Fall 10-19-2012

Student Self-Assessment in Middle School Mathematics: A Pilot Study

Zachary B. Warner
University at Albany, SUNY, zwarner@albany.edu

Fei Chen
University at Albany, fchen3@albany.edu

Heidi Andrade
University at Albany, handrade@albany.edu

Student Self-Assessment in Middle School Mathematics: A Pilot Study

Zachary B. Warner, Fei Chen, and Heidi Andrade

University at Albany, SUNY

Abstract

The aims of this study were to design a process of self-assessment for seventh grade mathematics students and test its effectiveness in improving performance. The treatment required the students to co-create the product quality criteria for constructed response items, use the criteria to formatively self-assess their work, and revise as needed. Although there was a statistically significant association between self-assessment and performance on only one of five items, the treatment group had a higher average score on three items and total score after controlling for prior achievement. The findings suggest that research with larger sample sizes is warranted.

Keywords: formative assessment, self-assessment, middle school mathematics

Student Self-Assessment in Middle School Mathematics: A Pilot Study

Self-assessment is a process during which students reflect on the quality of their work, compare it to explicitly stated criteria, judge how well their work reflects the criteria, and make appropriate revisions (Andrade, 2010). The emphasis here is on the word *formative*: Self-assessment is done on drafts of works in progress in order to inform revision and improvement; it is not a matter of having students determining their own grades. The study reported in this paper explored the relationship between middle school student's self-assessment experience and their performance on a summative mathematic test. The study investigated the following research questions:

1. Is self-assessment feasible in a middle school mathematics classroom?
2. Is the process for rubric-referenced self-assessment in writing (i.e., view a model, generate criteria, score with a rubric, revise) transferable to mathematics?
3. What is the direction and size of the effect on achievement when using self-assessment?

Literature Review

Self-Assessment

Student self-assessment is grounded in scholarship on self-regulated learning and feedback in learning. Self-regulated learning is the process whereby learners set goals for their learning and then attempt to monitor, regulate, and control their cognition, motivation, and behavior in order to reach their goals (Pintrich, 2000). A well-developed body of research suggests that self-regulation and academic achievement are closely related: Students who set goals, make flexible plans to meet them, and monitor their progress tend to learn more and do better in school than those who do not. Less effective learners, in contrast, have minimal self-regulation strategies and depend much more on external factors such as the teacher, peers, or the task for guidance and feedback (Pintrich, 2000; Zimmerman & Schunk, 2011). Research has also demonstrated that academic self-regulation is learnable. Studies have shown that children, including those with mild to moderate cognitive impairments (Brown & Palincsar, 1982), can learn to monitor and regulate their own learning more effectively.

A central purpose of both self-assessment and self-regulation is to provide learners with feedback they can use to deepen their understandings and improve their performances. Research has shown that, when effectively delivered, feedback tends to promote learning and achievement (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Butler & Winne, 1995; Crooks, 1988; Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Shute, 2008), yet most students get little informative feedback on their work (Black &Wiliam, 1998). The scarcity of feedback in most classrooms is due, in large part, to the fact that few teachers have the time to respond as often as they would like to each student's work. Fortunately, research shows that students themselves can be useful sources of feedback via self-assessment (Andrade, Du, & Mycek, 2010; Andrade, Du, & Wang, 2008; Ross, Rolheiser, & Hogaboam-Gray, 1999), particularly when the feedback provides task-specific information about the goals and processes of their learning, where they are in relation to those goals and processes, and what they can do to close the gap (Wiliam & Thompson, 2007).

Thoughtful self-assessment is often scaffolded by a rubric or checklist. Rubrics have become popular with teachers as a means of communicating expectations for an assignment, providing focused feedback on works in progress, and grading final products (Andrade, 2000; Jonsson & Svingby, 2007; Moskal, 2003; Popham, 1997). Although educators tend to define the word rubric in slightly different ways, a commonly accepted definition 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 checklist is a similar assessment tool that defines the scoring criteria but only allows for acknowledgement that a criterion has been met or not met; levels of quality are not described. Checklists can serve similar purposes to those of rubrics.

Rubric- or checklist-referenced self-assessment is a process of formative assessment in which students use a rubric or checklist to guide their judgments about the quality of their own work. The process begins with teachers and students viewing a model assignment, discussing its strengths and weaknesses, and co-creating scoring criteria. Next, students complete the assignment and self-assess their work using a careful, stepwise process guided by the rubric or checklist. Finally, students revise and improve their work. One of the major benefits of this type of self-assessment is that the process is the same for all student populations and can help all students to become more self-directed (Andrade & Warner, 2012).

