Examining Teacher Quality

Mary M. Kennedy
Michigan State University

The press to improve K-12 education has become more and more focused on questions of teacher quality. Governments at all levels seek improvements in the quality of the nation’s teaching force. The No Child Left Behind Act, the largest piece of federal education legislation, now includes very specific requirements about teacher qualifications. But as demands for increasing the quality of teachers become even more shrill, so do arguments about what quality actually is. Advocates on each side of this debate have formed organizations to pursue their points. We now have a National Commission on Teaching and America’s Future that advocates strong professional education for all teachers and a National Center on Teacher Quality that advocates a virtual elimination of professional education in favor of a stronger liberal arts education. Both are advocacy centers, not research centers.

The purpose of this research brief is to develop the concept of teacher quality in a way that can facilitate better research on it. The paper has four sections. The first section defines and describes four distinct aspects of teacher quality. The second and third sections discuss the relationships among them and provide examples of research that has examined relationships among them. The final section offers some notes on research methods in this area.

DEFINING TEACHER QUALITY

The term teacher quality has been used to refer to many different aspects of teaching. These are the four main aspects that have been examined in recent years.

- Qualifications. Qualifications include such things as what subject the teacher majored in, whether the teacher has a bachelor’s degree or a master’s degree, whether the teacher has passed required licensure tests, and so forth. Every state in the nation requires a specific set of qualifications that must be met before a teacher can be certified. But these requirements differ substantially from state to state. Exacerbating the problem of defining qualifications is the new federal legislation, No Child Left Behind, which requires that school districts provide a highly qualified teacher in every classroom.

- Effectiveness. Effectiveness refers to the teachers’ ability to produce student learning. Usually gains in student test scores that can be attributed to the teacher are assumed to provide evidence of effectiveness. For example, the state of Tennessee has a database that enables researchers to track student progress from grade to grade. Researchers there have used that data to show that individual teachers are remarkably varied in how much their students gain in any given year (Sanders & Horn, 1994). However, they do not say what kind of qualifications lead to effectiveness, nor what kind of teaching practices lead to effectiveness.
Quality of practice. When people evaluate the quality of teachers' classroom practices, they attend to such things as teachers' ability to engage students intellectually with content, their ability to increase student interest in academic subjects, their ability to include all students in the conversation, and so forth. This is the aspect of quality that most parents refer to when they say their children's teachers are "good" or "not so good." This is the aspect of quality that the National Council of Teachers of Mathematics (NCTM) aimed to improve when it introduced teaching standards (NCTM, 1991). However, research evidence regarding the relationship between particular kinds of practices and effectiveness is uneven at best.

Orientation. "Orientation" refers to how teachers understand and think about their work. It includes their beliefs and values, their goals, and their interpretations of classroom events as they unfold. This aspect of teacher quality is the least visible of the four but is frequently assumed to be an important contributor to both classroom practice and effectiveness. Although it is less often considered in policy conversations and in local school district administrative decisions, it has recently received a great deal of attention from researchers.

Debates about teacher quality are fueled by two interrelated problems. One is that these different aspects of teacher quality are confused with one another and are often used interchangeably, as if they meant the same thing or were so closely related that one can adequately represent another. All four of these aspects of quality have been used as "outcomes" in research on teacher quality without considering how they might differ from one another and which outcomes are closest to an "ultimate" outcome of student learning. While Wenglinski (2000) used Praxis scores (indicating qualifications) as an outcome in his study of teacher education, Coote, Shealy, and Arvoid (1998) used beliefs (indicating orientations) for their study of teacher change, Stigler and Hiebert (1999) used lesson structures (indicating quality of practice) as the outcome in their cross-national comparison, and Sanders and Rivers (1996) used student achievement test scores (indicating effectiveness) as an outcome in their study. I have examined this problem of variations in outcomes in a paper that contrasts outcomes according to how closely they approximate an ultimate outcome of student learning (Kennedy, 1999). I argue that virtually all measured outcomes are approximations to our goal of student learning, including test scores. However, I also argue that some outcomes, such as classroom practices and test scores, more closely approximate the ultimate outcome than others, such as teachers' orientations, do.

