2014

Middle School Mathematics Teacher Evaluation: The Role of Subject Matter in Supervisor Feedback

Christine Trinter
Virginia Commonwealth University, ctrinter@vcu.edu

Heather Carlson-Jaquez
Virginia Commonwealth University, carlsonha@vcu.edu

Follow this and additional works at: http://scholarscompass.vcu.edu/merc_pubs

Part of the Education Commons

Downloaded from http://scholarscompass.vcu.edu/merc_pubs/7

This Research Report is brought to you for free and open access by the MERC (Metropolitan Educational Research Consortium) at VCU Scholars Compass. It has been accepted for inclusion in MERC Publications by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.
A Review of Literature

Middle School Mathematics Teacher Evaluation: The Role of Subject Matter in Supervisor Feedback

By
Christine P. Trinter, Ph.D.
Heather Carlson-Jaquez

METROPOLITAN EDUCATIONAL RESEARCH CONSORTIUM
# MERC

## Metropolitan Education Research Consortium

### MERC Membership

- Chesterfield County Public Schools
- Colonial Heights City Schools
- Goochland County Public Schools
- Hanover County Public Schools
- Henrico County Public Schools
- Hopewell City Public Schools
- Powhatan County Public Schools
- Richmond City Public Schools
- Virginia Commonwealth University

### Background

Virginia Commonwealth University and the school divisions of Chesterfield, Colonial Heights, Goochland, Hanover, Henrico, Hopewell, Powhatan, and Richmond established the Metropolitan Educational Research Consortium (MERC) on August 29, 1991. The founding members created MERC to provide timely information to help resolve education problems identified by practicing professional educators. MERC currently provides services to over 12,000 teachers in eight school divisions. MERC has based funding from its membership. Its study teams are composed of university investigators and practitioners from the membership.

MERC is organized to serve the interests of its members by providing tangible material support to enhance the practice of educational leadership and the improvement of teaching and learning in metropolitan educational settings. MERC’s research and development agenda is built around four goals:

- To improve educational decision-making through joint development of practice-driven research questions, design and dissemination,
- To anticipate important educational issues and provide leadership in school improvement,
- To identify proven strategies for resolving instruction, management, policy and planning issues facing public education, and
- To enhance the dissemination of effective school practices.

In addition to conducting research as described above, MERC conducts technical and educational seminars, program evaluations, an annual conference and publishes reports and research briefs.

---

Copyright © 2014. Metropolitan Educational Research Consortium (MERC), Virginia Commonwealth University

The views expressed in MERC publications are those of individual authors and not necessarily those of the consortium or its members.
MIDDLE SCHOOL MATHEMATICS TEACHER EVALUATION:
THE ROLE OF SUBJECT MATTER IN SUPERVISOR FEEDBACK

TABLE OF CONTENTS

Background ........................................................................................................................................... p1
Teacher Evaluation Systems .................................................................................................................. p2
   Value Added Models ....................................................................................................................... p2
   Portfolios ......................................................................................................................................... p3
Recommended Evaluation Models ........................................................................................................ P3
   Virginia’s Teacher Evaluation Model ............................................................................................. p3
The Nature and Benefits of Observation and Feedback to Teachers ............................................... p3
   Subject-Specific Feedback ............................................................................................................. p5
Conclusion ............................................................................................................................................ p6
References ........................................................................................................................................... p7
Background: Mathematics Education as a Federal Priority

Mathematics Assessment Results on a Global Scale. For decades, mathematics education in the United States has been a focus of national attention. Since the implementation of the First International Mathematics and Science Study in the mid-sixties, critics have reported on the United States’ global standing on achievement tests. Almost two decades after the first administration of the International Mathematics and Science Study, the widely publicized report, A Nation at Risk: The Imperative for Educational Reform (National Commission on Excellence in Education, 1983) described various indicators for why the U.S. is at risk for losing its “preeminence in commerce, industry, science, and technological innovation” (p. 11). These risks include lack of achievement on standardized tests and an increase in remedial mathematics courses taught at public four-year colleges. More recently, results from the 2012 Programme for International Student Assessment (PISA) have garnered attention. The PISA evaluates 15 year olds from around the world on mathematical literacy, assessing both conceptual and procedural knowledge. On December 3, 2013 the Secretary General for the Organization for Economic Cooperation and Development, the governing agency for the PISA, presented the latest results from this test. The secretary general described U.S. student’s mathematics performance as “stark” explaining that very few of them reach proficiency Level 2, which only requires students to solve basic tasks using algorithms and whole numbers. Furthermore, the data show that U.S. students who completed this assessment struggled with tasks that demand complex mathematical thinking. Among the highest-level learners, only 2% of U.S. students reached the maximum performance level (conceptualization, generalization, using and applying mathematics creatively) while the OECD average was 3% and over 30% of the students from Shanghai, Hong Kong, Singapore, Chinese Taipei and Korea reached the highest level (OECD, 2013).

