PEABODY JOURNAL OF EDUCATION, 83: 562–591, 2008

ISSN: 0161-956X print / 1532-7930 online
DOI: 10.1080/01619560802414940

Routledge
Taylor & Francis Group

The Assessment of Mathematics and Science Teacher Quality

Patricia S. Moyer-Packenham
Utah State University

Johnna J. Bolyard
West Virginia University

Anastasia Kitsantas and Hana Oh
George Mason University

The purpose of this study was to examine the types of instruments being used to document mathematics and science teacher quality characteristics in 48 nationally funded mathematics and science education awards. Each of the 48 projects operationalized teacher quality and determined how to assess it. The main research questions examined the instruments awardees used to gather data on mathematics and science teacher quality, and the main characteristics of teachers examined by awardees. Results showed that awardees most frequently used surveys or questionnaires to assess characteristics of mathematics and science teacher quality. The most common teacher characteristics examined by awardees' included teacher behaviors, practices, and beliefs, followed by the assessment of subject and pedagogical knowledge, and the documentation of mathematics and science teachers' certification. A few new instruments were under development and in use to assess characteristics of teacher quality. Detailed information on the development and psychometric properties of the instruments used for these examinations was not available from the reports. Because awardees were at different stages in their funded activities and data collection efforts were ongoing at the time of this analysis, this study offers a preliminary and formative review of the use of assessments to document mathematics and science teacher quality characteristics among these awards.

In recent years, educators, researchers, and policymakers have sought to identify the characteristics of a highly qualified teacher (No Child Left Behind [NCLB],

Correspondence should be sent to Patricia S. Moyer-Packenham, Utah State University, College of Education, 2805 Old Main Hill, Logan, UT 84321. E-mail: patricia.moyer-packenham@usu.edu

2002). This goal presents a challenge because the literature on teacher quality is extensive and examines a wide range of empirical studies on teacher characteristics assumed to reflect teacher quality (Darling-Hammond, 2000; Darling-Hammond & Youngs, 2002; Rice, 2003; Wilson & Floden, 2003; Wilson, Floden, & Ferrini-Mundy, 2001). The goal is of particular importance to the mathematics and science education community where reports of international comparisons show that student performance in the United States is less than desirable in these subject areas (Hiebert et al., 2003). Student performance is often attributed to the quality, or lack thereof, of K-12 mathematics and science teaching. Although there is agreement that teacher quality is important, there is great variability in operationalizing the construct and even more variability in assessing it (Rice, 2003). Therefore, operationalizing and assessing quality, specifically in terms of mathematics and science teaching, is also yet to be clarified. This leads us to question, What have researchers learned about assessing mathematics and science teacher quality?

Current reform efforts have brought increased funding for national initiatives focusing on the quality of teachers in mathematics and science (see, e.g., http://www.ed.gov/ or http://nsf.gov/). This funding has resulted in some of the most cutting edge research on mathematics and science teacher quality in funded awards throughout the country, including the National Science Foundation's Math and Science Partnership (NSF MSP) Program. The NSF states the following as goals of the MSP Program:

> MSP serves students and educators by emphasizing strong partnerships that tackle local needs and build grassroots support to:
>
> - Enhance schools' capacity to provide challenging curricula for all students and encourage more students to succeed in advanced courses in mathematics and the sciences;
> - Increase the number, quality and diversity of mathematics and science teachers, especially in underserved areas;
> - Engage and support scientists, mathematicians, and engineers at local universities and local industries to work with K-12 educators and students;
> - Contribute to a greater understanding of how students effectively learn mathematics and science and how teacher preparation and professional development can be improved; and
> - Promote institutional and organizational change in education systems—from kindergarten through graduate school—to sustain partnerships' promising practices and policies. (NSF, 2007)

The study presented here was designed to examine one aspect within these goals, namely, the instruments used by the MSP awards as part of their efforts toward documenting mathematics and science teacher quality (Item 2). In the

2005 NSF Committee of Visitor's review of the MSP Program, in the section of the report focusing on "Results: Outputs and Outcomes of NSF Investments," the Committee of Visitor's review indicated,

> Processes for measuring growth in teacher content knowledge and effectiveness are less well-developed, but NSF should pay attention to pre- and post-testing of teachers, to classroom observation, and in general to ensuring that across projects the growth of teacher knowledge can be measured. (NSF, 2005, p. 17)

Our study is an effort to respond to this review by initially examining the types of instruments used by awardees in the MSP Program to gather data on characteristics of mathematics and science teacher quality. Our investigation focused on three areas: (a) the *characteristics* of mathematics and science teacher quality being assessed; in other words, how mathematics and science teacher quality was defined and operationalized by awardees in the MSP Program; (b) the *instrumentation* being used by awardees for teacher assessment; and (c) the *psychometric properties* of the instruments.

In the following sections, we describe the literature that led to the assignment of categories of instruments, describe instruments used to assess mathematics and science teacher quality by awardees in the MSP Program, and review the teacher quality characteristics the awardees examine. Because awardees were at different stages in their funded activities and data collection efforts were ongoing at the time of this analysis, this study offers a preliminary and formative review of the use of instruments to document mathematics and science teacher quality characteristics among these awards.

WHAT TEACHER QUALITY CHARACTERISTICS ARE EXAMINED IN RESEARCH?

There are six characteristics commonly identified by researchers in studies examining the quality of mathematics and science teachers (Bolyard & Moyer-Packenham, 2008). These characteristics include teacher behaviors, practices, and beliefs; subject knowledge; pedagogical knowledge; experience; certification status; and general ability. Among these characteristics are variables gathered through assessment measures (i.e., responses to test items or teaching performance during an observation) and nonassessment measures (i.e., highest degree obtained or number of years of teaching experience; American Statistical Association, 2007). A definition of teacher quality is sometimes defended by the relationship that research has found between a teacher variable and some other variable, often student achievement. As we present some of the relevant research findings, it is important to keep in mind the controversy involved in such a definition. Teachers

are associated with high or low student achievement test scores even when they are not in control of the characteristics of the students assigned to their classes, and they are not in control of other events that happen to their classrooms that are unpredictable.

Teachers' behaviors, practices, and beliefs provide important information about mathematics and science teacher quality. This aspect of teacher quality is usually the subject of studies using observational methods or self-report data. For example, in one observational study researchers found that 15% of observed mathematics and science lessons were categorized as high quality, whereas 27% and 59% were labeled medium and low quality, respectively (Hiebert et al., 2003; Weiss & Pasley, 2004; Weiss, Pasley, Smith, Banilower, & Heck, 2003). Some observational studies show associations between practices of high school science teachers and better classroom discipline (Druva & Anderson, 1983) and kindergarten teachers' instructional practices and student gains in mathematics (Guarino, Hamilton, Lockwood, & Rathbun, 2006). Further results indicate that teachers often decide how to teach content and those decisions are influenced by teachers' beliefs. For example, Staub and Stern (2002) found that elementary students of teachers who held more constructivist beliefs did better on word problem tests than students whose teachers used a more direct-instruction approach. Other research indicates a positive relationship between teachers' reported use of standards-based instruction and student achievement (Hamilton et al., 2003).

Subject knowledge is a highly valued characteristic of mathematics and science teachers and refers to the teacher's knowledge of mathematics and science content. Reviews of research indicate links between teachers' subject preparation and effectiveness, although these results are not always clear (Darling-Hammond, 2000; Darling-Hammond & Youngs, 2002; Rice, 2003; Wilson & Floden, 2003; Wilson et al., 2001). Results of studies examining the relationship between teachers holding subject specific degrees and student achievement vary, although mathematics results are generally positive (Chaney, 1995; Goldhaber & Brewer, 1997a, 2000; Rowan, Chiang, & Miller, 1997). Similarly, studies measuring teachers' subject knowledge using undergraduate or graduate coursework in the subject generally show a positive relationship with students' mathematics achievement (Chaney, 1995; Monk, 1994; Monk & King, 1994). Effects of subject matter coursework in science are often dependent upon the area of science studied (i.e., physical, earth, or life sciences; Chaney, 1995; Druva & Anderson, 1983; Monk & King, 1994). The data suggest a generally positive relationship between subject-specific mathematics and science coursework and student achievement. Some authors describe the intersection of subject-specific knowledge and pedagogy as *pedagogical content knowledge* (Shulman, 1986) or *mathematical knowledge for teaching* (Hill & Ball, 2004); however, this aspect of teacher knowledge is yet to be widely utilized as a research variable in studies on teacher quality.

Teachers' *pedagogical knowledge*, or knowledge of teaching, is often researched as evidence of teacher quality using data such as degrees in education, educational coursework, and scores on exams measuring professional knowledge. Researchers have reported positive effects of teachers' pedagogical knowledge and preparation (Adams & Krockover, 1997; Ferguson & Womack, 1993; Grossman & Richert, 1988; Grossman et al., 2000; Guyton & Farokhi, 1987; Hansen & Feldhusen, 1994; Valli & Agostinelli, 1993). Generally, studies of teachers' pedagogical knowledge find positive relationships between education training and teacher effectiveness (Darling-Hammond, 2000). Courses taken in subject-specific pedagogy (i.e., mathematics education or science education) also appear to have a positive impact, particularly in mathematics at the middle and secondary level (Chaney, 1995; Monk, 1994). However, other results show little or no relationships (e.g., Rivkin, Hanushek, & Kain, 2005). Wilson and Floden (2003) noted that much of the research focuses on teacher education programs rather than on specific courses or experiences.