Self-Assessment in Mathematics

There is some evidence that formative assessment has the potential to help students develop the skills they need to reason and communicate mathematically. Brookhart, Andolina, Zuza, and Furman (2004) found that third grade students' involvement in their own assessment improved not only their mathematics performance but also narrowed the gap between their predicted and actual performance. Ross, Hogaboam-Gray, and Rolheiser (2002) report positive results of self-assessment of word problem solutions in fifth and sixth grade mathematics classes. Working with older mathematics students, Stallings and Tascione (1996) found that the process of self-assessment can "engage students in evaluating their progress, aid in developing their communication skills, and increase their mathematics vocabulary" (p. 548).

International research also demonstrates positive effects of self-assessment on student performance in mathematics, both in formative and summative assessments. Fontana and

Fernandes (1994) found improvements in performance by Portuguese students ages eight to 14 when self-assessment was part of a multi-faceted treatment. Black, Harrison, Lee, Marshall and Wiliam's (2004) study of formative assessment practices in math and science classes for 11-15 year olds in England revealed a strong, positive relationship between formative assessment, including self-assessment and achievement. These authors concluded that "the development of self-assessment by the student might have to be an important feature of any programme of formative assessment" (p. 14). There is even some evidence that self-assessment is associated with improved performance on summative external examinations: A study of the influence of self-assessment training on high school students' scores on external exams in the Bahamas showed a consistently positive effect of treatment (MacDonald & Boud, 2003).

The primary goal of this research was to investigate student self-assessment in the mathematics classroom. Grade seven was chosen because of the crucial role of the middle school years in students' mathematical development. Middle school represent an important transition point as students enter early adolescence and develop conceptions of themselves as mathematical learners. Their self-images may be positive or negative, depending on their experience, and their internal beliefs will affect their attitudes and performance in mathematics in the ensuing years, potentially influencing career and life opportunities (National Council of Teachers of Mathematics, 2000).

Methodology

Participants

The participants were 49 seventh grade students in a public middle school in upstate New York. Classes were randomly assigned into the treatment (n = 24) or comparison (n = 25) group.

Prior achievement was evaluated for both groups using the previous year's mathematics course grades on a 0.0 - 4.0 scale (treatment grade mean = 3.1, comparison grade mean = 3.0).

Procedures

Preliminary field tests of a complete rubric revealed that the rubric was unnecessarily complex for the kinds of problem solving to be done by the seventh graders in the study: A checklist designed with the classroom teacher was sufficient. In class, the purposes and features of the self-assessment process were discussed and modeled for students in the treatment group, and students' questions and concerns were addressed. The students were then engaged in a process of co-constructing product criteria for constructed-response items by analyzing one high and one low quality solution to sample problems. The resulting criteria were then included in the checklist students used to self-assess their work on new items. The treatment students were presented with items and asked to solve them, self-assess according to the partially co-created Math Process and Product Checklist (Appendix A), and revise as necessary. Students were reminded to make use of the Math Process and Product Checklist by marking different symbols for each item. Students in the comparison group were given the same items and asked to complete them as they would under the usual conditions, with no checklist or formal self-assessment.

Instruments and Measures

Math process and product checklist. The checklist used by students in the treatment condition to self-assess their work contained both process and product criteria. Based on the New York State standards and research on how students approach problem solving, the process components of the instrument were developed and refined by the authors in collaboration with the participating teacher. The product criteria listed in the checklist were co-created by the teacher together with the students in the treatment group.

Test items. Students worked out solutions to five geometry items from previous state tests that aligned with the New York State Grade 7 mathematics curriculum. Items were selected based on their relationship to the current classroom unit and their utility for observing and measuring self-assessment (i.e., constructed-response items). The items used in this study required students to find the volume and surface area of solid figures.

Scoring rubrics. Operational test items for NYS are typically scored holistically. Because this study sought information regarding practical as well as statistical differences, analytic scoring rubrics that addressed each aspect of an item (e.g., concept, calculation, labeling) were developed by content experts using model rubrics from NYS Regents examinations in mathematics. The rubrics had a range of 0-5 points for four items and 0-2 points for one item (Item 5).

Results

Descriptive statistics are presented in Table 1. Adjusted means by experimental condition are presented in Table 2. One-way analysis of covariance was used to analyze the data. Prior achievement was statistically controlled while investigating differences in item scores. Treatment or control assignment served as the independent variable. In separate analyses, the score for each item (i.e., Item 1, Item 2, etc.) as well as the total score for all five items were investigated as dependent variables. The results of each ANCOVA are presented in Table 3.