The second problem with research in this area is that, even when researchers try to distinguish these different aspects of quality, and try to understand their relationships better, these relationships are very difficult to ascertain empirically because we cannot experimentally manipulate any aspect of teacher quality in order to see how such manipulations affect other aspects of quality. Researchers cannot manipulate qualifications by randomly assigning teachers to institutions, programs, majors, grade point averages or test scores, they cannot randomly assign teachers to different belief systems, to different classroom practices, or to different outcomes. Virtually every study addressing questions of teacher quality must use naturally occurring populations, people who developed their own qualifications, orientations and practices. Therefore any researcher approaching this issue must address a variety of potentially confounding variables.

1 The closest researchers can come to random assignment of any aspect of quality is when they randomly assign teachers to professional development programs that are designed to influence classroom practice. Even in this situation, they cannot ensure that every "graduate" will indeed engage in the researchers' definition of quality practice.
HYPOTHESES ABOUT TEACHER QUALITY

Because these four aspects of teacher quality have not been carefully distinguished from one another in the literature, there has been very little effort toward understanding how changes in one aspect of quality might lead to changes in another. Still, studies that do look at relations among aspects of quality seem to assume that the pattern of influence is as shown below.

Although this hypothesized pattern of relationships has not been explicitly articulated in the literature, it is implied in the patterns of studies that are done. For example many longitudinal studies of preservice teacher education focus on orientations, trying to document the orientations candidates have when they arrive in their teacher education programs and trying to document whether and in what ways their orientations change while they participate in these programs. Other studies look at the relationship between qualifications and effectiveness, or the relationship between the quality of practice and effectiveness. All studies that examine the relationship between two aspects of teacher quality assume that causation runs from the left-most aspect to the right-most aspect, not the reverse. In the next section, I review examples of these studies.

RESEARCH ON TEACHER QUALITY

Researchers have examined virtually every combination of aspects of quality, more combinations than I have depicted with arrows in the diagram above. Below I briefly describe a sample study or line of work that illustrates each combination.

Orientations

Descriptions of orientations. Orientations are difficult to describe and to capture empirically. Most research on orientations relies on qualitative methods to uncover teachers' tacit beliefs and values about teaching, mathematics, and student learning. An example of such work is Ball's (1988a, 1988b, 1990a, 1990b, Ball & Feiman-Nemser, 1988) early work on preservice mathematics teachers' beliefs. Using both interviews and questionnaires, Ball examined how teachers think about specific mathematical ideas and about how to teach these ideas to students. Her work highlights the importance of early learning in the formation of beliefs and raises issues about the role and potential of teacher education to alter these beliefs.

Influence of orientations on qualifications. Though my chart suggests that orientations precede and therefore influence teachers' qualifications, most researchers who examine this relationship are interested in bi-directional influence. That is, researchers know that teachers' orientations influence their response to their teacher education programs. However, the aim of the research is not usually to document this influence but instead to see whether the program can alter students' orientations. An example of a study that examines the relationship between
orientations and qualifications is Cooney, et al.'s (1998) longitudinal study of four preservice teacher candidates. As teacher educators, these authors wanted to persuade their students that mathematics was constructed and did not appear in nature fully formed. As researchers, they were interested in the beliefs their students had when they arrived and in the extent to which these beliefs changed over time. The authors describe three phases in their students' learning process: (a) their initial understanding of math and teaching; (b) their search for affirmation of their initial understandings, and (c) their anticipation of student teaching. The second stage is especially difficult for teacher educators, for it makes students resistant to program messages. Some students worked harder at affirmation and never changed, others tried to affirm but eventually yielded to the program point of view.