Some studies report positive relationships between teachers' years of *experience* and teacher effectiveness (Ehrenberg & Brewer, 1995; Ferguson, 1991; Fetler, 1999; Goldhaber & Brewer, 1997b; Greenwald, Hedges, & Laine, 1996; Hanushek, 1992, 1996). Reviewing studies examining the relationship between teacher experience and student achievement, Rice (2003) concluded a positive relationship between these variables, which was more pronounced during the first years of teaching at the elementary level and more constant at the secondary level. Although characteristics such as teacher experience and education are commonly identified as favorable characteristics in the teacher hiring process, some researchers argue that little of the variation in teacher quality is explained by these variables (Rivkin et al., 2005).

Mathematics and science teachers' *certification* status is used as an indicator of knowledge gained from teacher preparation (Darling-Hammond, 2000; Darling-Hammond & Youngs, 2002). Certification refers to the types of teaching certificates one holds (e.g., secondary mathematics certificate, algebra endorsement, or physical science certification). Researchers compare those who are fully certified and those who hold provisional or emergency certification (Darling-Hammond, 2000; Fetler, 1999; Goldhaber & Brewer, 2000). Several studies indicate an advantage in favor of fully certified teachers on measures of student achievement and teacher performance evaluations (Darling-Hammond, 2000; Fetler, 1999). Mathematics student achievement has been found to be positively associated with having a teacher who is certified in-field (Goldhaber & Brewer, 1997b; Hawk, Coble, & Swanson, 1985).

Teacher's *general intellectual abilities*, that is, those verbal and quantitative abilities that frequently qualify individuals for higher education, are also considered aspects of teacher quality. Studies generally report a positive relationship between measures of teachers' general and verbal abilities and their effectiveness

(Ehrenberg & Brewer, 1994; Ferguson, 1991; Ferguson & Ladd, 1996; Greenwald et al., 1996; Hanushek, 1971; Strauss & Sawyer, 1986). Other studies indicate mixed or negative results (Ehrenberg & Brewer, 1995; Hanushek, 1992; Murnane & Phillips, 1981).

In this section we have classified the characteristics that researchers of teacher quality have included in their studies. We now turn to the ways that these characteristics have been measured.

WHAT INSTRUMENTS ARE USED TO MEASURE TEACHER QUALITY?

Although much of the literature on teacher quality focuses on characteristics of teachers, there is less focus on the instrumentation used to gather data on those characteristics. In many cases, proxies, or substitutes for teacher quality characteristics, are used to measure the mathematics and science teacher quality construct, prompting different interpretations of the results in these studies (Darling-Hammond & Youngs, 2002; Wilson & Floden, 2003; Wilson et al., 2001). Some proxies are a better representation of the teacher quality characteristic than others. For example, studies use teachers' college majors as evidence of pedagogical and subject knowledge. However, a college major does not illuminate specific knowledge gained through such training or account for variations in programs among colleges and universities. The use of certification status is also common (Darling-Hammond, 2000; Goe, 2002; Goldhaber & Brewer, 2000). Yet states set their own certification criteria, and therefore, the skills and knowledge represented by a teacher's certification varies from state to state. Another difficulty is that teacher quality researchers sometimes use several variables that are highly correlated with each other. For example, education levels are highly correlated with age, experience, and general ability, and certification is often correlated with educational training and subject knowledge background (Darling-Hammond & Youngs, 2002). Combined with variations in units of analysis and methodological approaches, researchers may obtain conflicting results based on the same teacher characteristics.

Common instruments used to gather data on teacher quality in mathematics and science include written surveys and questionnaires, behavioral observations, exams, interviews, portfolios, and archival records. Researchers use written *surveys and questionnaires* to gather information about teachers' classroom practices and beliefs about teaching and learning (Darling-Hammond, Chung, & Frelow, 2002). Some surveys gather information on beginning teachers' professional concerns and opinions about their preparation (Darling-Hammond et al., 2002; Houston, Marshall, & McDavid, 1993; Sandlin, Young, & Karge, 1992). Surveys are sometimes used to gather information about teachers' entry into the profession (Andrew,

1990; Andrew & Schwab, 1995; Darling-Hammond et al., 2002), their perceptions of teaching as a profession (Lutz & Hutton, 1989), and their intention to remain in the profession (Darling-Hammond et al., 2002). Other surveys collect background information on teachers to use as representations of teacher quality characteristics (i.e., number of graduate and undergraduate courses taken, undergraduate institution, certification status, and major; Andrew, 1990; Andrew & Schwab, 1995).

Behavioral observations are often used to gather information on teachers' pedagogical knowledge and instructional practices. Observation protocols gather information on teachers' classroom management and instructional skills (Sandlin et al., 1992) and look for evidence of the use of best practices (Hawk et al., 1985; Miller, McKenna, & McKenna, 1998). Some observation data are examined to determine relationships between teachers' preservice preparation and their practices, knowledge, and beliefs (Adams & Krockover, 1997; Ferguson & Womack, 1993; Grossman, 1989; Grossman & Richert, 1988; Grossman et al., 2000; Hansen & Feldhusen, 1994). Generally these studies involve small sample sizes and combine observational data with data gathered through other sources. Observations of teacher behaviors and classroom practices provide a rich source of data, and there are several studies that have examined teachers' practices on a large scale (see, e.g., Weiss et al., 2003).

Scores on *exams* have been used to measure teacher characteristics such as subject knowledge, pedagogical or professional knowledge, and general or verbal ability. Exams are of two types: those used to measure subject knowledge created specifically for a study, and standardized exams such as the National Teachers Examination Subject Area Specialty exams (Hawk & Schmidt, 1989; Rowan et al., 1997) and the Praxis Subject Area exams. Exams used to measure teachers' pedagogical or professional knowledge include state and national certification exams such as the National Teachers Examinations Test of Professional Knowledge exam (Hawk & Schmidt, 1989). Some researchers have developed exams designed to measure the mathematical knowledge that teachers use in their work, or *mathematical knowledge for teaching* (MKT) (see, e.g., Hill & Ball, 2004). Scores on college entrance exams, such as ACT and SAT, and tests of verbal aptitude or basic literacy, are often used to measure teachers' general or verbal ability (Ferguson, 1991; Ferguson & Ladd, 1996; Hanushek, 1992).

Interview protocols are used to gather information on characteristics such as teachers' pedagogical knowledge and beliefs on teaching and learning. Interview data are often examined to determine relationships between teachers' preservice preparation and their practices, knowledge, and beliefs. Interview protocols are commonly used in conjunction with other instruments such as observations and surveys (Adams & Krockover, 1997; Ferguson & Womack, 1993; Grossman, 1989; Grossman & Richert, 1988; Grossman et al., 2000; Hansen & Feldhusen, 1994).

Portfolios and other written documents are analyzed as evidence of teachers' pedagogical skills and knowledge (Guyton & Farokhi, 1987). For example, one

study analyzed classroom artifacts (lesson plans and other teaching documents) from 10 beginning teachers to determine impacts of teacher education (Grossman et al., 2000). To apply for National Board Certification, teachers create teaching portfolios that contain videotapes of their teaching, evidence of student learning products, and a detailed analysis of their teaching practices (National Board for Professional Teaching Standards, http://www.nbpts.org).

Archival records often contain background information on teachers including degree completion, college transcripts and grade point average, college entrance exam scores, scores on professional certification exams, certification status, and years of experience. Data on certification status, degree completion, and graduate and undergraduate courses taken are often used as evidence of teachers' pedagogical and/or subject matter preparation (Chaney, 1995; Darling-Hammond, Holtzman, Gatlin, & Heilig, 2005; Fetler, 1999; Laczko-Kerr & Berliner, 2002; Monk, 1994; Rowan et al., 1997). The information is often gathered in and accessed through state and national databases.

In this section we have reviewed a variety of instruments commonly used to gather data on the quality of individual teachers. At this point we turn our attention to the characteristics of teacher quality identified by awardees in the MSP Program and the instruments used by awardees to assess those characteristics. Our analysis focused on the following research questions: (a) What instrumentation is being used by awardees to assess teacher quality characteristics? Two subquestions emerged from this research question: Are the instruments locally or externally developed? What information is available regarding the psychometric properties of the instruments being used? The second research question was (b) What teacher characteristics are being assessed by the instruments? Subquestions included the following questions: How is subject knowledge (mathematics, science, and MKT) measured? In this case it was hypothesized that standard content tests would be used to assess subject knowledge. How is pedagogical knowledge measured? It was hypothesized that surveys and observations would be used to assess pedagogical knowledge. In a further analysis we examined similarities and differences among the awardees in terms of when they received their awards (i.e., Cohort I, II, and III awards, distributed to partnerships between 2002 and 2004).