Table 1

Descriptive Statistics for Items 1-5 and Total Score

Variable	**N**	**Score Range**	**Mean**	**SD**
Item 1	49	0 – 5	3.94	1.162
Item 2	49	0 – 5	3.00	1.837
Item 3	49	0 – 5	4.14	1.291
Item 4	49	0 – 5	4.16	1.143
Item 5	49	0 – 2	1.10	0.586
Total Score	49	3 – 22	16.35	4.381

Table 2

Adjusted Means for Items 1-5 and Total Score Treatment (Self-Assessment) and Comparison Groups

	Adjusted Means	
	Self-Assessment	Comparison
Item 1	4.07	4.11
Item 2	3.10	3.17
Item 3	4.57	4.10
Item 4	4.31	4.21
Item 5	1.35	1.01
Total Score	17.40	16.60

Only item 5 showed statistically significant differences between the treatment and comparison groups. The significance is likely due to higher relative variability: Because only two score points were available for Item 5, missing a label or making a minor calculation error (both one-point deductions) had a larger impact on the overall score for the item. However, the treatment group's higher scores on items 3 and 4 and the total score (see Table 2), although not statistically significant, suggest that further investigation with a larger sample size might yield statistically, and more importantly, practically significant differences between students engaged in self-assessment and those following standard classroom procedures.

Table 3

ANCOVA Results for Self-Assessment Association with Student Achievement on Items 1-5 and Total Score, Controlling for Prior Achievement

Variable	Source	*SS*	*df*	*MS*	*F*
Item 1	Regression	3.46	1	3.46	3.71
	Self-Assessment	0.02	1	0.02	0.02
	Error	40.19	43	0.93	
Item 2	Regression	37.07	1	37.07	15.03***
	Self-Assessment	0.06	1	0.06	0.02
	Error	106.04	43	2.47	
Item 3	Regression	1.66	1	1.66	1.57
	Self-Assessment	2.53	1	2.53	2.39

	Error	45.49	43	1.06	
Item 4	Regression	0.07	1	0.07	0.07
	Self-Assessment	0.12	1	0.12	0.12
	Error	42.66	43	0.99	
Item 5	Regression	0.74	1	0.74	0.87
	Self-Assessment	1.31	1	1.31	5.45*
	Error	10.35	43	0.24	
Total Score	Regression	107.49	1	107.49	9.63**
	Self-Assessment	7.26	1	7.26	0.65
	Error	479.92	43	11.16	

* $p < 0.05$
** $p < 0.01$
*** $p < 0.001$

Conclusions and Educational Implications

This pilot study investigated the conditions necessary for successful student self-assessment in middle school mathematics. With the help of a participating teacher, a checklist containing mathematical process and product criteria was developed to guide students through the self-assessment process. The product criteria were co-created with the students. After going over the assignment, students were engaged in a process of partially co-creating their scoring guide for self-assessment by analyzing model solutions to sample problems in terms of their strengths. The strengths identified by students and the teacher and researcher were integrated into the checklist as criteria for successful performance in mathematics. The Math Process and Product Checklist experienced iterative refinement after field testing items and self-assessments to collect information about their empirical feasibility and effectiveness.

In response to the first research question about the feasibility of rubric-referenced self-assessment in middle school mathematics, the results of this study show that this form of self-assessment is feasible even given time, curriculum, and other typical classroom constraints. However, the process of self-assessment should be more highly scaffolded in order to ensure fidelity of treatment, perhaps by asking students to use colored pencils to mark on their answer sheets the evidence that the criteria on the checklist are met. Further study is needed to outline best practices for criteria-referenced self-assessment in mathematics but the pilot study suggests that this is a promising line of inquiry.

The second research question examined the process used in previous, full-scale studies of self-assessment in writing (e.g., Andrade, Du, & Wang, 2008; Andrade, Du, & Mycek, 2010) in terms of transferability to mathematics. Students in this pilot study reviewed a model, co-created scoring criteria, and used a checklist to self-assess, then revised as needed. This was in line with the process employed in the writing studies. Thus, it can be concluded that the process for self-assessment in middle school mathematics can look like effective self-assessment in other content areas: The process appears to be adaptable across academic subjects.

The final research question inquired about the magnitude and direction of the effect of self-assessment on achievement. The results of the pilot study suggest some small differences in performance between students who self-assessed and those that did not with the students engaging in self-assessment scoring slightly higher on three out of five items and for the total score. Only one item showed statistical significance for the differences between groups.

However, from the perspective of practical significance, these finding suggest that further, larger-scale investigations are appropriate. The differences in the total score variable translate to approximately one-half of a letter grade (e.g., B to B+). Overall, the results of the pilot study imply that self-assessment in mathematics classes could have the potential to promote learning and achievement.

The study has a limitation typical of most pilot studies; its small sample size. The homogeneity of student participants from one single school also prevent us from drawing conclusions that can be generalized to a middle school student population. The absence of demographic information and other covariates that would allow for more thorough statistical analysis constitute another limitation. As noted above, however, further investigation into student self-assessment is warranted. Given the weight placed on student achievement in mathematics (CCSSI, 2010; NCLB, 2002), an extension of the principles of self-assessment to mathematics instruction seems timely and appropriate. Giving students the perspective of a rater allows them to approach their work with the assessment criteria in mind. This can lead to products that are more aligned with the intended objectives of the assignment and, we predict, higher quality products and student achievement.