Influence of orientations on quality of practice. Researchers care about orientations mainly because they believe orientations have a great bearing on teaching practice and that orientations are responsible for whether, and to what extent, teachers respond appropriately to ideas such as those laid out in the NCTM Standards documents. An example of a study that examines the relationship between orientations and the quality of classroom practice is a study done by Peterson, Fennema, Carpenter, and Loe (1989). These researchers developed a measure of what they called a "cognitive-based perspective." and then sought to see whether it was related to teachers' classroom practices. The authors found a positive relationship between the presence of a "cognitive-based perspective" and student scores on the problem-solving portion of an achievement test. They also found that there was no relationship between teachers' cognitive orientation and the number facts portion of the achievement test. Their measure of orientation focused on four specific beliefs: that students construct their own mathematical knowledge; that mathematics instruction should be organized to facilitate children's construction of knowledge; that children's development of mathematical ideas should provide the basis for sequencing topics for instruction; and that mathematical skills should be taught in relation to understanding and problem solving.

Qualifications

Descriptions of qualifications. An example of research that describes teacher qualifications is the line of work being carried out by Ingersoll (1996, 1999, 2001, 2002). Ingersoll is interested in the frequency with which teachers are assigned to teach classes for which they are not qualified. He varies his definition of qualifications from one analysis to the next, sometimes focusing on subject matter major or minor and sometimes on whether the teacher is certified to teach in a given subject. For instance, in his 1999 paper, he describes the number of teachers who are teaching out of their field and also the number of students who are being taught by teachers who are out of their field. In this analysis, he found that 33% of mathematics teachers were teaching outside of their field and that 27% of students were taught by teachers who were teaching outside their field.

Influence of qualifications on other qualifications. An example of a study that looks at the contribution of early qualifications to later qualifications is Wenglinsky's (2000) study of predictors of Praxis scores. Wenglinsky regressed numerous predictors against Praxis scores, including students' own earlier test scores, students' SES backgrounds, and institutional and program characteristics. He found, for instance, that programs with relatively more diverse faculty produced students who obtained higher Praxis scores, that private institutions outperformed public institutions, and that universities outperformed colleges.
Influence of qualifications on quality of practice. An example of a study examining the relationship of qualifications to the quality of teaching practice is Claridge's (1990) comparison of beginning teachers. This researcher was particularly interested in differences among beginning teachers who appeared to be relatively more and less accomplished, but also reported that those who had fewer courses in pedagogy were less able to keep students on track, to provide feedback to them or to assess their learning. Valli and Agostinelli (2000) conducted a study using a very different approach. They described a teacher who was teaching without a teaching license and then decided to return to college and obtain one. They compared his practice before and after he obtained his license.

Influence of qualifications on effectiveness. A very large body of work exists examining the relationship between qualifications and effectiveness. The large number of studies in this area is due in part to the availability of large-scale data bases such as the data provided by the National Educational Longitudinal Study (NELS). The NELS data base includes information on both teacher qualifications and student achievement. Examples of research of this type are Monk's (1994; Monk & King, 1994) studies of how the secondary teachers' college course taking contributes to achievement of the teachers' students. Instead of looking at what teachers majored or minored in their subjects, Monk looked at the number of credit hours they had taken in mathematics or in methods of teaching mathematics. Among other things, Monk found that teachers' course credits yielded diminishing returns on student achievement. The relationship grows stronger as the number of courses increases from none to five courses and then diminishes so that additional courses do not particularly add to future student achievement. He also found that teachers' qualifications mattered more when they taught advanced courses than when they taught remedial courses.

