In the transition to middle school, and during the middle school years, students’ motivation for mathematics tends to decline from what it was during elementary school. Formative assessment strategies in mathematics can help support motivation by building confidence for challenging tasks. In this study, the authors developed and piloted a professional development program, Learning to Use Formative Assessment in Mathematics with the Assessment Work Sample Method (AWSM) to build middle school math teachers’ understanding of the characteristics of high-quality formative assessment processes and increases their ability to use them in their classrooms. AWSM proved to be feasible to implement in the middle school setting. It improved teachers’ practice of formative assessment, especially in their feedback practices, regardless of their pedagogical content knowledge at entry. Results from focus groups suggested that teachers were better able to implement ungraded practice and student self- and peer-assessment after AWSM, and that students were more willing to engage in complex problem solving.

In the transition to middle school, and during the middle school years, students’ motivation for mathematics tends to decline from what it was during elementary school. At this age, students report less valuing of mathematics and lower effort and persistence in mathematics problem solving over time (Pajares & Graham, 1999; Valas, 2001; Wigfield, Eccles, & Pintrich, 1996). Middle school students also report lower confidence in their mathematics ability than before (Clarke, Roche, Cheeseman, & van der Schans, 2014; Pintrich & Schunk, 1996), influenced in part by exposure to a larger peer group with whom they begin to compare themselves, more perceived competition in the classroom environment, and more rigorous standards for evaluation (Eccles & Midgley, 1989). Some students perceive that they are judged more on innate ability in middle school than on improvement (Anderman & Midgley, 1997), and come to believe that if they lack ability, their effort will not help (Valas, 2001). Students also lose confidence that they can self-regulate, in that they feel less able to organize their thoughts, understand complex tasks, and choose and evaluate problem-solving strategies (Pajares & Graham, 1999).

**Low Confidence in Mathematics Affects Teaching and Learning**

These feelings have an effect on teaching and learning in the middle school mathematics classroom. Students who are not confident that they can solve complex problems, or who do not see the point of putting forth effort to do so, try to avoid those tasks or pressure teachers to make the work simpler for them (Clarke et al., 2014). This lack of mathematics self-efficacy is a predictor of lower mathematics achievement outcomes (Pajares & Graham, 1999). Some middle schoolers will attempt to engage in mathematics learning only when tangible rewards are offered (Rowan-Kenyon, Swan, & Creager, 2012). Students’ difficulties with motivation and persistence, therefore, become a problem of practice for middle school teachers.

In this environment, middle school mathematics teachers can feel hesitant giving students challenging and complex work. If they do so anyway and students encounter difficulty, some teachers oversimplify the task or tell students how to solve it (Clarke et al., 2014; Ferguson, 2009). This is especially true when teachers work with lower-achieving students (Zohar, Degani, & Vaaknin, 2001). However, the Common Core and other contemporary U.S. mathematics standards require that students be able to solve complex problems and to explain their reasoning, so teachers need effective instructional strategies to support students in these practices.

**Instruction Can Build Confidence for Challenging Mathematics Tasks**

Research has supported some features of instruction and the classroom environment that make it easier for middle school mathematics teachers to pose challenging tasks and for students to engage, persist, and succeed in them. Rather than lower the level of cognitive demand of difficult problems, teachers have found success by supporting...
small-group discussions and peer collaboration (Ferguson, 2009; Rowan-Kenyon et al., 2012; Ryan & Patrick, 2001; Schunk & Pajares, 2001; Zohar et al., 2001). When teachers demonstrated to students that they are supportive and will try to help students as individuals, and when teachers had clear expectations that students would respect one another’s ideas in the classroom, students were less disruptive and better able to self-regulate (Ryan & Patrick, 2001; Valas, 2001). A classroom focused more on mastery and improvement rather than competition also supported students’ academic confidence (Ryan & Patrick, 2001).

Formative Assessment Strategies Can Help Students Be More Confident Mathematics Learners

Formative assessment is an evidence-based process of gathering information on three questions: (a) Where am I going? (b) How am I doing now? (c) Where do I go next? to support a learning cycle (Hattie & Timperley, 2007; Sadler, 1989). Therefore, the most important formative assessment practices involve (a) students’ understanding of their learning goals and targets, (b) the criteria by which they will know how they are progressing with their learning, and (c) what needs to be done next to move learning forward. Feedback is an active part of the process and can address the task, the student’s processing of the task, suggestions for what to work on next, and scaffolds for the individual student (Hattie & Timperley, 2007). Literature supports prioritizing these three dimensions of formative assessment (see Figure 1).