METHODS

Data Sources

The data sources in this study came from funded partnerships in the NSF-MSP Program awarded between fiscal year (FY) 2002 and FY2004. The NSF describes the following four components that make up the MSP Program:

- Comprehensive Partnerships implement change across the K-12 continuum in mathematics, science, or both.
- Targeted Partnerships focus on improved student achievement in a narrower grade range or disciplinary focus in mathematics and/or science.
- Institute Partnerships develop mathematics and science teachers as school- and district-based intellectual leaders and master teachers.
- Research, Evaluation, and Technical Assistance (RETA) activities assist partnership awardees in the implementation and evaluation of their work (NSF, 2007).

Our study examined data from 48 awards in three of these categories including 12 Comprehensive Partnerships, 28 Targeted Partnerships, and 8 Institute Partnerships. RETA awards were not included in the analysis because of the nature and scope of their work in "assisting" the other award categories.

Each partnership is required to address the quality of the mathematics and science teaching force and to document its progress toward the teacher quality goals and benchmarks it has established. Awardees submit Annual and Evaluation Reports describing this progress. In this analysis, researchers reviewed 123 Annual and Evaluation Reports provided to the NSF, with the length of each report ranging from 29 to 707 pages. These reports, along with awardees Web sites, published papers, and presentations, were the secondary source documents for the analysis. Data reviewed for this article were obtained from documents available to researchers between January 2005 and February 2006.

DEFINING INSTRUMENT AND TEACHER QUALITY CHARACTERISTICS CATEGORIES

Based on the review of research, we determined a set of categories for types of instruments and a set of categories for teacher quality characteristics. The following sections define each of these categories and describe how they were used in the analysis.

Instrument Categories

To focus the scope of the analysis, researchers determined the following critera for the instruments that would be included in the analysis. One criterion was that the instrument needed to gather data on *teacher* quality, and the analysis was confined to instruments used with teachers. Teachers were defined as those whose primary instructional responsibilities were in the classroom with students for at least 50% of a school day. There were a variety of instruments in use among the

awards that collected data on attributes of other school positions (i.e., principals, administrators, curriculum specialists). Researchers selected those instruments that collected data on *teachers* for inclusion.

Another criterion was that the instruments needed to be used to collect data on *individual characteristics* of teachers. Individual teacher characteristics identified in the research included teacher behaviors, practices, and beliefs; subject knowledge; pedagogical knowledge; experience; certification status; and general ability (Bolyard & Moyer-Packenham, 2008). Instruments that collected data on teacher quantity and diversity, such as numbers of participants in courses and demographics on teacher race and ethnicity, were beyond the scope of our analysis because they focused on characteristics of teachers as a group or population rather than on the quality of the individual teacher. In-depth examinations of teacher quantity and diversity are the focus of other investigations in the MSP Program Evaluation (Moyer-Packenham, Bolyard, Oh, Kridler, & Salkind, 2006; Moyer-Packenham, Parker, Bolyard, Kitsantas, & Huie, 2008; Tyler & Vitanova, 2007).

We used the definition of an *instrument* based on research compiled by Prus and Johnson (1994) for categorizing instruments. This categorization system included six types of instruments: (a) written surveys and questionnaires, (b) behavioral observations, (c) exams, (d) exit and other interviews, (e) portfolios, and (f) archival and other records. By using this system of categorization, we limited the scope of the analysis, thereby excluding some types of data that were collected by the awardees. For example, many teachers in the partnerships attended courses and workshops to improve their knowledge and practices. When the MSPs reported *offering* a course or *numbers of teachers taking* a course, we had no way of knowing what teacher characteristics were impacted and what types of instruments were used in the course, and therefore course participation was not captured in this analysis. However, when the awardees reported their use of exams, interviews, or any other instruments to document teacher characteristics during or following courses, these instruments were included in our analysis. This type of focused examination ensured that the teacher characteristics assessed were linked directly by the awardees themselves with the instruments used to document the characteristics.

In this section we provide specific detail on the instrument categories as they relate to the present study. A *survey or questionnaire* was a document where respondents replied to questions or comments in writing, often choosing from a given set of answers (Fraenkel & Wallen, 1993). *Behavioral observations* included instruments, such as protocols, which categorize teacher behaviors and performances in a natural setting such as a classroom (Miles & Huberman, 1984; Prus & Johnson, 1994; Schloss & Smith, 1999). *Exams* were those instruments administered to teacher–participants as part of the awardees' activities. This category often included instruments designed to test knowledge in one or more areas (i.e., mathematics or science; Fraenkel & Wallen, 1993; Prus & Johnson, 1994), through multiple-choice, short-answer, and essay formats, among others,

and included instruments developed externally (by an individual or group outside the award) and those developed locally (by an individual working within the award; Fraenkel & Wallen, 1993; Lopez, 1998; Prus & Johnson, 1994). *Exit and other interviews* required participants to discuss their perceptions, beliefs, knowledge, or experiences often in a face-to-face setting with questions posed by an interviewer (Fraenkel & Wallen, 1993; Prus & Johnson, 1994). *Portfolios* included collections of work samples and other documents produced and compiled by teachers over time, with the portfolios most often assessed using a rubric (Hart, 1994; Paulson, Paulson, & Meyer, 1991; Prus & Johnson, 1994). *Archival records* included documents regarding background and demographic information, or other file data (Prus & Johnson, 1994). In our study, this information was often provided by an existing file compiled by a university or school district and included data on teacher certification status, teacher exam scores, and years of experience. When the score from an exam was gathered from instruments not administered by awardees during their activities, and was obtained from external database sources, these were categorized as archival records rather than exams.

The final category, *unspecified*, was added and included instruments for which awardees did not provide sufficient information to determine the assessment being used. In these cases, awardees described assessing a particular teacher characteristic but did not specify the instrument used in the assessment. A cross-checking method was used to search Web sites, conference papers, and other available documents in an attempt to identify these instruments. The *unspecified* category was used when no additional information was available following this search. Researchers looked for examples of the instruments among the documents to determine the content of each instrument.

Teacher Quality Characteristics Categories

Researchers used the following six categories for teacher quality characteristics identified in a literature review conducted by Bolyard and Moyer-Packenham (2008): (a) teacher behaviors, practices, and beliefs; (b) subject knowledge; (c) pedagogical knowledge; (d) experience; (e) certification status; and (f) general ability. An additional category, *unspecified,* was used when the specific teacher characteristic being assessed could not be determined based on the descriptive information provided by awardees. As in the case of instruments, a cross-checking method was used to search other available documents for this information. The following section describes each of the teacher quality characteristics categories as they relate to our study.

The category teacher behaviors, practices, and beliefs was further defined in two subcategories: *teacher behaviors and practices* and *teacher beliefs.* The teacher behaviors and practices category included what the teacher does in the classroom, for example, questioning strategies, instructional equity, classroom management,

and use of time. Teacher beliefs included beliefs about students' learning, such as beliefs about the way students learn content and beliefs about who can and cannot learn, and beliefs about content, such as teachers' views on the nature of the content and the best methods for teaching it. *Subject knowledge* refers to teachers' knowledge and understanding of concepts and topics related to specific content (Ferguson & Womack, 1993; Monk, 1994). In our study, subject knowledge refers to knowledge of mathematics and science content, and MKT. MKT, as defined by Hill and Ball (2004), is the specialized kind of content knowledge needed to teach mathematics and is part of the work of one of the RETA awards in the NSF-MSP Program. *Pedagogical knowledge* refers to knowledge of teaching and learning including knowledge of students' cognitive development, learning theories, and instructional approaches and strategies. *Experience* is defined as the total number of years a teacher has been teaching and/or the number of years a teacher has taught a specific grade level or subject area, although researchers note that experience can also include the substance, variety, and quality of one's experiences. *Certification* describes teachers' certification status (including whether they are emergency, provisionally, or fully certified), whether a teacher is certified in the field in which they are teaching, and whether teachers are *highly qualified* as defined by NCLB (2002). *General ability* refers to teachers' general intellectual academic and verbal abilities, often including evidence of language and mathematical proficiency.

Procedures

Researchers conducted a preliminary analysis of the secondary source documents that focused on understanding the major themes of teacher quality, quantity, and diversity among the work of awardees prior to our study. This preliminary analysis indicated that the awardees in this program were engaged in a variety of activities designed to influence teacher quality, quantity, and diversity and that they had implemented numerous strategies for assessing their progress. The prior examination showed that the data collected on teacher quality primarily focused on changes in teachers' subject and pedagogical knowledge, their practices and beliefs, and their certification status. The data on teacher quantity focused on numbers of teachers participating in MSP activities and activities of the schools and universities associated with the MSP award. Data on teacher diversity focused on reporting race and ethnicity of participating teachers. Overall, the preliminary analysis showed that interventions identified by the awardees as influences on teacher quality, quantity, and diversity characteristics included new programs and coursework; professional development; teacher leadership; recruiting; preservice training; compensation; retention; linking science, technology, engineering, and mathematics (STEM) faculty with teachers; and induction. These results are discussed in another Math and Science Partnership Program Evaluation (MSP-PE) manuscript (Moyer-Packenham et al., 2006).