Quality of Practice

There have been studies examining the relationship between classroom practice and student outcomes. However, many of these studies do not provide a definition of quality of practice, but instead use student outcomes to determine which practices are effective. This was the strategy of the process-product line of research, in which researchers documented numerous specific teaching practices and then correlated them all with student achievement in an effort to determine which were most highly correlated. This research strategy conflates these two aspects of teaching quality, so that quality of practice = whatever works. Other researchers, however, have defined quality of practice and then tested their definitions by examining their relationships to student outcomes. For example, Saxe, Gearhart and Seltzer (1999) defined quality of practice as practice that is aligned with reform principles. They created "reform" classrooms by first recruiting teachers whose textbooks were aligned with reform ideas, and then providing these teachers with professional development. In their observations, they focused on whole-group discussions and documented the extent to which classroom practices elicited and built on students' thinking, and the extent to which conceptual issues were addressed in discussions of how to solve problems. They found that their conception of quality of practice made a difference in students' problem-solving ability, but mainly for students who were better prepared to begin with, and that it made no difference to students' computational abilities.
Effectiveness

Probably the most widely cited research that focuses on effectiveness is William Sanders' work (Sanders & Horn, 1994, 1995, 1998; Sanders & Rivers, 1996). Unlike other researchers, who frequently try to see which other aspects of quality contribute to effectiveness, Sanders and his colleagues measured no aspect of teacher quality except student achievement, and simply assure that their data measures teacher quality.

SOME NOTES ON METHOD

Several observations can be made about how the aspects of teacher quality discussed in the brief are defined, measured, and interrelated. In the following paragraphs, I pose several methodological problems that need to be dealt with in designing future research.

Problems Associated with Attributing Cause

The problem of self-selection. Because random assignment is difficult, virtually all studies of influences in this domain are susceptible to confounding variables and alternative explanations. For instance, Sanders' program of research on teacher effectiveness assumes that differences among classrooms in student achievement are caused by teachers, and can therefore be labeled as "teacher effects." Since the researchers have access to prior student achievement data, they can partial out classroom differences in prior achievement. However they cannot partial out differences in classroom personalities, nor the effect of one particularly resistant student on the teachers' ability to teach the rest of the class.

The problem of backward inference. There is a pervasive problem in the research literature in meeting the first two rules of causal inference, which are that the cause must be distinct from the effect and that the cause must precede the effect. These matter especially in studies of the relationship between orientations and quality of practice, because many such studies infer teachers' orientations from practice itself. For example, when Schoenfeld (1999) developed his model of teacher reasoning, he inferred teachers' orientations from their practices, arguing that in order to engage in a particular practice, a teacher would have to believe that ______. Such a strategy may be sufficient for purposes of model development, but a test of the influence of orientations on quality of practice could only occur if orientations were measured independently from practice, as they were in the Peterson et al (1989) study, and should also be measured before practice is observed.

The problem of open searches. By an "open search" study, I mean a study that stipulates an outcome (usually student achievement) and then uses multiple regression to test numerous possible influences on that outcome. Process-product research studies represent this approach. In these studies, researchers observe teaching practices, often documenting numerous specific practices, and then test the full range of potentially relevant practices to see which ones actually are associated with the outcome of interest. There are two problems with this approach. One is that the variables ultimately "discovered" may represent nothing more than chance. In a study of this sort, the greater the number of variables tested, the greater the likelihood that something will appear to have a relationship which wouldn't appear in a different sample or in the population as a whole. The second problem is that the variables that can be "discovered" in this way depend heavily on what the researchers measure. This makes sense if the researchers have a theory that guides their choice of measures, but also is a problem when researchers choices depend on what a
Exemining Teacher Quality

possible to measure, what is convenient to measure, or what ideas the researcher happened to come up with.

Problems Associated with Measurement

The problem of relying on extant databases. Many studies in this area rely on extant databases to test questions about teacher quality, even if these databases were not originally intended for this purpose. One result of this reliance is that researchers focus on qualifications that are not well-defined and are not necessarily meaningful. This occurs particularly with the variable “certification,” which tends to be included in data bases because it is easy to obtain. However, the criteria for what makes a teacher “certified” vary both across states and across time, so that its meaning as a variable lacks precision. In correlational research, if a measure is not reliable itself, it will not correlate highly with other measures, so we are not really gaining valid knowledge from such correlations.