Assessment-centered teaching can have profound effects on student learning and motivation (DiRanna et al., 2008). Well-designed formative assessment is associated with major gains in student achievement across all ages and subjects, and has its greatest positive impact on students who struggle in mathematics (Black & Wiliam, 1998a, 1998b). To understand learning targets, students need to have clear knowledge of where they are going with their learning and avoid the inefficiency and frustration of trial and error (Sadler, 1989). Students who struggle in school particularly benefit from understanding learning targets, because they do not intuit the targets on their own (Black & Wiliam, 1998a, 1998b; James et al., 2006). Clear success criteria enable students to understand how they are doing on learning targets, so they can judge and self-regulate the quality of their work (e.g., Andrade, Du, & Mycek, 2010; Brookhart, Andolina, Zusa, & Furman, 2004; Sadler, 1989). When teachers involve students in the assessment process, students perceive more control of and more responsibility for their own learning (Rieg, 2007). Allowing students to help generate success criteria gives them a feeling of empowerment and makes assessment of their work seem less punitive and more constructive (Brookhart, 1997). In turn, students feel more competent and are more likely to engage in learning (Stiggins, Arter, Chappuis, & Chappuis, 2006).

Formative Assessment Is Particularly Difficult in Mathematics Without Support

Formative assessment requires teachers to elicit students’ existing ideas as students make their thinking visible to help further understanding. In mathematics, it is often difficult to interpret students’ understanding, knowledge, and learning of complex content. Interpretations of student work as evidence of learning may be powerfully influenced by teachers’ own understandings of mathematics concepts and how the concepts can be communicated, and expectations.
regarding the performance of whole classrooms and individual students (Even, 2005; Morgan & Watson, 2002). 

**Professional Development in Formative Assessment in Mathematics Needs Improvement**

Much of current teacher professional development (PD) in formative assessment is delivered in the form of one-time workshops, which are ineffective at changing teachers’ day-to-day professional practices (McLaughlin & Talbert, 2006). Even more extensive PD that includes books and videos does not necessarily help teachers change their practice. In a study of one such program, implementation was low, and while teacher knowledge grew, teacher practice did not change (Randel, Beesley, Apthorp, Clark, & Wang, 2016; Randel et al., 2011). Assessment work sample method (AWSM) is intended to support mathematics teachers in using, not just learning about, high-quality formative assessment strategies linked to the overall formative assessment process.

**Program**

*Learning to Use Formative Assessment in Mathematics with the AWSM* is a PD program that builds middle school mathematics teachers’ understanding of the characteristics of high-quality formative assessment processes and increases their ability to use them in their classrooms. We were inspired to create AWSM following the results from previous research (Randel et al., 2011, 2016) on a formative assessment PD program developed for all content areas and levels of learning. This program did not change teacher practice in mathematics, in part because it had few examples of how to implement formative assessment in that content area. By contrast, AWSM provides PD that builds formative assessment practices and skills specifically in mathematics. AWSM is about helping teachers develop effective instructional practices embedded in everyday teacher lessons and uses actual student and teacher work samples to connect the PD to content, instructional practice, and local context. The sessions aim to prompt teachers to examine their current assumptions and decide which aspects of their practice to change, provide time to learn new approaches, and stimulate reflection that leads to strategy implementation.

In AWSM, reviewing and discussing samples of student work help teachers shift from thinking of teaching as something teachers do to a focus on learning as something students do, because it is in student work that student thinking is made visible (e.g., Hattie, 2009). The AWSM approach uses the *Assessment Work Sample Method*, a way to collect, create, discuss, and learn about a formative assessment process using student work samples that help ground the learning in daily practice (Clare, Valdés, Pascal, & Steinberg, 2001; Matsumura et al., 2006). The work samples include a teacher cover sheet that conveys the goals of the lesson, the type of knowledge/skill being developed, the success criteria for meeting learning goals, and general information that will help reviewers understand the “what and why” of the assignment. Attached to the cover sheet are four pieces of student work, two that achieved the learning goals and/or targets and two that did not (see Figure 2 for work sample excerpt). In AWSM, work samples in each session are the foundation of discussions around formative assessment and mathematics teaching and learning.