Building on this prior analysis, the team of researchers examined the secondary documents to locate information on the *instruments* in use by MSP awardees. The prior analysis indicated that there were numerous instruments in use among the awards. The challenge faced by researchers was in extracting this information because it was scattered in a variety of different locations throughout the reports. Researchers found that some awardees described numerous instruments, whereas others included little information about their instruments in the reports. In many cases, the actual instruments themselves were described by awardees but were not included in the reports.

Researchers used the previously described definitions for instrumentation and teacher quality characteristics to sort and classify the data, compiling the following information for each instrument: the name of the award using the instrument, the name of the instrument, the teacher quality characteristic assessed, type of instrument, source of the instrument (local or external to the award), information on psychometric properties, and instrument availability (whether a copy of the instrument was included in the reports or other documents). The research team scanned reports from the RETA awards of the MSP Program to cross check for instruments that might be under development in the RETAs and determine if these were in use by awardees. Instruments were categorized along two dimensions: the type of instrument used and the teacher characteristics assessed by the instrument. These categories were analyzed by examining relationships and using descriptive and chi-square tests.

RESULTS

The first research question examined all instruments being used by awardees. A total of 282 instruments were identified across the 48 awards. Figure 1 shows the distribution of these instruments. This is an average of almost six instruments reported per award (5.88) at the time of our preliminary analysis. As Figure 1 shows, every awardee identified at least 1 instrument (three reported only 1), and some reported as many as 10, 12, or even 15 instruments.

As shown in Table 1, the majority of instruments used across the 48 awards were survey/questionnaires (37.9% of all the instruments identified) used by 87.5% of the awards. These were followed by exams (16.0%) used by 62.5% of awards, behavioral observations (14.2%) used by 62.5% of awards, exit and other interviews (10.6%) used by 50% of awards, portfolios (7.1%) used by 29.2% of awards, archival records (10.6%) used by 45.8% of awards, and finally instruments that were unspecified (3.5%) used by 16.7% of awards.

The 107 surveys and questionnaires that were identified collected data on a wide range of topics from several different teacher audiences. One example was a survey intended for teacher participants focusing on their perceptions of changes

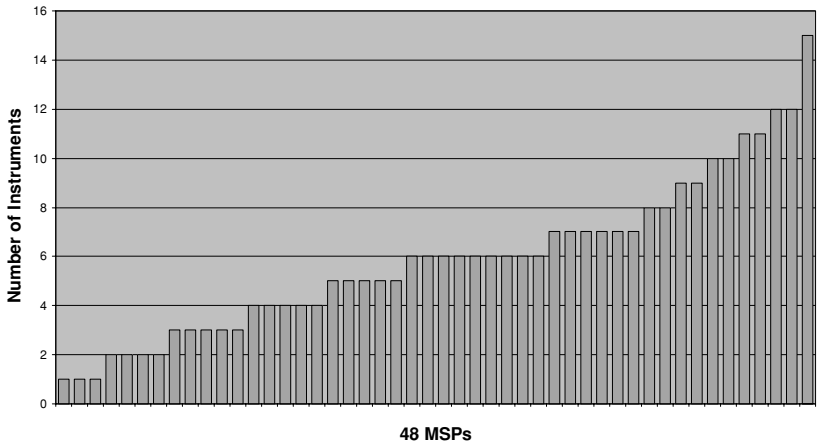

FIGURE 1 Distribution of instruments across Math and Science Partnership (MSP) awards.

in their knowledge, skills, and practices as a result of participation in an activity. In this example, the survey asked teachers about their perceptions of changes in their own knowledge rather than assessing their knowledge directly. One of the 45 exams assessed respondents' mathematical knowledge about precalculus concepts. Another exam was designed to measure growth in secondary teachers' knowledge of algebra and geometry. In the 40 behavioral observations, a variety of instruments asked observers to record information including demonstrated level of teachers' subject knowledge, tools and strategies employed, cognitive level of tasks, instructional equity, and lesson implementation. One example of the 30

TABLE 1
Frequency and Percentage of Instruments Used Across the Awards

	Award Frequency[a]		
Instrument	Not Used	Used	Instrument Frequency[b]
Written surveys and questionnaires	6 (12.5%)	42 (87.5%)	107 (37.9%)
Exams	18 (37.5%)	30 (62.5%)	45 (16.0%)
Behavioral observations	18 (37.5%)	30 (62.5%)	40 (14.2%)
Exit and other interviews	24 (50.0%)	24 (50.0%)	30 (10.6%)
Portfolios	34 (70.8%)	14 (29.2%)	20 (7.1%)
Archival records	26 (54.2%)	22 (45.8%)	30 (10.6%)
Unspecified	40 (83.3%)	8 (16.7%)	10 (3.5%)

[a] $N = 48$; [b] $N = 282$

interviews used by awardees included an interview protocol designed to elicit information on changes in the teachers' own practices and their students' learning as a result of participation in the partnership. Among the 20 portfolios were those that analyzed teachers' writing in online logs to document successes, challenges, and concerns as teachers implemented award goals over time. Others focused on teachers' lesson plans to document changes in teachers' practices. The 30 archival records were documents that contained summative information about teacher licensure and certification status, years of experience, levels of education, grades and examination scores, and general ability measures (i.e., SAT or GRE scores).

In addition to examining type and frequency, researchers also determined whether the instruments were locally or externally developed, see Table 2. This examination was constrained to the documents available for analysis and was therefore limited in its scope. Locally developed instruments were those developed by awardees, whereas externally developed instruments were those developed by someone external to the award. This analysis revealed that the same number of surveys and questionnaires were locally developed and externally developed (30, or 28.0%), and 47 (43.9%) were not identified. Among the behavioral observations, 12 (30.0%) were locally developed, 17 (42.5%) were externally developed, and 11 (27.5%) were not identified. Exams tended to be externally developed (25, or 55.6%), whereas 9 (20.0%) were locally developed, and 11 (24.4%) were not identified. Most of the interview instruments were not identified (18, or 60.0%), 8 (26.7%) were locally developed and 4 (13.3%) were developed externally. In terms of the portfolios, 9 (45.0%) were locally developed, 2 (10.0%) were developed externally, and 9 (45.0%) were not identified. Finally, 1 (10%) of the unspecified documents was locally developed and 9 (90.0%) were not identified.

Next researchers examined the psychometric properties of the locally developed instruments. These results are also presented in Table 2. For 26 (86.7%) and 27 (90.0%) surveys and questionnaires there was no information reported about the validity and reliability, respectively. However, 4 (13.3%) reported validity information and 3 (10.0%) reported reliability. Similar patterns were observed for the behavioral observations instruments (1 of 12 reported validity and reliability) and exams (2 and 1 of 9 reported validity and reliability, respectively). No psychometric properties were reported for interviews or portfolios. Archival records were not included in this table because psychometric properties can not be established for this type of instrument.

Researchers conducted further investigations of the number of awards using exam instruments to measure types of subject-specific knowledge, including mathematics, science, and MKT; see Table 3. Among the 48 MSPs were awards that focused on mathematics only, science only, and a combination of mathematics and science. There were 40 awards that included mathematics, and 27 awards that included science. Of the 40 awards that included mathematics, 17 awards (42.5%)

TABLE 2
Frequency and Percentage of Instrument Properties Across the Awards

Instrument	*N*	Frequency	Locally Developed: Psychometric Properties				Externally Developed Frequency	Development Not Identified Frequency
			Validity Not Reported	Validity Reported	Reliability Not Reported	Reliability Reported		
Written surveys and questionnaires	107	30 (28.0%)	26 (86.7%)	4 (13.3%)	27 (90.0%)	3 (10.0%)	30 (28.0%)	47 (43.9%)
Behavioral observations	40	12 (30.0%)	11 (91.7%)	1 (8.3%)	11 (91.7%)	1 (8.3%)	17 (42.5%)	11 (27.5%)
Exams	45	9 (20.0%)	7 (77.8%)	2 (22.2%)	8 (88.9%)	1 (11.1%)	25 (55.6%)	11 (24.4%)
Exit and other interviews	30	8 (26.7%)					4 (13.3%)	18 (60.0%)
Portfolios	20	9 (45.0%)					2 (10.0%)	9 (45.0%)
Unspecified	10	1 (10.0%)					0 (0.0%)	9 (90.0%)

Note. The 30 archival records are not included in this table.