While policies for education reform have been around since the 1964 Elementary and Secondary Education Act, these kinds of international test scores lead policymakers to question why mathematics education in the U.S. continually falls short at both the national and the international field (U.S. Department of Education, 2008). Considering that teachers play a significant role in student learning (Alton-Lee, 2006; Hanushek, 2011), policymakers have placed heavy emphasis on teacher evaluation measures.

Education Reform Policies. In 1994, the Improving America’s Schools Act was re-instated and required states to develop standards and report data on student evaluation measures linking “adequate yearly progress” to assessment results (Improving Americas Schools Act, 1994). This Act directed attention to mathematics achievement at the state and local level. Following this, mathematics teachers became a central focus of national education reform with the passing and implementation of No Child Left Behind (NCLB) (2002), which sought to motivate students and teachers to increase student achievement and close achievement gaps. In particular, section 2201 of NCLB states that schools need to “to improve the academic achievement of students in the areas of mathematics and science” by: encouraging improvements of math teacher education; developing lifelong learning programs and professional developments for math and science teachers; build partnerships between teachers and professionals in math and science fields to develop teacher skills; and develop a more rigorous math and science curriculum (Definition A). NCLB also required that teachers meet expressed criteria to be labeled as “highly qualified”. In particular, teachers must hold a state mandated teaching license obtained by passing teaching licensure exams and completing a specified number of college credits for the content in which they teach.

More recently, President Obama reauthorized the Elementary and Secondary Education Act (2010) stating: “We are calling on states and districts to develop and
implement systems of teacher and principal evaluation and support, and to identify effective and highly effective teachers and principals on the basis of student growth and other factors.” Furthermore, this Act purports that teacher evaluation should drive promotion and retention while informing professional development and improving student learning. The Race to the Top (RTTT) initiative once again set math as a priority, and sought to reward states that offered rigorous mathematics courses, created community partnerships to give students applied learning opportunities and prepared students for advanced study in science, technology, math and engineering (STEM) fields (U.S. Department of Education, 2009). RTTT also sought to reward states who increased student achievement and decreased achievement gaps between subgroups in math, as well as ensure that “high-poverty and/or high-minority schools...have equitable access to highly effective teachers” (U.S. Department of Education, 2009, p. 9). Again, this initiative focused on “highly effective teachers” who are defined and evaluated, in part, by student growth measures. RTTT states “supplemental measures may include, for example, multiple observation-based assessments of teacher performance (p. 13).” This definition of highly effective teacher is aligned with the current teacher evaluation system in the state of Virginia (Virginia Department of Education, 2011) which includes a new component, Standard 7.3, that links student performance on “state provided growth measures” to teacher evaluation (p. 12). Though federal and state legislation are both explicit about the need for teacher evaluation and accountability, both governing bodies give local education agencies (LEAs) latitude to create a customized evaluation system that incorporate the general guidelines outlined. As a result, school districts nationwide have been investing time and resources into developing teacher evaluation instruments and protocols to assist administrators in documenting teacher effectiveness. These evaluations play a significant role in the professional growth and careers of many educators.