**AWSM Includes Recommended Characteristics of PD in Formative Assessment**

High-quality PD in formative assessment should be intensive and ongoing, connected to practice, collaborative, content-focused, adapted to local context, active, systemically supported, and coherent (Trumbull & Gerzon, 2013). It should offer teachers choices about the instructional practices they will focus on, the flexibility to adapt formative assessment to their context, provide teachers with a lot of time to change practices, and offer both support and accountability from peers (Leahy & Wiliam, 2012). AWSM has these recommended features, as well as a content focus on mathematics. Specifically, AWSM places teachers in collaborative learning groups, features ongoing meetings throughout the year, and uses work samples to connect the PD to content, instructional practice, and local context. The sessions aim to prompt teachers to examine their current assumptions and decide which aspects of their practice to change, provide time to learn new approaches, and stimulate reflection that leads to strategy implementation.

AWSM is structured around nine face-to-face meetings which include a two-day introductory workshop and eight sessions of about 45 minutes each. The introductory workshop is designed to help participants gain a shared understanding of the formative assessment process and prepare to create positive classroom environments while the eight short sessions are designed to support teachers as they implement the process. During Part One of the introductory workshop, participants build their understanding of formative assessment as one component of a larger assessment system, with an emphasis on providing descriptive feedback to move the learning forward. Participants discuss the characteristics of positive classroom culture and why it is essential, and consider structures that will help them create such a culture. Participants learn about fixed and growth-oriented mindsets (Dweck, 2006) and the type of feedback that will promote a growth-oriented mindset that builds student self-efficacy for mathematics. They discuss physical, social, and emotional factors that impact classroom culture and share successes and challenges of creating such cultures in their classrooms.
In Part Two of the introductory workshop, participants plan for implementing formative assessment by first examining authentic student work. In collaborative groups, teachers analyze the intended learning goals and success criteria from anonymous work samples and then compare the teachers’ intended learning goals to accompanying student work. Using criteria, participants determine if the teacher’s intended learning and the actual student tasks are strongly aligned, partially aligned, or weakly aligned. Mathematical content as well as the inferred cognitive demand of both the learning goal and student task are reviewed and discussed. Through this analysis, participants clarify their understanding of Dimension 1 (learning goal and aligned student task) and Dimension 2 (success criteria...
aligned to the learning goal and student task). These dimensions are the foundation for the formative assessment process because without clarity about what is to be learned and clear criteria for goal attainment, the feedback process can be derailed and learning impeded. When these dimensions are clear, students are supported in taking greater ownership for learning.

During Part Three of the introductory workshop, participants apply the learning from Parts One and Two to their own instructional planning. During this segment, participants analyze an instructional unit they plan to implement to determine whether the learning goals, student tasks, and success criteria are strongly aligned. Participants consider strategies for how to best communicate learning expectations to students and identify/adapt tools to help students track their own progress. Independent work time allows participants to initiate revisions to their instructional unit if needed and receive feedback from peers and workshop facilitators.

The short sessions are scheduled around grade level team times and are organized as teacher learning communities with a facilitator that has both mathematics and formative assessment expertise. Dimension 3, providing descriptive feedback, is the focus of Sessions 1–4, but the program continues to reference topics from the introductory workshop such as class culture, student attitudes, clarity of learning goals, and success criteria because it is important for participants to understand how these dimensions intersect. Participants learn what effective descriptive feedback entails and how to provide both oral and written feedback. Cues, questions, and recommendations for next steps help students take more responsibility for learning than simply providing correct solutions or step-by-step actions. Participants also learn that students should be active in the feedback process, and strategies for implementing effective peer- and self-assessment are discussed. In Sessions 1–5, anonymous student work samples remain at the core of the PD. Participants assess work samples to determine the degree to which written teacher comments will move student learning forward. During this process participants make clear connections between AWSM Dimensions 1–3, as providing effective feedback is dependent on clearly articulated learning goals and success criteria. Participants examine work samples that incorporate student peer and self-assessment, and they use the AWSM rubric (adapted from Matsumura, Garnier, Pascal, & Valdes, 2002) to independently assess anonymous work samples and practice providing written feedback.

Rather than analyze the student work of an unknown teacher, Sessions 6–8 ask participants to share their instructional practice more directly. In these sessions, participants generate their own student work samples, present the work to colleagues, and request feedback on how to improve their own implementation of the process. These sessions are powerful as the entire formative assessment process is fully integrated. Participants reflect on their own progress for implementing formative assessment and identify next steps. The formative assessment process is modeled during each of the sessions. Clear goals and success criteria are provided, procedures are modeled and discussed, and feedback is incorporated to support learning among the participants. Clarifying this approach and calling out the modeling helps participants think about what their students need from them to engage in an ongoing feedback process.

In AWSM, the expertise of the session facilitator should not be underestimated. The mathematics teachers bring targeted questions to these sessions, and the facilitator must be prepared to reference research, share how they themselves were able overcome challenges with using formative assessment, and continue to build and maintain trust with participants.