TABLE 3
Awardees' Use of External and Local Exams to Measure Types of Subject Knowledge

	Subject Knowledge Category							
	Exams Combined		Externally Developed Exams		Locally Developed Exams		Development Not Identified	
	Not Used	Used	Not Used	Used	Not Used	Used	Not Used	Used
Math content[a]	23 (57.5%)	17 (42.5%)	33 (82.5%)	7 (17.5%)	36 (90.0%)	4 (10.0%)	32 (80.0%)	8 (20.0%)
Science content[b]	14 (51.9%)	13 (48.1%)	21 (77.8%)	6 (22.2%)	23 (85.2%)	4 (14.8%)	23 (85.2%)	4 (14.8%)
MKT[a,c]	27 (67.5%)	13 (32.5%)	27 (67.5%)	13 (32.5%)	40 (100%)	0 (0.0%)	40 (100%)	0 (0.0%)

Note. Some awards use more than one type of exam with different sources of development; therefore numbers in the rows do not sum.
[a] $N = 40$ *Math-focused* Math and Science Partnerships.
[b] $N = 27$ *Science-focused* Math and Science Partnerships.
[c] Mathematical Knowledge for Teaching (MTK) as defined by Hill and Ball (2004).

used mathematics content exams to measure subject knowledge and 13 awards (32.5%) used the MKT instrument. Of the 27 awards that included science, 13 awards (48.1%) used science content exams to measure subject knowledge. Next we determined whether awards used exam instruments that were locally or externally developed. This analysis revealed that seven (17.5%) awards measuring mathematics content used exams that were externally developed, whereas four (10.0%) awards used locally developed exams, and eight (20.0%) used mathematics content exams whose development was not identified. In regards to exams measuring science content, six (22.2%) awards used exams that were externally developed, whereas four (14.8%) awards used locally developed exams, and four (14.8%) used exams where development was not identified. All of the awards that measured MKT (13 or 32.5%) used an exam that was developed external to the award.

The second research question examined the teacher characteristics being assessed by the instruments. Table 4 provides the frequencies of the teacher characteristics measured and not measured. Based on these results, 41 (85.4%) awards focused on assessing teacher behaviors, practices, and beliefs, with some awards focusing specifically on teachers' behaviors and practices only (37 or 77.1%), and others focusing on teachers' beliefs only (31 or 64.6%). Thirty-nine (81.3%) awards reported assessing subject knowledge, including 27 of 40 (67.5%) mathematics awards measuring mathematics knowledge, 18 of 27 (66.7%) science

TABLE 4
Frequency and Percentage of Teacher Characteristics Examined by All Instruments Across the Awards

Teacher Characteristic	Frequency	
	Not Measured	Measured
Teacher behaviors, practices, and beliefs (combined)[a]	7 (14.6%)	41 (85.4%)
Teacher behaviors and practices	11 (22.9%)	37 (77.1%)
Teacher beliefs	17 (35.4%)	31 (64.6%)
Subject knowledge (combined)[a]	9 (18.8%)	39 (81.3%)
Math content[b]	13 (32.5%)	27 (67.5%)
Science content[c]	9 (33.3%)	18 (66.7%)
MKT[b,d]	27 (67.5%)	13 (32.5%)
Pedagogical knowledge	11 (22.9%)	37 (77.1%)
Certification	18 (37.5%)	30 (62.5%)
Experience	30 (62.5%)	18 (37.5%)
General ability	44 (91.7%)	4 (8.3%)
Unspecified	29 (60.4%)	19 (39.6%)

Note. $N = 48$.
[a]Combined totals reflect the number of awards measuring one or more characteristics in that category. [b] $N = 40$ *Math-focused* Math and Science Partnerships. [c] $N = 27$ *Science-focused* Math and Science Partnerships. [d]Mathematical Knowledge for Teaching (MKT) as defined by Hill and Ball (2004).

awards measuring science knowledge, and 13 of 40 (32.5%) mathematics awards measuring MKT. Pedagogical knowledge was assessed by 37 (77.1%) awards, whereas teacher certification was documented by 30 (62.5%) awards. Teacher experience and general ability were documented by 18 (37.5%) and four (8.3%) awards, respectively. Finally, 19 (39.6%) awards described instruments that measured teacher characteristics that could not be identified based on the descriptions in the reports.

Table 5 depicts the frequencies and percentages of the subquestions for research question two answering what teacher characteristics are being assessed. Regarding the first subquestion, how subject knowledge (combined) was assessed, nine awards used surveys and/or questionnaires, nine used behavioral observations, 30 used exams, four used interviews, five used portfolios, one used an archival record, and two awards did not specify. Pedagogical knowledge was assessed using surveys and/or questionnaires by 24 awards; 20 awards used behavioral observations, four used exams, 12 used interviews, seven used portfolios, two used archival records, and one award did not specify. Mathematics knowledge was assessed using surveys and/or questionnaires by five awards, whereas seven awards used behavioral observations, 17 used exams, three used interviews, four used portfolios, one used an archival record, and one did not specify. Science knowledge was assessed using

TABLE 5
Awardees' Use of Instruments to Document Types of Teacher Knowledge

	Instrument Uses													
	Surveys		Observations		Exams		Interviews		Portfolios		Archival Records		Unspecified	
Teacher Knowledge Category	Not Used	Used	Not Used	Used	Not Used	Used	Not Used	Used	Not Used	Used	Not Used	Used	Not Used	Used
Subject knowledge (combined)[a]	39 (81.3%)	9 (18.8%)	39 (81.3%)	9 (18.8%)	18 (37.5%)	30 (62.5%)	44 (91.7%)	4 (8.3%)	43 (89.6%)	5 (10.4%)	47 (97.9%)	1 (2.1%)	46 (95.8%)	2 (4.2%)
Math content[b]	35 (87.5%)	5 (12.5%)	33 (82.5%)	7 (17.5%)	23 (57.5%)	17 (42.5%)	37 (92.5%)	3 (7.5%)	36 (90.0%)	4 (10.0%)	39 (97.5%)	1 (2.5%)	39 (97.5%)	1 (2.5%)
Science content[c]	22 (81.5%)	5 (18.5%)	24 (88.9%)	3 (11.1%)	14 (51.9%)	13 (48.1%)	26 (96.3%)	1 (3.7%)	25 (92.6%)	2 (7.4%)	26 (96.3%)	1 (3.7%)	26 (96.3%)	1 (3.7%)
MKT[b,d]	40 (100%)	0 (0.0%)	40 (100%)	0 (0.0%)	27 (67.5%)	13 (32.5%)	40 (100%)	0 (0.0%)	40 (100%)	0 (0.0%)	40 (100%)	0 (0.0%)	40 (100%)	0 (0.0%)
Pedagogical knowledge[a]	24 (50.0%)	24 (50.0%)	28 (58.3%)	20 (41.7%)	44 (91.7%)	4 (8.3%)	36 (75.0%)	12 (25.0%)	41 (85.4%)	7 (14.6%)	46 (95.8%)	2 (4.2%)	47 (97.9%)	1 (2.1%)

Note. Some awards use multiple instruments; therefore numbers in the rows do not sum.
[a] $N = 48$ Math and Science Partnerships. [b] $N = 40$ *Math-focused* Math and Science Partnerships. [c] $N = 27$ *Science-focused* Math and Science Partnerships.
[d] Mathematical Knowledge for Teaching (MKT) as defined by Hill and Ball (2004).

surveys and/or questionnaires in five awards, whereas three awards used behavioral observations, 13 used exams, one used an interview, two used portfolios, one used an archival record, and one did not specify. Finally, MKT was measured using an exam in 13 of the 40 mathematics awards.

Chi-square tests were used to test the hypotheses that (a) standard content tests would be used to measure subject knowledge (mathematics, science, and MKT), rather than observations, surveys, portfolios, or interviews, whereas (b) surveys, observations, and interviews would be used to assess teacher's pedagogical knowledge rather than exams. Support for this hypothesis was found. First, in terms of mathematics knowledge, a significant χ^2 (6, $N = 146$) = 12.80, $p < .05$ was obtained, showing that exams were more often used to capture teacher content knowledge in mathematics. Similar results were revealed for science, χ^2 (6, $N = 146$) = 15.01, $p < .05$, and MKT, $\mathfrak{g}\chi^2$(6, $N = 146$) = 33.08, $p < .001$. Moreover, as hypothesized, awards used surveys and observations, χ^2 (6, $N = 146$) = 90.00, $p < .001$, to assess teachers' pedagogical knowledge, which is significantly different from the way that subject knowledge was measured.

Finally, in regards to the last research question researchers examined the data for similarities and differences among the awards in the types of teacher characteristics examined and the number and type of instruments used by Cohort I, II, and III awards (awarded between FY2002 and FY 2004). The first subquestion focused on the types of teacher characteristics assessed by different cohorts of awards. Essentially, this examination showed that the awards in each cohort were using similar instruments to gathering data on the same teacher quality characteristics, and no overall significant differences emerged for teacher characteristics; see Table 6. At a descriptive level, frequencies showed that 90.9% of Cohort I assessed teachers' behaviors, practices, and beliefs, as compared with 85.7% of Cohort II and 75.0% of Cohort III. This trend was similar for the assessment of subject knowledge by the awards in Cohorts I (77.3%), II (71.4%), and III (50%). Although 68.2% of Cohort I and 75.0% of Cohort III awards assessed pedagogical knowledge, a larger portion of the Cohort II awards (92.9%) assessed this characteristic. The assessment of certification status (Cohort I, 63.6%; Cohort II, 64.3%; and Cohort III, 58.3%) and teacher experience (Cohort I, 36.4%; Cohort II, 42.9%; and Cohort III, 33.3%) were similar across the three cohorts. All cohorts focused less on collecting data on general ability (Cohort I, 0.0%; Cohort II, 14.3%; and Cohort III, 16.7%).

