3) is then multiplied by the total of table C and D. Leo uses a similar strategy, but he multiplies table C and D by the width separately.

In this example, the FT lacked precision and clarity in their explanation by leaving off units (e.g., square feet) and did not connect their explanation with the expressions.

Discussion & Conclusion

In this study, we used a MW assessment tool previously identified for use in K12 school settings in the context of undergraduate elementary mathematics methods courses at three U.S. institutions. The analytic rubric used was efficient and demonstrated acceptable within 1 point IRR for formative assessment purposes, but the rubric needs revision for further research or high-stakes decision making. Yet, use of a common rubric in MW, or a close facsimile, can promote generalization of findings across research studies so that the field may develop a better understanding of what FTs and K12 students need to be successful with MW. Future studies may consider examining whether or not the rubric is responsive to student learning (i.e., sensitive to growth), what kinds of instructional decisions teachers make based upon student performance on the rubric, and if the rubric can be readily used by students to self-evaluate and revise their MW.

Findings from this study indicate that FTs' need additional instruction in their teacher preparation program on mathematical concepts and vocabulary—some of which they had already received in their teacher preparation program (e.g., operations and number)—along with clarity and precision in writing. Given prior research, we hypothesized FTs would likely perform better on the writing dimensions (at or near ceiling) than the mathematics dimensions and were somewhat surprised when this did not occur. Such findings highlight the need to simultaneously bolster FTs' writing competencies alongside mathematics. To support FTs in crafting MW responses, instruction should incorporate writing strategies, such as mnemonic strategies (e.g., PSOLVE, Hauk & Isom, 2009) and graphic organizers, alongside mathematics while leveraging FTs' literacy practices. Furthermore, the planning and organization components inherent in effective writing instruction (e.g., PSOLVE, Hauk & Isom, 2016) could also support students' problem-solving skills. Furthermore, instruction focused around MW should include explicit attention to identifying what academic language is embedded in task statements and why it is important to use academic language to clearly, coherently, and precisely communicate mathematical thinking to others.

Since MW is a form of mathematical discourse and represents one way students communicate their mathematical understandings, it is critical FTs can generate high-quality MW responses if they are to support students in this work. Moreover, given the prominence of MW on standardized high-stakes assessments, ensuring FTs are sufficiently prepared to support students is a matter of educational equity.

References

Boaler, J. (2002). The development of disciplinary relationships: knowledge, practice and identity in mathematics classrooms. *For the Learning of Mathematics, 22*(1), 42–47.

Casa, T. M., Firmender, J. M., Cahill, J., Cardetti, F., Choppin, J. M., Cohen, J., …Zawodniak, R. (2016). Types of and purposes for elementary mathematical writing: Task force recommendations. Retrieved from https://mathwriting.education.uconn.edu/wp-content/uploads/sites/1454/2016/04/Types_of_and_Purposes_for_Elementary_Mathematical_Writing_for_Web-2.pdf

Casa, T. M., MacSwan, J. R., LaMonica, K. E., Colonnese, M. W., & Firmender, J. M. (2019). An analysis of the amount and characteristics of writing prompts in Grade 3 mathematics student books. *School Science and Mathematics*, *119*(4), 176–189.

Chval, K. B., Smith, E., Trigos-Carrillo, L., & Pinnow, R. J. (2021). *Teaching Math to Multilingual Students, Grades K-8: Positioning English Learners for Success.* Corwin.

Freeman, B., Higgins, K. N., & Horney, M. (2016). How students communicate mathematical ideas: an examination of multimodal writing using digital technologies. *Contemporary Educational Technology, 7*(4), 281-313.

Gillespie, A., Graham, S., Kiuhara, S., & Hebert, M. (2014). High school teachers' use of writing to support students' learning: A national survey. *Reading and Writing, 27*, 1043−1072.

Hauk, S., & Isom, M. A. (2009). Fostering college students' autonomy in written mathematical justification. *Investigations in Mathematics Learning, 2*(1), 49–78.

Ivanič, R. (1998). *Writing and identity.* John Benjamins.

Kilic, H., & Dogan, O. (2022). Preservice mathematics teachers' noticing in action and in reflection. *International Journal of Science and Mathematics Education*, *20*(2), 345-366.

Kuzle, A. (2013). Promoting writing in mathematics: Prospective teachers' experiences and perspectives on the process of writing when doing mathematics as problem solving. *Center for Educational Policy Studies Journal, 3*(4), 41–59. https://doi.org/10.26529/cepsj.222

Marks, G., & Mousley, J. (1990). Mathematics education and genre: Dare we make the process writing mistake again? *Language and Education, 4*(2), 117–135.

Namakshi, N., Warshauer, H. K., Strickland, S., & McMahon, L. (2022). Investigating preservice teachers' assessment skills: Relating aspects of teacher noticing and content knowledge for assessing student thinking in written work. *School Science and Mathematics*, *122*(3), 142-154.

Namkung, J. M., Hebert, M., Powell, S. R., Hoins, M., Bricko, N., & Torchia, M. (2020). Comparing and validating four methods for scoring mathematics writing. *Reading & Writing Quarterly, 36*(2), 157–175. https://doi.org/10.1080/10573569.2019.1700858

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics* (Vol. 1). Reston, VA: Author.

Powell, S. R., Hebert, M. A., & Hughes, E. M. (2021). How educators use mathematics writing in the classroom: A national survey of mathematics educators. *Reading and Writing* (Vol. 34). Springer Netherlands. https://doi.org/10.1007/s11145-020-10076-8