Methods

In school years 2012–2013 and 2013–2014, we worked with seven middle schools in an urban district in Colorado, with a total of 47 mathematics teachers. One pilot school received the AWSM PD in the 2012–2013 school year, and the remaining six schools received the PD in the 2013–2014 school year.

Research Questions

The overall evaluation of AWSM addresses three primary research questions: (a) to what extent can AWSM be implemented with fidelity in an authentic education delivery setting? (b) to what extent does AWSM show promise for improving teacher practice of mathematics formative assessment? and (c) to what extent does AWSM show promise for increasing student achievement in mathematics? The research involved teacher measures of formative assessment practice, student measures of mathematics achievement, observations of professional development sessions, and teacher questionnaires and focus groups.

Data Collection

To answer the aforementioned research questions, data were collected from (a) teacher work samples for assessing teacher practice in formative assessment, (b) a pretest of content and pedagogical knowledge in mathematics, (c) a
district-wide mathematics assessment to assess the impact of AWSM on student mathematics achievement, and (d) teacher focus groups to assess teacher perceptions of the impact of AWSM on their knowledge and practice of formative assessment. Data collection for each of these outcomes measures are discussed in turn below.

**Work samples.** To provide a measure of teacher practice of formative assessment, participating teachers provided samples of student work that were graded and/or marked with teacher feedback. Student work samples were linked to an assignment that could be a homework/seatwork assignment, performance assessment, or quiz/in-class assessment. For each assignment, teachers completed a cover sheet to describe the assignment (including directions given to students); the source of the assignment (e.g., teacher created; from a textbook, workbook, curriculum publisher, etc.); why the particular assignment was selected; the purpose of the assignment (e.g., check student progress; provide practice, unit post-test); the student learning goals for the assignment; student accommodations for the assignment (e.g., addressing different skill levels; whether students received help); how the assignment was assessed (e.g., rubric; grading system; student reflection); and criteria for deciding whether students met learning objectives. For each assignment, participating teachers provided four samples of student work—two that met learning objectives and two that did not meet learning objectives.

Participating teachers submitted work samples at the beginning of the school year, prior to the AWSM professional development sessions (pretest), and again at the conclusion of AWSM program (post-test). Two analysts both scored all work samples using the AWSM rubric (see Figure 3 for sample) on a scale from 1 to 4, indicating the level of quality of formative assessment practices on six separate dimensions: (a) focus of the goals on student learning; (b) alignment of learning goals and task; (c) alignment of learning goals and assessment criteria; (d) clarity of the student assessment criteria; (e) feedback type; and (f) feedback integrates student involvement. When scores for the two raters were discrepant, the raters discussed the work sample and reconciled their scores to produce one score per work sample per dimension.

<table>
<thead>
<tr>
<th>Focus of the goals on student learning</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goals are highly focused on student learning. Goals are:</td>
<td>• fully elaborated (all aspects and parts of the learning goal are included and described);</td>
<td>• appropriate to student learning needs (grade/course appropriate); and</td>
<td>• may or may not be clearly communicated to students.</td>
<td>No learning goals are provided for the task.</td>
</tr>
<tr>
<td></td>
<td>• clear statements (not an activity, topic, phrase, or question);</td>
<td>• appropriate to student learning needs (grade/course appropriate); and/or</td>
<td>• may not be clearly communicated to students.</td>
<td>• no clear description of what students should know or be able to do; or</td>
</tr>
<tr>
<td></td>
<td>• may or may not be clearly communicated to students.</td>
<td></td>
<td></td>
<td>• topic rather than learning goal may be listed.</td>
</tr>
</tbody>
</table>

![Figure 3. Sample of two dimensions of AWSM work sample rubric (adapted from Matsumura et al., 2002). Higher scores indicate higher formative assessment quality.](image-url)
Measure of pedagogical content knowledge. The learning mathematics for teaching (LMT) assessment (Learning Mathematics for Teaching Project, 2011) was the measure of teacher content and pedagogical knowledge in mathematics. The purpose of the LMT assessment is to capture teachers’ mathematical knowledge for teaching (MKT). The LMT is not an assessment of “common” content knowledge alone, but knowledge that is specific and useful for teaching student mathematics. The assessment was designed to discriminate teachers’ MKT across the entire ability range; therefore, items were chosen such that the average teacher answers 50% of them correctly. The developers stressed that the assessment not for drawing conclusions about the competency of individual teachers, but rather for purposes such as measuring the effectiveness of professional development. The LMT content area assessments included Middle School Number Concepts and Operations (NCO) and Middle School Patterns, Functions and Algebra (PFA), as these assessments were aligned to AWSM PD content.