Teacher Evaluation Systems

Teacher evaluation systems have recently created a national stir with teacher expulsions for some and merit pay for others. A variety of teacher evaluation methods are employed in schools. Two widely used measures are value-added models and teacher portfolios. Value-added models “use statistical methods to measure changes in student scores over time while considering student characteristics and other factors often found to influence achievement” (Darling-Hammond et al., 2012, p8). Portfolios offer a more global perspective of a teachers’ proficiency and may include student and teacher artifacts as well as observations by administrators (Sartain, Stoelinga, & Brown, 2011). Each model offers unique strengths and weaknesses in terms of giving teachers quality feedback that can improve their teaching as well as providing administrators with reliable and valid data with which to make personnel decisions. Though current federal policy advocates for value-added models (Milanowski A., 2011; Yeh, 2012), teacher evaluation methods employed in individual school systems are by and large decided by local agencies (U.S. Department of Education, 2009).

Value-Added Models

Many states have adopted value-added models (VAM) for teacher evaluation based upon the recommendation of federal policy (Milanowski A., 2011; Yeh, 2012). VAM attempt to estimate, through statistical analysis of test scores, the likely contribution of a teacher to student learning (Milanowski A., 2011). Though some studies have found evidence to support using VAM for teacher evaluation (e.g. (Milanowski A., 2004), others argue that the model lacks sufficient reliability and has too much measurement error to be used to make high-stakes decisions, is unfair to teachers of populations who traditionally score lower on standardized assessments, in some cases reduces student achievement, and is cost ineffective (Milanowski A., 2011; Yeh, 2012; Darling-Hammond et al., 2012). If VAM are to be used, the literature suggests that they should be used in
conjunction with other forms of evaluation to provide a more comprehensive view of teacher performance (Milanowski A., 2011).

Portfolios

Portfolios are collections of classroom artifacts, and can include documents such as lesson plans, class assignments, student work, and photographs or video evidence of what is occurring in the classroom (Darling-Hammond & Snyder, 2000). Portfolios can provide a broad overview of the teachers’ contribution to classroom learning as well as a set of documents for teachers to reflect upon their practice (Darling-Hammond & Snyder, 2000; Moss, et al., 2004); however, caution should be taken to assure reliability and validity when scoring portfolios for use as a method of teacher evaluation (Schutz & Moss, 2004). Despite this caution, portfolio assessments of teacher performance have been used successfully by organizations such as the National Board for Professional Teaching Standards. They provide a longitudinal process by which teachers are encouraged to be reflective, provide multiple examples of teaching and learning, and collect products which can be shared for the sake of furthering teaching and learning (Darling-Hammond & Snyder, 2000).

Recommended Evaluation Models

Many researchers recommend evaluation models that include multiple methods of data collection in an effort to account for the limitations of each measure. For example, Milanowski (2011) notes that “outcome measures don’t provide enough information to improve teacher performance” and “instructional practice measures that aren’t linked to effects on learning are likely to lose their rigor and relevance” (pp. 19-20). Rockoff & Speroni (2011) found evidence to support that first year teachers who receive quality subjective evaluations produce greater gains in student achievement with future students, but recommend both subjective evaluations by trained professionals and objective performance data to identify low teacher quality. Similarly, Darling-Hammond et al. (2012) report that effective systems utilize trained evaluators, provide frequent evaluation and feedback, and integrate measures (e.g. observations, videos, artifacts) that link what teachers do to what happens as a result.

Virginia’s Teacher Evaluation Model

Virginia’s recently added Standard 7, or student-based outcome measures, shifts evaluation focus, in part, to student academic progress (Virginia Department of Education, 2011). The Commonwealth, however, has retained other professional and instructional components such as: formal, informal and walk-through evaluations; student surveys; and portfolios and artifacts as part of the evaluation model. Further, Virginia’s teacher evaluation guidelines include two components to support teacher improvement: support dialogue and performance improvement plan. Both components include dialogue between evaluators and teachers in an effort to improve teacher performance and subsequent student achievement. Each of these components align with the previously described, literature-based, recommended evaluation models.