In regards to the second part of the final research question, no significant differences were detected among the frequency of instruments within each instrument category among the Cohort I, II, and III awards; see Table 7. Descriptively, more instruments were used in each instrument type in relation to the year that the MSP was awarded their funding (i.e. Cohort I, awarded 2002, 138 instruments; Cohort II, awarded 2003, 92 instruments; and Cohort III, awarded 2004, 53 instruments). There were also more documents available for analysis from the awards that were

TABLE 6
Frequency and Percentage of Teacher Characteristics Examined by Cohort I, II, and III Awards

Teacher Characteristic	Cohort I (Awarded 2002)[a]		Cohort II (Awarded 2003)[b]		Cohort III (Awarded 2004)[c]	
	Not Measured	Measured	Not Measured	Measured	Not Measured	Measured
Teacher behaviors, practices, and beliefs	2 (9.1%)	20 (90.9%)	2 (14.3%)	12 (85.7%)	3 (25.0%)	9 (75.0%)
Subject knowledge	3 (13.6%)	19 (86.4%)	2 (14.3%)	12 (85.7%)	4 (33.3%)	8 (66.7%)
Pedagogical knowledge	7 (31.8%)	15 (68.2%)	1 (7.1%)	13 (92.9%)	3 (25.0%)	9 (75.0%)
Certification	8 (36.4%)	14 (63.6%)	5 (35.7%)	9 (64.3%)	5 (41.7%)	7 (58.3%)
Experience	14 (63.6%)	8 (36.4%)	8 (57.9%)	6 (42.9%)	8 (66.7%)	4 (33.3%)
General ability	22 (100%)	0 (0.0%)	12 (85.7%)	2 (14.3%)	10 (83.3%)	2 (16.7%)

[a] $N = 22$. [b] $N = 14$. [c] $N = 12$.

funded earlier, and these awards had more data collection activities accumulated over the years they had invested in their award. Therefore, the earlier the MSP was awarded, the more documents there were available for researchers to analyze, resulting in a larger number of instruments reported. However, when the proportions were compared for each instrument type, the three cohorts were all using instruments in similar proportions. These results indicate that, although the make-up of the three cohorts contained different types of partnerships, the types of instruments used and the teacher quality characteristics assessed were similar among the cohorts.

LIMITATIONS

Researchers acknowledge several limitations in our study. A major limitation was our exclusive use of secondary source documents to gather data about the instruments in use by these awardees. Because this was a preliminary analysis of the MSP-PE, researchers were constrained to the use of documents provided by the awardees to the funding agency through annual reports, evaluation reports, published papers, presentations, and project Web sites. This limited our data in several ways. First, awardees were not required to describe and include samples of their instruments and assessments or their psychometric properties in their reports to the funding agency. For this reason, the information on the instruments was reported voluntarily by awardees and is potentially an underrepresentation of the actual

TABLE 7
Frequency and Percentage of Instruments Used by Cohort I, II, and III Awards

Instrument	Cohort I (Awarded 2002)[a]	Cohort II (Awarded 2003)[b]	Cohort III (Awarded 2004)[c]
Written surveys and questionnaires	50 (36.2%)	41 (45.1%)	16 (30.2%)
Behavioral observations	22 (15.9%)	10 (11.0%)	8 (15.1%)
Exams	18 (13.0%)	16 (17.6%)	11 (20.8%)
Exit and other interviews	15 (10.9%)	10 (11.0%)	5 (9.4%)
Portfolios	11 (8.0%)	4 (4.4%)	5 (9.4%)
Archival records	15 (10.9%)	8 (8.8%)	7 (13.2%)
Unspecified	7 (5.1%)	2 (2.2%)	1 (1.9%)

Note. *N*s indicate the number of instruments in each of the Cohort I, II, and III awards.

[a] $N = 138$. [b] $N = 91$. [c] $N = 53$.

amount of instruments in use. In addition, researchers were not able to interact with the awardees at the time of this analysis because the MSP-PE was in its early stages and had not yet gained permission to collect data directly from awardees. This prevented researchers from interviewing awardees to determine instruments in use that may not have been identified in the secondary source documents.

Another limitation is the element of time. While researchers were gathering and analyzing data from the secondary documents, awardees were going on with their work and developing and using additional instruments to collect data on characteristics of teacher quality. For example, one RETA has designed a knowledge assessment for middle school science teachers, focusing on *Force and Motion*, *Plate Tectonics*, and *Flow of Matter and Energy in Living Systems* (Smith, 2007). This assessment has an inventory of 1,170 items covering K-12 physical science and earth science content standards. Although this assessment was not identified by any of the awardees at the time of our investigation, it may be in use by awardees at the time our study is in print. Therefore, the results reported here represent a previous point in time along the continuum of the ongoing work of these awards. Additional analyses of the instrumentation among awardees will be enhanced by the MSP-PE's ability to gather new data directly from awardees in the future.

Although our study was limited in its scope, we believe that it serves a useful purpose in providing an initial examination of the instrumentation in use among awardees in the MSP Program, thereby providing a formative assessment and impetus for comprehensive reporting on instrumentation for assessing characteristics

of teacher quality. The identification of the instruments in use by awardees in this study is also a useful first step toward determining how to design further examinations of the growth of teacher content knowledge, which was an important goal put forth in the Committee of Visitor's review of the MSP Program (NSF, 2005).

DISCUSSION

The results of our study show the instrumentation used by awardees to assess teacher quality characteristics in a national mathematics and science program. The findings illustrate teacher characteristics of most importance to awardees and the instruments used to gather data on those characteristics. Several key findings emerged from the analysis.

What the Results Reveal About the Assessment of Teacher Quality Characteristics

These results reveal that awardees in this program are engaged in the assessment of teacher quality using a variety of different types of instruments to document the growth of several teacher characteristics. Although much of the pure research in the general domain of teacher quality uses characteristics such as years of experience, general ability, and certification status as representations of teacher quality, awardees in our study were more likely to assess (a) teachers' behaviors, practices, and beliefs; (b) subject knowledge; and (c) pedagogical knowledge (85.4%, 81.3%, and 77.1% of awards, respectively). In the context of this awards program these results are not surprising. These are characteristics for which the awardees have identified specific goals for improvement as part of their work. The awards are funded based on a set of project-specific goals and plans for demonstrating and assessing progress toward those goals. It makes sense that awardees would focus assessments of teacher quality on subject matter knowledge; pedagogical knowledge; and behaviors, practices, and belief, because these are characteristics of teachers over which awardees' work may have some influence.

Exams were used most often to assess subject knowledge, and surveys and observations were used most often to assess pedagogical knowledge. The use of exams to assess subject knowledge was true for all three types of subject knowledge (mathematics, science, and MKT). The use of exams is a common and preferred method for assessing subject knowledge in academic settings, including schools and universities. Because each of these awards is a partnership among schools and universities, with discipline faculty involved in the teacher knowledge development work of the award, using exams is viewed as a practical and objective measure for this characteristic. More than half of the awards in our study used exams that were developed externally. Reasons for this may be that exams are

more available to the awardees as resources from external sources than other types of instruments. In addition, the development of exams is a complex and time-intensive process that involves a variety of psychometric processes to validate the instruments. The use of surveys or observations to assess pedagogical knowledge was consistent among many of the awards in the program. In some cases, a combination of surveys, observations, and exams was used to gather data on teacher characteristics for partnership activities. Teacher quality is a complex construct, and it was not uncommon for awardees to utilize various instruments to collect data on different teacher characteristics in the hopes that these data could be triangulated to illuminate teacher change. The use of various instruments reveals that awardees are aware of the complexity inherent in documenting teacher growth and that they are attempting to focus on that growth as it relates to teachers' participation in partnership activities.

The Quality of the Instruments

Almost every award used surveys and questionnaires, with almost one third of these developed locally by awardees. However, the awards in this analysis were not required to provide comprehensive information about the instruments in use at the time of this review, and therefore much of the information on the psychometric properties of the locally developed instruments was unknown. In contrast, 28% of the instruments were identified as externally developed, which means that the potential for these instruments to have psychometric properties is promising. An additional 37% of the instruments in use did not have their development identified, and perhaps some of these have available psychometric properties as well. Because the development of so many of the instruments was not identified, and because many were not available for direct review, researchers could not reach any general conclusions about the quality of these instruments.