Student achievement. For the impact analysis of the AWSM on student achievement, the outcomes were mathematics scores on the Measures of Academic Progress (MAP) assessment administered at two-time periods during the school year (fall and spring). We used a quasi-experimental difference-in-differences approach (Cook, Shadish, & Wong, 2008), comparing students of teachers who participated in the AWSM PD to their previous year’s students prior to AWSM exposure. Two separate analyses were conducted for proximal and distal outcomes. The proximal outcome analysis was conducted at the end of the training year and included 7 schools, 47 teachers, 2,281 counterfactual group students, and 3,896 treatment group students. The distal outcome analysis was conducted one-year post-training and included 1 school, 3 teachers, 312 counterfactual group students, and 330 treatment group students.1 Both the proximal and distal analyses utilized 2-level hierarchical linear models (HLM) to account for students nested within teachers. The proximal analysis included school dummy codes at level 2 to account for shared variance between students at the school level.

Teacher focus groups. Schools were to have two teacher focus groups in the fall 2013 semester and two in the 2014 spring semester. All but two schools had the four planned focus groups; due to scheduling difficulties two of the schools had three focus groups, resulting in a total of 22 focus groups for the 2013–2014 school year. Each focus group lasted for about 45 minutes. The focus groups were audio recorded and transcribed.

The semistructured focus group protocol addressed the research questions by asking for teachers’ feedback about AWSM implementation (“How have the program facilitators supported you before and during AWSM?”), perceptions of impact on their classroom practice (“To what extent is AWSM building your ability to use formative assessment in your classroom?”), perceptions of impact on student learning (“What strategies seem most effective for student learning?”), and recommendations for improvement. Although researchers had the same protocol for all focus groups, the phase of AWSM implementation influenced which questions they asked at each group. For example, researchers did not ask about results from teachers’ use of formative assessment strategies in the first focus group because it was too early then for teachers to have substantial experience of using the strategies.

Audio recordings of focus groups were transcribed, and the transcripts were imported into MAXQDA for analysis. Two coders developed initial codes based on the question topics in the focus group protocol. As coding progressed, the two coders conferred frequently on the coding scheme and the meaning of the codes, and collapsed some coding categories while expanding others into sub-codes. The final list of codes addressed session spacing and timing, fostering teacher collaboration, session mathematics content topics, facilitator support, understanding of formative assessment, need for professional development in formative assessment, implementation of formative assessment strategies, planning for formative assessment, formative assessment for differentiation, and results from formative assessment strategies. The emphasis of the analysis was to represent the points of view of the respondents, while also including some representative comments.

After each iteration of AWSM, the evaluators analyzed focus groups feedback and provided the developers with formative information to revise the PD materials. For example, the final version of AWSM included more content information in the introductory summer session and an earlier introduction of the work sample rubric, in response to teacher feedback.

Analysis and Results

Results below are provided for the teacher work samples, teacher pedagogical content knowledge assessment, student achievement, and focus groups.

Teacher Work Samples

Table 1 displays descriptive statistics by work sample dimension as well test statistics and effect sizes for comparisons of pre- and post-test work sample scores. Teachers showed the greatest improvement in feedback:
Table 1
Descriptive Statistics, Test Statistics, Effect Sizes for the Work Sample Pretest and Post-Test by Rubric Dimension (N = 47)

<table>
<thead>
<tr>
<th>Rubric Dimension</th>
<th>Pretest Mean (SD)</th>
<th>Post-Test Mean (SD)</th>
<th>Test Statistic</th>
<th>p-Value</th>
<th>Cohen’s d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focus of the goals on student learning</td>
<td>2.23 (.70)</td>
<td>2.52 (.56)</td>
<td>.80</td>
<td>.43</td>
<td>.45</td>
</tr>
<tr>
<td>Alignment of learning goals and task</td>
<td>2.21 (.72)</td>
<td>2.32 (.57)</td>
<td>2.22</td>
<td>.03</td>
<td>.16</td>
</tr>
<tr>
<td>Alignment of learning goals and assessment criteria</td>
<td>1.38 (.49)</td>
<td>2.01 (.70)</td>
<td>5.02</td>
<td>&lt;.0001</td>
<td>1.03</td>
</tr>
<tr>
<td>Clarity of the student assessment criteria</td>
<td>1.43 (.50)</td>
<td>2.24 (.70)</td>
<td>6.54</td>
<td>&lt;.0001</td>
<td>1.35</td>
</tr>
<tr>
<td>Feedback type</td>
<td>1.34 (.60)</td>
<td>2.53 (.60)</td>
<td>9.81</td>
<td>&lt;.0001</td>
<td>1.98</td>
</tr>
<tr>
<td>Feedback integrates student involvement</td>
<td>1.17 (.38)</td>
<td>2.49 (.67)</td>
<td>12.48</td>
<td>&lt;.0001</td>
<td>2.42</td>
</tr>
<tr>
<td>Average score across all dimensions</td>
<td>1.63 (.54)</td>
<td>2.35 (.61)</td>
<td>8.74</td>
<td>&lt;.0001</td>
<td>1.79</td>
</tr>
</tbody>
</table>