The Nature and Benefits of Observation and Feedback to Teachers

A commonly employed method for promoting dialogue between evaluators and teachers and one that is included in recommended models (Darling-Hammond & Snyder, 2000; Moss, et al., 2004) as well as Virginia’s Teacher Evaluation plan (Virginia Department of Education, 2011) is observation and feedback from administrators. This method is often included in teacher evaluation frameworks that look at multiple aspects of teaching and learning (Sartain, Stoelinga & Brown, 2011). An important component of this process is the feedback that the teacher receives from the evaluator (Scheeler, Ruhl, & McAfee, 2004; Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012). Assessment research tells us that feedback is most effective when it communicates current level of achievement in relation to
goals and provides steps to attaining goals (McMillan, 2011). Furthermore, quality feedback can be described as timely, specific and frequent (Northcraft, Schmidt, & Ashford, 2011; Price, Handley, Millar, & O’Donovan, 2010). Previous studies with college students have found that feedback also provides for interaction where the evaluator can express care and respect for those who are being evaluated, which can help those who are receiving the feedback calm anxiety and regulate emotions (Rowe, 2010). Teacher evaluations should include specific and clear feedback so that teachers can improve their practice using the results (Milanowski A., 2011).

A review of literature on feedback to teachers conducted by Scheeler, Ruhl and McAfee (2004) found 208 articles were published on feedback to teachers between 1970-2004; however, only 4% of those articles focused on in-service teachers, with the rest focusing on pre-service teachers. The authors narrowed the focus of their review by choosing articles that had an independent variable that was a dimension of feedback (nature of feedback, temporal dimensions of feedback, and who gives feedback) and were true experimental or quasi-experimental. They conclude that “feedback is better than no feedback, immediate feedback is better than delayed feedback, and feedback that is immediate, specific, positive and corrective holds the most promise for bringing about lasting change in teaching behavior” (p. 405). Though some studies on feedback to teachers consider the method of delivery of the feedback and who gives the feedback, these studies lack the validity required to make broad generalizations (Scheeler, Ruhl, & McAfee, 2004).

Sartain, Stoeinga & Brown (2011) studied the effectiveness of a teacher evaluation framework that employed the observation/feedback model between administrators and teachers. The researchers found that some areas of the protocol, with regard to both observation and feedback, were reliable while others were less consistent. In particular, teachers were observed by both a researcher and an administrator with each observer using a common scale to rate the instruction. When reporting on the higher end of the scale (proficient versus distinguished instruction), there was significant discrepancy between the observation ratings. Administrators were more likely than the researcher to rate a teacher as “distinguished”. In this same study, the conversations between administrators and teachers were observed and the analysis showed that principals were more likely to ask “low end” questions that did not invoke reflective conversation versus “high end” questions that sparked deeper discussion about the instruction. Administrators explained that they took into account their relationship with the teacher as well as the teacher’s prior evaluations when rating the teacher.

Considering the personal nature of the observer-feedback evaluation cycle, the experiences and perceptions of teachers and observers are also noteworthy. Studies that have looked at teacher and administrator perceptions and experiences emphasize the need for multiple observers; specific, written feedback coupled with dialogue; and adequate time for the full cycle to be effectively employed (Collins, 2004; Ovando, 2005; Ovando & Ramirez, 2006). Specifically, in one qualitative study, teachers and administrators had differing perceptions of the nature of the given feedback following teacher observations (Collins, 2004). Teachers in this study believed that when instruction was satisfactory, they received no feedback from administrators. This was problematic for teachers as they expressed a need for feedback, regardless of the nature of instruction. The administrator believed that negative written feedback may lower morale and result in poor performance and therefore, limited the written feedback. Collins recommends that the evaluation process should be modified to include supplemental observers such as department heads and senior teachers. These observers would be subject experts and together with the administrator’s observation, the evaluation process would be more comprehensive and would include sharing written feedback documents with teachers.
Using action research methodology, Ovando (2005) examined the experiences of teachers and administrators during their observation and feedback cycle. Administrators reported that in order to effectively provide written feedback they should develop knowledge of quality instruction, scripting skills and appropriate professional language during graduate work. Similar to Collins (2004) finding, administrators commented on the importance of adequate time to write the feedback and include the strengths and weaknesses of instruction and teachers noted the importance of specificity in written feedback. Additionally, they appreciated face-to-face conversation about the observation and the written feedback. Some of the components for effective feedback include post-observation conferences between the administrator and teacher that (1) focus on the strengths of the instruction, (2) are based on observable actions and (3) result in professional development goals for the teacher (Ovando, 2005).