In future research and development work that includes the creation and use of instruments to assess teacher quality, reporting psychometric properties of the instrumentation will be informative to researchers and educators. When conclusions are reached in any assessment of teacher quality characteristics without reporting sufficient information about the instrumentation used in the assessment, careful attention must be given to the trustworthiness of the results. Inclusion of this information in publications by the awardees will be a necessary part of the interpretation of any findings. In the case of these data, previously discussed limitations prevented researchers in our study from determining if the instruments did not have psychometric properties or if this information was simply not included in the secondary source documents because it was not required.

The limited amount of information being widely distributed on the instruments currently in use by the awardees is a drawback to others engaged in mathematics and science teacher quality work. Researchers in our study recommend that

awardees organize and expand the collection of MSP instruments available. Although one of the RETA Awards (http://www.addingvalue.org) currently lists several resources for instrumentation, and the MSP Toolbox/Materials section of the MSP Net Web site (http://hub.mspnet.org/index.cfm/join) lists some instruments, this resource could be expanded more broadly. In addition to awardees posting their instruments for shared access, the site could be a place to post standards for the selection of high-quality instruments in an effort to support awardees and enhance the quality of data gathered in the MSP Program. Standards for selecting instruments should include basic questions such as, What criteria were used to select the instruments? How do we know that this instrument is gathering evidence that will help us to determine whether or not we have reached our project's benchmarks and goals? These are good practices to adopt in evidence-based designs and beneficial when instruments are discussed and shared with the broader research community.

The Development of Instruments That Fill Needed Niches

An important idea that emerged from these findings for the general field of teacher education research is that there are a limited number of instruments available that effectively measure mathematics and science teacher quality characteristics. As NCLB set the goals for teacher accountability, and educators sought to achieve "Highly Qualified" teacher status, greater focus was placed on assessing the quality of mathematics and science teachers. National and international comparisons in mathematics and science painted a less than favorable picture of the quality of America's mathematics and science teaching force. As a result, benchmarks were set to ensure that every mathematics and science classroom would have a highly qualified teacher. A need developed for assessments of teacher characteristics that better reflected teacher quality. As part of this process, important questions have emerged. For example, What instruments are specific to measuring the quality of mathematics and science teachers? Are there measures of mathematics and science teacher quality that can be tied to student outcomes? Is it possible to develop instruments to assess the multidimensional characteristics needed to effectively teach mathematics and science?

Prior research has indicated that there are gaps in the instrumentation available to measure types of teacher knowledge. Developing and testing these instruments is time-consuming and expensive work. But there is evidence among these awards that instruments are under development and in use by awardees in this program. For example, the MKT assessment, which was not developed solely with funds from this program, is the result of ongoing research from a variety of funding sources including an NSF-MSP RETA award (Hill & Ball, 2004; Hill, Rowan, & Ball, 2005). This instrument filled a needed niche for assessing subject and teaching knowledge for mathematics at the elementary level, whereas previous assessments focused on measuring mathematics subject knowledge alone. Because the MKT

instrument is being used and tested in settings across 13 of the awards, it provides an opportunity for its developers to gather data on its use across a variety of mathematics teaching and learning environments.

Although the goals of awardees were not specifically focused on the development of new instruments, almost one fourth of the instruments identified in this analysis were reported as developed locally (69 instruments) by the awardees themselves. About 10% of these had also reported some psychometric properties at the time of this analysis. Among these instruments are assessments that have the potential to fill needed niches for collecting data on other teacher quality characteristics. These newly developed instruments appear in a number of different categories (surveys, observations, exams, interviews, and portfolios) and may be particularly useful to schools and universities because they were developed by awardees in the program and used in applied settings. New instruments that assess mathematics and science teacher quality at the end of preservice training at the university, for the purpose of hiring mathematics and science teachers for K-12 school positions, or to identify areas of needed in-service training for teachers, would benefit the field of education and the assessment of mathematics and science teacher quality.

CONCLUSION

At the beginning of this research our team posed the following question: What have researchers learned about assessing mathematics and science teacher quality? The results of our study shed some light on the answers to this question. Our findings indicate that there are a variety of instruments in use for assessing characteristics of mathematics and science teacher quality, including exams, surveys, observations, and interviews. The characteristics of mathematics and science teachers most commonly assessed among these awards included teacher behaviors, practices, and beliefs; subject knowledge; and pedagogical knowledge, which the research indicates are teacher characteristics commonly associated with student achievement outcomes. There are also a number of instruments that have been developed and are under development for assessing characteristics of mathematics and science teachers. These developing instruments may fill gaps that currently exist in instrumentation, providing researchers and educators with better ways to assess mathematics and science teacher quality.

ACKNOWLEDGMENT

This study is one in a series of substudies for the Math and Science Partnership Program Evaluation (MSP-PE) conducted for the National Science Foundation's

MSP Program. The MSP-PE is conducted under Contract No. EHR-0456995. Since 2007, Bernice Anderson, Ed.D., Senior Advisor for Evaluation, Directorate for Education and Human Resources, has served as the National Science Foundation Program Officer. The authors are Patricia S. Moyer-Packenham, Utah State University (formerly of George Mason University), Johnna J. Bolyard, West Virginia University, Anastasia Kitsantas and Hana Oh, George Mason University.

The MSP-PE is led by COSMOS Corporation in current partnership with George Mason University (GMU) and Brown University. Robert K. Yin (COSMOS) serves as Principal Investigator and Jennifer Scherer (COSMOS) serves as one of three Co-Principal Investigators. Additional Co-Principal Investigators and their collaborating institutions (including discipline departments and math centers) are Patricia Moyer-Packenham (USU, formerly GMU) and Kenneth Wong (Brown). Any opinions, findings, conclusions, and recommendations expressed in this article are those of authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Adams, P. E., & Krockover, G. H. (1997). Beginning science teacher cognition and its origin in the preservice secondary science teacher program. *Journal of Research in Science Teaching, 34,* 633–653.

American Statistical Association. (2007). *Using statistics effectively in mathematics education research*. Retrieved February 25, 2007, from http://www.amstat.org/research_grants/pdfs/ SMERReport.pdf

Andrew, M. D. (1990). Differences between graduates of 4-year and 5-year teacher preparation programs. *Journal of Teacher Education, 41*(2), 45–51.

Andrew, M. D., & Schwab, R. L. (1995). Has reform in teacher education influenced teacher performance? An outcome assessment of graduates of an eleven-university consortium. *Action in Teacher Education, 17*(3), 43–53.

Bolyard, J. J., & Moyer-Packenham, P. S. (2008/this issue). A review of the literature on mathematics and science teacher quality. *Peabody Journal of Education, 83*(4), 509–535.

Chaney, B. (1995) *Student outcomes and the professional preparation of eighth-grad teachers in science and mathematics* (National Science Foundation Rep. No. RED-9255255). Rockville, MD: Westat. (ERIC Document Reproduction Service No. ED 389530)

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Educational Policy Analysis Archives, 8*(1). Retrieved February 2005 from http://epaa.asu.edu/epaa/v8n1

Darling-Hammond, L., Chung, R., & Frelow, F. (2002). Variation in teacher preparation: How well do different pathways prepare teachers to teach? *Journal of Teacher Education, 53*, 286–302.

Darling-Hammond, L., Holtzman, D. J., Gatlin, S. J., & Heilig, J. V. (2005). *Does Teacher Preparation Matter? Evidence about Teacher Certification, Teach for America, and Teacher Effectiveness*. Retrieved August 16, 2005, from Stanford University, Center for Teacher Education and School Reform Web site: http://schoolredesign.net/srn/news/certification.html

Darling-Hammond, L., & Youngs, P. (2002). Defining "highly qualified teachers": What does "scientifically-based research" actually tell us? *Educational Researcher, 31*(9), 13–25.

Druva, C. A., & Anderson, R. D. (1983). Science teacher characteristics by teacher behavior and by student outcome: A meta-analysis of research. *Journal of Research in Science Teaching, 20*, 467–479.

Ehrenberg, R. G., & Brewer, D. J. (1994). Do school and teacher characteristics matter? Evidence from High School and Beyond. *Economics of Education Review, 13*(1), 1–17.

Ehrenberg, R. G., & Brewer, D. J. (1995). Did teachers' verbal ability and race matter in the 1960s? Coleman revisited. *Economics of Education Review, 14*(1), 1–21.

Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation, 28*, 465–498.

Ferguson, R. F., & Ladd, H. F. (1996). How and why money matters: An analysis of Alabama schools. In H. F. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (pp. 265–298). Washington, DC: Brookings Institution.

Ferguson, P., & Womack, S. T. (1993). The impact of subject matter and education coursework on teaching performance. *Journal of Teacher Education, 44,* 55–63.

Fetler, M. (1999). High school staff characteristics and mathematics test results. *Education Policy Analysis Archives, 7*(9). Retrieved December 15, 2006, from http://epaa.asu.edu/epaa/v7n9

Fraenkel, J. R., & Wallen, N. E. (1993). *How to design and evaluate research in education* (3rd ed.). New York: McGraw-Hill.