“feedback integrates student involvement” (d = 2.42) and “feedback type” (d = 1.98). Participants’ change scores also showed high effect sizes (greater than 1) on the two dimensions involving assessment criteria: “alignment of the learning goals and assessment criteria” (d = 1.04) and “clarity of the student assessment criteria” (d = 1.35). Participants showed the least improvement in the two dimensions involving learning goals: “focus of the goals on student learning” (d = .16) and “alignment of learning goals and task” (d = .45), as they scored relatively high at baseline. Scores did not differ significantly by teacher experience level.

Tests of statistical significance were conducted using HLM to account for nesting of teachers within school. Seven separate models were estimated for each of the work sample dimensions as well as the average score across dimensions. For the average work sample scores, the analysis revealed a significant increase from pre- to post-test (see Figure 4). For the analyses of individual dimensions, dimensions representing assessment criteria and feedback (Dimensions 3–6) showed statistically significant increases from pre- to post-test. For dimensions representing learning goals, change in “focus of learning goals on student learning” (Dimension 1) was statistically significant from pre- to post-test; however, change in “alignment of learning goals and task,” while in the positive direction, was not statistically significant (see Table 1 for test statistics and p-values for each comparison). Overall, these findings suggest that the AWSM professional development program increased teachers’ formative assessment practice, particularly in the areas of assessment criteria and feedback.

**Pedagogical Content Knowledge**

Baseline pedagogical content knowledge was to be used as a covariate in the student achievement model to assess the extent to which it affected the impact of the professional development on teacher practice and student achievement. Across both content areas, there was wide variability among teachers on assessment scores. Specifically, standardized
scores ranged from −1.52 to 2.12 on the PFA assessment (mean = .41, SD = .80) and from −1.64 to 1.91 on the NCO assessment (mean = .31, SD = .87). A comparison of scores between the two subtests revealed no statistically significant difference between NCO and PFA. In order to compare scores from the AWSM field test sample to the normed LMT sample, one-sample t-tests were performed. Results indicate that the study sample scored significantly higher than the normed sample on both the NCO subtests, indicating that the study sample had greater pedagogical content knowledge at AWSM course entry than did the population of teachers who had previously taken the LMT. LMT scores did not significantly predict change in teacher work sample scores, indicating that prior mathematics pedagogical content knowledge was not a moderator of these effects. Therefore, AWSM improved teachers’ formative assessment practice regardless of their mathematics instruction entry skills.

**Student Achievement**

For proximal student achievement (PD year, 7 schools), results indicate that the counterfactual group scored higher than the treatment group at the fall pretest ($p < .0001$, Glass’s $Δ = -.08$) and at the spring outcome ($p < .0001$, Glass’s $Δ = -.07$). However, the treatment x time interaction was not significant, suggesting that there was no difference in growth from fall pretest to spring post-test for students who were exposed to the AWSM PD program versus the previous year’s students of the same teachers who were not exposed to AWSM.

For distal student achievement (1-year post training, 1 school), there was no significant difference between the treatment and counterfactual groups at the fall pretest or the spring outcome. In addition, the treatment x time interaction was not significant (see Figure 5).

**Focus Groups**

In early focus groups, teachers reported that they clearly communicated learning goals and success criteria to students, and used ungraded practice quizzes so students would understand how well they were mastering the learning goals. Teachers reported that this helped clarify their teaching and let students know what they should expect to learn. Further, having clear learning goals and success criteria facilitated communication with students and parents. Teachers said that formative assessment data helped them work with students at various levels of understanding, by changing the structure of the classroom so that students were working on different assignments based on their level of mastery of the learning goal. This helped each group focus on the next step in the learning progression.