Subject-Specific Feedback

With the current focus in mathematics education on process standards, student mathematical dialogue, justification and modeling (Common Core State Standards-Math, 2012; NCTM, 2000; VA Department of Education Standards of Learning, 2009), it is critical that administrators direct their attention to more than pedagogical and behavioral concerns in instruction but also value subject matter in both the content and the practice of disciplines (Nelson & Sassi, 2000). In 1989, The National Council of Teachers of Mathematics published its Professional Standards for Teachers of Mathematics. These standards outlined best practices with regard to teaching mathematics and the evaluation of, support for and development of mathematics educators. These standards are still upheld today as guideposts for exemplary mathematics teaching and learning (Jacobs, J., et al., 2006). In this document, the NCTM described eight evaluation standards and stated “each standard serves as a statement about what should be observed regardless of who is doing the observing” (Introduction section). According to the NCTM, evaluations of teachers’ competence should adhere to these standards and the process of evaluation described. The eight standards are grouped into two headings: (1) the process of evaluation and (2) the foci of evaluation. Central to the process of evaluation is the inclusion of multiple observations from more than a single observer with the teacher involved as a reflective practitioner, providing information to the observer about the teacher’s goals and a self-analysis of teaching. The goal of the observations and post observation dialogue should be to provide information for a professional development plan and improve instruction and not to simply check a box to fulfill a school district teacher evaluation protocol. The NCTM recommends five content-related standards for evaluators to use when obtaining information through observation and assessing classroom teaching. In particular, the assessment of teaching should show that a teacher

- demonstrates a sound knowledge of mathematical concepts and procedures;
- represents mathematics as a network of interconnected concepts and procedures;
- emphasizes connections between mathematics and other disciplines and connections to daily living; engages students in tasks that promote the understanding of mathematical concepts, procedures, and connections;
- engages students in tasks that promote the understanding of mathematical concepts, procedures, and connections;
- and engages students in mathematical discourse that extends their understanding of mathematical concepts, procedures, and connections (Standard Four, Mathematics Concept, Procedures and Connections)

Additionally, evaluation should include evidence that the teacher emphasizes and models problem solving, mathematical reasoning, communication and discourse among students. Finally, the NCTM states that
mathematics teachers should promote a positive mathematical disposition, encourage students to persevere, foster confidence, appropriately assess students’ understanding of mathematics and create a productive and respectful learning environment.

These recommendations for teacher evaluation support the research literature, which recommends evaluation models that include multiple sources for data collection and they also align with Virginia’s teacher evaluation plan. Despite the NCTM’s recommendation for mathematics specific dialogue and evidence of content mastery, very few studies have taken a look at subject-specific observation and feedback (McDonald, 2008; Nelson & Sassi, 2000). Nelson & Sassi (2000) examined the nature of administrators’ observations of a video-recorded fifth grade mathematics lesson and found that administrators appreciated different aspects of the lesson during their first observation then during a second viewing, eight months later. During the first observation, administrators were appreciating the structural features of the lesson including “orderliness, good classroom management, understandable and well-executed structural components to the lesson and teacher behaviors such as wait time and gender equities (p. 565).” After viewing the video a second time and at least 8 months into a professional development seminar for administrators on observation and supervision of elementary mathematics, the administrators were observing subject-specific features of the lesson. For example, administrators noticed the nature of the students’ mathematical discourse. The observation shifted from teacher action and surface features of instruction to the development of ideas. The findings from this study also indicate that sense making develops differently in different disciplines and content and pedagogy are intertwined in teachers’ instructional decision making. The relationship between content and pedagogy is unique to each discipline due to subject specific procedures, language and concepts (Nelson & Sassi, 2000) and this must be taken into consideration when preparing supervisors for observing and evaluating mathematics teachers.

Conclusion

For decades, mathematics education has received national attention. Due to more recent policy shifts, teacher evaluation is a priority at the federal, state and local levels. While many studies have examined teacher evaluation, there is a void in the literature pertaining to the subject matter knowledge of observers and the type of post-observation feedback that is provided to teachers. Teacher evaluation protocols are not content specific; therefore, administrators are observing and giving feedback to teachers in all content areas regardless of the subject-matter background of the principal. Considering federal policy charges teacher evaluation to drive professional development, promotion and retention, it is critical to understand the types of feedback that are being given to teachers from their evaluators. This information has implications for the credibility of evaluation systems and the usefulness of feedback from administrators.
References