References

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

Andrade, H. (2010). Students as the definitive source of formative assessment: Academic self-assessment and the self-regulation of learning. In H. Andrade & G. Cizek (Eds.), *Handbook of formative assessment* (pp. 90-105). New York: Routledge.

Andrade, H., & Warner, Z. (2012). Beyond “I give myself an A”: Meaningful, rubric-referenced student self-assessment. *Educator’s Voice, V*(42), 42-51.

Andrade, H., Du, Y., & Mycek, K. (2010). Rubric-referenced self-assessment and middle school students’ writing. *Assessment in Education, 17*(2), 199-214.

Andrade, H., Du, Y., & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students’ writing. *Educational Measurement: Issues and Practices, 27*(2), 3-13.

Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*(2), 213–238.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). *Assessment for learning: Putting it into practice.* Berkshire, England: Open University Press.

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80(2), 139-148.*

Brookhart, S. M., Andolina, M., Zuza, M., & Furman, R. (2004). Minute math: An action research study of student self-assessment. *Educational Studies in Mathematics*, *57*(2), 213-227.

Brown A. L., & Palincsar, A. S. (1982). Inducing strategic learning from texts by means of informed, self-control training. *Topics in Learning and Learning Disabilities, 2*(1), 1-17.

Butler, D., & Winne, P. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research,65*(3), 245-281.

Common Core State Standards Initiative (2010). Common core state standards for mathematics. Available from: http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf

Crooks, T. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58*(4), 438-481.

Fontana, D., & Fernandes, M. (1994). Improvements in math performance as a consequence of self-assessment in Portuguese primary school pupils. *British Journal of Educational Psychology*, *64*, 407-417.

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*(2), 130–44.

Kluger, A., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254-284.

MacDonald, B., & Boud, D. (2003). The impact of self-assessment on achievement: The effects of self-assessment training on performance in external examinations. *Assessment in Education, 10*(2), 209-220.

Moskal, B., & Leydens, J. (2000). Scoring rubric development: validity and reliability. *Practical Assessment, Research & Evaluation, 7*(10). Available at http://PAREonline.net/getvn.asp?v=7&n=10.

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

No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002).

Pintrich, P. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 452-502). San Diego, CA: Academic.

Popham, J. W. (1997). What's wrong – and what's right – with rubrics. *Educational Leadership, 55*(2), 72–5.

Ross, J. A., Hogaboam-Gray, A., & Rolheiser, C. (2002). Student self-evaluation in grade 5-6 mathematics effects on problem-solving achievement. *Educational Assessment, 8*(1), 43-59.

Ross, J., Rolheiser, C., & Hogaboam-Gray, A. (1999). Effects of self-evaluation training on narrative writing. *Assessing Writing, 6*(1), 107-132.

Shute, V. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153-189.

Stallings, V., & Tascione, C. (1996). Student self-assessment and self-evaluation. *Mathematics Teacher, 89*(7), 548-55.

Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53-82). Mahwah, NJ: Erlbaum.

Zimmerman, B., & Schunk, D. (Eds.) (2011). *Handbook of self-regulation of learning and performances.* New York: Routledge.

Appendix A

1	**Understand the task**	I can clearly state what the problem is asking me to find.
2	**Explain what is known**	I can clearly explain the given information (what I know from the problem). I use words, numbers, and diagrams as appropriate.
3	**Plan an approach**	I can clearly describe my chosen strategy, which is efficient and sophisticated (e.g., "I will make a table," "make an organized list", "draw a diagram,").
4	**Solve the problem**	I use my plan to solve every part of the problem. If my strategy doesn't work, I try a new one. I write out all the steps in my solution so the reader doesn't have to guess at how or why I did what I did. I use words, numbers, and diagrams/charts/graphs, as appropriate. My work is clearly labeled.
5	**Explain the solution**	I clearly explain my solution and why I believe it is correct using precise and correct math terms and notations. I check to make sure my solution is reasonable. I check for possible flaws in my reasoning or my computations. If I can, I solve the problem in a different way and get the same answer.
6	**Check the solution**	I check my solution according to the scoring criteria below. **Scoring Criteria:** (✓) _______ Appropriate formula _______ Diagram with clear labels (if appropriate) _______ All work shown and connected to final answer _______ Correct calculations and order of operations _______ Final answer clearly identified _______ Answer labeled with units (if appropriate) _______ Answer correctly rounded to the requested decimal place (if appropriate) If my solution is incorrect, I find my mistake, determine a new plan, solve the problem, and justify my new answer.