According to teachers, AWSM helped them realize that students should be involved in the formative assessment process to increase student understanding and accountability and enable them to be more independent learners. For example, teachers gave students success criteria to use in self-assessing their level of understanding relative to the learning goal. They also paired students for more formal peer assessment. This required teachers to teach students how to provide feedback to one another and to monitor the peer feedback process. As teachers had more practice with involving students in the formative assessment process, they reported more successes.

In later focus groups, as AWSM progressed and teachers got specific suggestions in implementing formative assessment strategies, teachers at all schools began using the strategies more often and in a more efficient way. For example, teachers at all schools used more ungraded activities and found ways of providing feedback. They reported implementing low-risk ungraded quizzes as checkpoints; students working on these quizzes to get ready for the summative assessment could redo their work until they achieved mastery. In order to give feedback efficiently, one teacher arranged his/her classroom in centers to allow for more opportunities for one-on-one feedback. Another

---

**Figure 5.** Adjusted means for mathematics achievement growth from fall pretest to spring outcome for treatment versus counterfactual groups.
teacher devised a system of two-letter codes that corresponded to items of feedback, such as “check to see if your answer is realistic (number sense).” This teacher explained, “You know, I’m trying to find as efficient a way as I can to make sure they’re not practicing doing it wrong.” One teacher commented that whenever he/she provided students with ungraded written feedback on homework assignments, students subsequently performed better on those topics in summative assessments.

At every participating school, teachers reported that their students initially resisted ungraded practice and expected a grade for every activity. Some of these teachers said that they persisted with the technique until students came to understand the advantages it has for them, such as not being assigned a grade until they had had sufficient practice. Teachers also linked ungraded practice activities to supporting the effort, focus, and perseverance mathematical practices of the Common Core, and also told students that the process and be able to tell whether their answer was correct. One described the process: “I worked off of that [template] for weeks, developing more success criteria and going into a simple rubric . . . And having that as an electronic file that I could tweak and turn and change for each one of my classes, that was a huge deal.”

Teachers at five schools reported changing their feedback practices to avoid focusing primarily on getting right answers. They described their struggling students as initially reluctant to engage in complex problem solving out of fear of getting a wrong answer; these students would avoid the task or try to get teachers to simplify it or tell them exactly what to do. Several teachers would supply answers to problems so that students would focus instead on the process and be able to tell whether their answer was correct. One said, “I try to implement, through the year, I don’t care what the answer is. Show me how you get to the correct. One said, “I try to implement, through the year, I don’t care what the answer is. Show me how you get to the correct. One said, “I try to implement, through the year, I don’t care what the answer is. Show me how you get to the correct.

At five schools, teachers said they were using peer assessment extensively. They described having students partner to discuss their approaches to homework problems or to go over in-class activities. One commented that this technique made it easier for students to get help when they were reluctant to seek help from the teacher. Teachers found that it was important to model and demonstrate to students how to give effective peer feedback. For example:

Well, I used the Two Stars and a Wish [a peer feedback organization strategy], and I just put that right on the paper so they kind of knew ahead of time that they were going to get peer feedback. And the first time we did it, I kind of gave an example and showed some things that I wanted. I gave some counterexamples like, “Don’t say, ‘Nice handwriting,’” or something like that.

Teachers who used peer assessment sometimes found that students could better explain errors and next steps to peers. One said:

When you walk around the room and you listen to their discussions, they’re really good discussions. Sometimes they can point something out to another student that’s at their level better than I can. Sometimes, we come in it from the teacher point of view, and the kid point of view they’re struggling along with them and they find ways to get through to them and communicate to them that we hadn’t thought about.

Another teacher said, “I’ve had several kids, when working together, say ‘Oh, now I understand it because so and so explained it to me.’ And it’s like that’s good because sometimes they’ll say something a little bit differently, and it makes more sense.” Some of the teachers pointed out that effective peer feedback helps the student giving feedback as well as the one receiving it in any interaction. For example:

[T]heir feedback was actually better than their actual work on the original assessment . . . As far as the actual, the task itself, there were some kids who did really, really well and some kids not so much, but the actual feedback, even the kids who didn’t do so well on the [assignment], they started recognizing. “Oh yeah, that would have been a better approach” and so they actually . . . were learning even though they didn’t understand how to do it themselves initially. I think by them reading someone else’s they actually made sense of it.

Some teachers mentioned that peer feedback helped students understand what teachers are looking for when
they ask students to show their work—the steps in the students’ approach to problems. One mentioned pairing students who show their work for peer assessment with those who generally do not, because peer feedback about how to show work was more effective at getting students to try it than was teacher encouragement.