Goe, L. (2002). Legislating equity: The distribution of emergency permit teachers in California. *Education Policy Analysis Archives, 10*(42). Retrieved June 2, 2005, from http://epaa.asu.edu/v10n42/

Goldhaber, D. D., & Brewer, D. J. (1997a). Evaluating the effect of teacher degree level on educational performance. In W. J. Fowler (Ed.), *Developments in school finance, 1996* (pp. 197–210). Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Goldhaber, D. D., & Brewer, D. J. (1997b). Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *The Journal of Human Resources, 32*, 505–523. Retrieved May 25, 2005, from JSTOR database.

Goldhaber, D. D., & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis, 22*(2), 129–146.

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research, 66*, 361–396.

Grossman, P. L. (1989). Learning to teach without teacher education. *Teachers College Record, 91,* 191–207.

Grossman, P. L., & Richert, A. E. (1988). Unacknowledged knowledge growth: A re-examination of the effects of teacher education. *Teaching and Teacher Education, 4*, 53–62.

Grossman, P. L., Valencia, S. W., Evans, K., Thompson, C., Martin, S., & Place, N. (2000). Transitions into teaching: Learning to teach writing in teacher education and beyond. *Journal of Literacy Research, 32*, 631–662.

Guarino, C. M., Hamilton, L. S., Lockwood, J. R., & Rathburn, A. H. (2006). *Teacher qualifications, instructional practices, and reading and mathematics gains of kindergarteners. Research and development report* (NCES No. 2006-031). Washington, DC: National Center for Education Statistics.

Guyton, E., & Farokhi, E. (1987). Relationships among academic performance, basic skills, subject matter knowledge, and teaching skills of teacher education graduates. *Journal of Teacher Education, 38*(5), 37–42.

Hamilton, L. S., McCaffrey, D. F., Stecher, B. M., Klein, S. P., Robyn, A., & Bugliari, D. (2003). Studying large-scale reforms of instructional practice: An example from mathematics and science. *Educational Evaluation and Policy Analysis, 25*, 1–29.

Hansen, J. B., & Feldhusen, J. F. (1994). Comparison of trained and untrained teachers of gifted students. *Gifted Child Quarterly, 38,* 115–121.

Hanushek, E. (1971). Teacher characteristics and gains in student achievement: Estimation using micro data. *The American Economic Review, 61*, 280–288.

Hanushek, E. (1992). The trade-off between child quantity and quality. *The Journal of Political Economy, 100*(1), 84-117. Retrieved January 20, 2006, from JSTOR database.

Hanushek, E. (1996). A more complete picture of school resource policies. *Review of Educational Research, 66,* 397–409.

Hart, D. (1994). *Authentic assessment: A handbook for educators.* Menlo Park, CA: Addison-Wesley.

Hawk, P. P., Coble, C. R., & Swanson, M. (1985). Certification: It does matter. *Journal of Teacher Education, 36*(3), 13–15.

Hawk, P. P., & Schmidt, M. W. (1989). Teacher preparation: A comparison of Traditional and alternative programs. *Journal of Teacher Education, 40*(5), 53–58.

Hiebert, J., Gallimore, R., Garnier, H., Givvin, K. B., Hollingsworth, H., Jacobs, J., et al. (2003). *Teaching Mathematics in Seven Countries: Results from the TIMSS 1999 Video Study* (No. NCES 2003–013 Revised). Washington DC: U.S. Department of Education, National Center for Education Statistics.

Hill, H. C., & Ball, D. L. (2004). Learning mathematics for teaching: Results from California's mathematics professional development institutes. *Journal for Research in Mathematics Education, 35*(5), 330–351.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 371–406.

Houston, W. R., Marshall, F., & McDavid, T. (1993). Problems of traditionally prepared and alternatively certified first-year teachers. *Education and Urban Society, 26*(1), 78–89.

Laczko-Kerr, I., & Berliner, D. C. (2002). The effectiveness of "Teach for America" and other under-certified teachers on student academic achievement: A case of harmful public policy. *Educational Policy Analysis Archives, 10*(37). Retrieved March 4, 2005, from http://epaa.asu.edu/epaa/v10n37/

Lopez, C. L. (1998). Assessment of student learning. *Liberal Education, 84*(3), 36–43.

Lutz, F. W., & Hutton, J. B. (1989). Alternative teacher certification: Its policy implications for classroom personnel practice. *Educational Evaluation and Policy Analysis, 11*, 237–254.

Miles, M. B., & Huberman, A. M. (1984). *Qualitative data analysis*. Beverly Hills: Sage.

Miller, J. W., McKenna, M. C., & McKenna, B. A. (1998). A comparison of alternatively and traditionally prepared teachers. *Journal of Teacher Education, 49*, 165–176.

Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review, 13*(2), 125–145.

Moyer-Packenham, P. S., Bolyard, J. J., Oh, H., Kridler, P., & Salkind, G. (2006). Representations of teacher quality, quanitity, and diversity in a national mathematics and science program. *Journal of Educational Research & Policy Studies, 6*(2), 1–40.

Moyer-Packenham, P. S., Parker, J., Bolyard, J. J., Kitsantas, A., & Huie, F. (2008, May). *Examining strategies that promote teacher diversity in mathematics and science*. Research Paper Presentation, Twelfth Consultation of the International Consortium for Research in Science and Mathematics Education (ICRSME), Quito, Equador.

Monk, D. H., & King, J. A. (1994). Multilevel teacher resource effects in pupil performance in secondary mathematics and science: The case of teacher subject matter preparation. In R. G. Ehrenberg (Ed.), *Choices and consequences: Contemporary policy issues in education* (pp. 29–58). Ithaca, NY: ILR Press.

Murnane, R. J., & Phillips, B. R. (1981). What do effective teachers of inner-city children have in common? *Social Science Research, 10*, 83–100.

National Science Foundation. (2005). *Memorandum: Committee of visitors report for the math and science partnership program*. Retrieved December 20, 2006, from http://www.nsf.gov/od/oia/ activities/cov/ehr/2005/MSPcov.pdf

National Science Foundation. (2007). *MSP goals, structure and composition*. Retrieved February 21, 2007, from http://www.nsf.gov/ehr/MSP/nsf05069_3.jsp

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

Paulson, L. F., Paulson, P. R., & Meyer, C. (1991). What makes a portfolio a portfolio? *Educational Leadership, 48*(5), 60–63.

Prus, J., & Johnson, R. (1994). Assessment and testing myths and realities. *New Directions for Community Colleges, 88,* 69–83.

Rice, J. K. (2003). *Teacher quality: Understanding the effectiveness of teacher attributes.* Washington, DC: Economic Policy Institute.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417–458.

Rowan, B., Chiang, F., & Miller, R. J. (1997). Using research on employees' performance to the study effects of teachers on students' achievement. *Sociology of Education, 70*(4), 256–284.

Sandlin, R. A., Young, B. L., & Karge, B. D. (1992). Regularly and alternatively credentialed beginning teachers: Comparison and contrast of their development. *Action in Teacher Education, 14*(4), 16–23.

Schloss, P. J., & Smith, M. A. (1999). *Conducting research.* Upper Saddle River, NJ: Prentice-Hall.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14.

Smith, P. (2007). *Assessing teacher learning about science teaching*. Retrieved February 26, 2007, from http://atl.mspnet.org/

Staub, F. C., & Stern, E. (2002). The nature of teachers' pedagogical content beliefs matters for students' achievement gains: Quasi-experimental evidence from elementary mathematics.*Journal of Educational Psychology, 94,* 344–355.

Strauss, R. P., & Sawyer, E. A. (1986). Some new evidence on teacher and student competencies. *Economics of Education Review, 5,* 41–48.

Tyler, J., & Vitanova, S. (2008/this issue). Does MSP participation increase the supply of math teachers? Developing and testing an analytic model. *Peabody Journal of Education, 83*(4), 536–561.

Valli, L., & Agostinelli, A. (1993). Teaching before and after professional preparation: The story of a high school mathematics teacher. *Journal of Teacher Education, 44,* 107–114.

Weiss, I. R., & Pasley, J. D. (2004). What is high-quality instruction? *Educational Leadership, 61*(5), 24–28.

Weiss, I. R., Pasley, J. D., Smith, P. S., Banilower, E. R., & Heck, D. J. (2003). *Looking inside the classroom: A study of K-12 mathematics and science education in the United States.* Chapel Hill, NC: Horizon Research.

Wilson, S., & Floden, R. E. (2003). *Creating effective teachers: Concise Answers for Hard Questions. An Addendum to the Report "Teacher preparation research: Current knowledge, gaps, and recommendations*." Education Commission of the States. (ERIC Document Reproduction Service No. ED 476366)

Wilson, S. M., Floden, R., & Ferrini-Mundy, J. (2001). *Teacher preparation research: Current knowledge, gaps, and recommendations. A research report prepared for the U.S. Department of Education.* Seattle: Center for the Study of Teaching and Policy, University of Washington. Retrieved March 2005 from http://www.ctpweb.org/