The problem of qualitative data. Most of the research on teachers’ orientations is qualitative and relies on observations of practice, structured and unstructured interviews, and occasionally task protocols. These studies are highly interpretive and the particular orientations examined vary from one study to the next. They help us understand the nature and nuances of orientations but cannot help us establish their role in determining either the quality of teachers’ practices or the effectiveness of their practices. For this task, we need to move to larger-scale studies and clearly defined and measurable constructs. That is, we need to move from a general sense that orientations matter to specific hypotheses about specific orientations that can be clearly defined and reliably measured.

Problems Associated with Ecological Validity

The problem of valid intervals. Teachers normally teach students for standard units of time—usually an academic year, but occasionally one semester. When teacher qualifications are studied in relationship to student effectiveness, it is important that the time interval used to estimate changes in student achievement match the time interval of instruction. In many studies, the interval between the “before” and “after” test points is not coterminal with the interval of teaching. For instance, several researchers (e.g., Adams & Singh, 1998, Goldhaber & Brewer, 199, Rowan, Chang, & Miller, 1997) have relied on the NELS:88 to examine relationships between teachers’ qualifications and their effectiveness. However, NELS:88 measures student achievement only in grades 8, 10, and 12, so any relationships between achievement gains and teacher qualifications are using a two-year achievement gain interval to study teachers who taught the students for only one of those two years.

The problem of valid units of analysis. In part because researchers rely on existing data bases, they often study school or district averages rather than individual teachers. For instance, Strauss and Sawyer (1986) examined the relationship of teachers’ NTE scores to student pass rates on a high school competency test in South Carolina, using districts in South Carolina as his units of study. That is, he related district average NTE scores with district average student outcomes. From his analysis, he concluded that teachers’ test scores were contributing to student outcomes. However, districts represent relatively stable and self-sustaining socio-demographic communities. That is, one represents an agricultural community, another an industrial community, and a third a resort community. Each community is self-sustaining in that parents transfer their values and beliefs to their children, decide what level of support to offer their

79
schools, and elect their own school boards. These school boards, in turn, hire administrators and teachers. Moreover, as teachers search for jobs, they seek out positions in districts that match their own interests and values. Between teachers' job preferences and districts' cultural preferences, teachers get hired who will "fit it" to their communities. Causal relations are thus bi-directional, as teachers, once hired, perpetuate the districts' educational aims. So while it is possible to say that the teachers' average NTE scores predict student achievement, it would be equally possible to say that the districts' average student achievement predicts the NTE scores of the teachers that districts hire. The best way to avoid the problem of community to locate research within the classroom, looking only at teachers and their immediate students.

The problem of treatment validity. A few researchers have tried to examine questions about teacher quality by experimentally manipulating teaching tasks. For example, for his comparison of certified teachers with college students, Popham (1971) asked teachers to teach a unit that was developed by the researchers, and to teach it to a group of students who were not their regular students. In his effort to control confounding variables, Popham introduced new problems by creating an artificial situation in which teachers and students did not know one another, the unit was not a part of the normal curriculum, and the teachers had no flexibility in how they implemented the curriculum. Somewhat more recently, Berliner et al. (1988) also tried to ask teachers to perform artificial teaching tasks and found that, although novice teachers were willing to comply with this request, expert teachers bristled at the notion of being removed from the very situation in which their expertise rested.

The problem of teacher representativeness. One final aspect of ecological validity that deserves mention is the question of the representativeness of samples of teachers. This is a concern especially in case studies, where teachers tend to be selected because they demonstrate a specific phenomenon, not because they are representative or typical. And it is a concern especially in self-studies, in which a teacher is studying his or her own practice, as is the case with teachers such as Ball (1995) and Lampert (2001). In each of these situations, the teachers are intentionally not representative of the larger population of teachers, and while their findings can have great value to us, we need to be careful about generalizing them to broader populations.

REFERENCES