At two schools, teachers were initially concerned that students would be unkind to one another during peer feedback. This turned out to be a problem when peer feedback was given in written, anonymous form. However, when teachers then tried structured, face-to-face peer assessment, it worked well: “I find when they verbally feedback to their peer, they’re much nicer, it’s more constructive, and it’s actually a lot more helpful.”

At three schools, teachers also focused on student self-assessment. They gave students tracking sheets of success criteria so they could understand their progress. One said, “They’ll start on what they know, but then they actually take ownership of saying, ‘Oh, I haven’t mastered this.’ And then they start testing their own learning.” Another teacher described how he/she used red, yellow, and green cards that the students would periodically hold up to communicate their understanding of the learning objective. This helped him/her know when to move on and when to spend more time on a concept. Regarding self-assessment strategies, one teacher said, “it’s also teaching them to be a little bit more independent workers, and be problem solvers on their own, and figuring it out.” Teachers at these three schools said they looked forward to planning for student self and peer assessment from the beginning next year to establish routines and help students recognize their progress toward mathematics learning goals.

Discussion

The AWSM professional development was feasible to implement in an actual school setting. The work samples helped to ground assessment discussions in teachers’ actual practice. The emphasis on formative assessment as a process including learning goals, success criteria, aligned tasks, and feedback encouraged teachers to see it as a way of framing instruction, rather than as something separate from the teaching/learning cycle. The developers made adjustments to the AWSM schedule and materials over the three iterations in order to respond to participants’ performance and suggestions.

Effect sizes for changes in assessment work sample scores from pretest to post-test indicated that teachers overall demonstrated significant improvements in their formative assessment practices, with the greatest improvement in the two dimensions involving feedback and on the two dimensions involving assessment criteria. Participants showed the least improvement in the two dimensions involving learning goals, in part because of relatively high baseline scores. The pretest of mathematics pedagogical content knowledge did not moderate the pre–post difference in teachers’ assessment skill. Overall, these findings suggest that the AWSM professional development program improved teachers’ formative assessment practice, particularly in the areas of assessment criteria and feedback, and that it did so regardless of teachers’ prior mathematics pedagogical content knowledge.

Although literature on formative assessment documents learning gains associated with higher-quality practices (e.g., Black & Wiliam, 1998a, 1998b), in this study there was no statistically significant positive impact of AWSM on student achievement as measured by MAP during the training year, despite the change in teacher practice. MAP data on the post-training-year were available only for one school, and also did not show statistically significant impact. However, because AWSM was implemented over an entire school year and because it encouraged teachers to think of formative assessment as an approach to instructional practice affecting the whole teaching-learning cycle, it is unlikely that student impacts would be seen in a training year. Future research will focus on student achievement outcomes in the year prior to AWSM professional development implementation, when teachers can plan to use the AWSM strategies from the beginning of the year.

In focus groups, participating teachers said that their students initially had the same problems reported in the literature with student motivation for learning mathematics, such as difficulty with engagement and persistence, especially with challenging problems. Their students also were reluctant to be wrong, to show work, and to do work that was ungraded. When AWSM teachers were able to involve students in understanding learning targets and in tracking progress on success criteria, they found that students were more likely to seek clarification of the goals and contribute to development of the success criteria. Teachers were able to group students to work together supportively, and to implement student peer assessment and self-assessment, even with struggling students. With the peer assessment process, using success criteria focused students on providing feedback relative to levels of performance, not just about whether the answer was right. Teachers were able to implement ungraded assignments by explaining to students their purpose, although some students struggled with this because students were used to receiving “points” as an indication of proficiency. However, when practice activities in the classroom were no
longer high-stakes and graded, students were more willing to engage and persist in more complex problem solving. Because of the importance of complex problem solving to the mathematical practices aspect of contemporary standards, and literature supporting the role of formative assessment in motivation and persistence, future research with AWSM will focus on measures of motivation in mathematics, and performance in solving complex problems (which was not addressed by the assessments we used as the achievement outcome in this study).

Teachers were able through the AWSM sessions to share their assessment practices and learn from their peers. With peer support, they were able to adjust their mindsets about what students were able to do when they shifted from a teacher-centered classroom to a more student-centered approach. Finally, teachers recognized that implementing a formative assessment process did not mean adding another initiative to an already full plate of programs but was part of effective teaching and student learning.

References


---

**Authors’ Notes**

Keywords: attitudes/beliefs, learning processes, classroom discourse, teachers and teaching, professional development, student assessment, students and learning.

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A110392. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

1 Only one pilot school (three teachers), which received PD in the 2012–2013 school year, provided data for the distal outcome (one year post training).