Improving Measures of “Highly Qualified” Teaching: The Weaknesses of the NCLB Paradigm and the Importance of Multiple Methods of Evaluation

Katherine Hinkle
Education Law and Policy
Loyola University Chicago School of Law
I. Introduction

In 2001, the “No Child Left Behind Act” 1 (NCLB) created a standardized definition of a “highly qualified teacher.” NCLB required that teachers must have a bachelor's degree, full state certification, and prove that they are competent in each subject they teach to be deemed highly qualified. 2 The law defined competency as proof that a teacher knows the subject they are teaching, which can be established by a major in the subject they teach, credits equivalent to a major in the subject, passage of a state-developed test, HOUSSE, an advanced certification from the state, or a graduate degree. The law focused federal funding on the training and recruitment of these highly quality teachers and required states to demonstrate that there was a highly qualified teacher in every classroom by the 2005-2006 school year.

NCLB gave states a great deal of flexibility in defining these broad requirements more specifically and developing their own teacher tests and certification requirements to reflect state education standards. However, all state systems have one basic principle in common; each state deems teachers “highly qualified” purely through assessing their credentials. Yet, a growing body of research indicates that credentials alone do not actually reflect teacher effectiveness in the classroom3. Measuring teacher characteristics alone is not enough to determine who is highly qualified. Through a discussion of several measures of teacher quality, this paper will demonstrate that the credentials-only paradigm of “highly qualified” established in NCLB needs to change. I will argue that the best way to determine teacher quality is through a system of multiple measures. Adding tailored and sound teacher observations and properly used value

3 For the purposes of this paper, highly qualified will be defined as highly effective. Effectiveness is traditionally defined in terms of mean student achievement gains.
added measures to the most effective measures already established by NCLB would retain the few strengths of the existing paradigm while significantly improving upon its faults.

II. The Weaknesses of the NCLB Paradigm

The “highly qualified” requirement of NCLB is based on the principal that measuring teacher qualities, in the form of educational background and certification, is an adequate means of predicting student achievement outcomes. NCLB allows states the flexibility to define their certification processes and align their teacher testing with state educational goals and frameworks, but all state determinations of “highly qualified” teaching must be based on teacher testing and credentials. However, research has shown that this model is ineffective.

States have traditionally required some variety of certification for teachers, but as a result of NCLB, certification has become an even more rigorous process for the states and their teachers. For example, Illinois requires individuals to complete a state-approved certification program, a basic skills test, a content area test, and the Assessment of Professional Teaching test to achieve certification.\(^4\) NCLB assumes that by requiring teachers to meet high certification requirements set by each state, only the most qualified and well-educated individuals will be in the classroom. This assumption is founded on the idea that qualifications and education levels are always valid measures of teacher effectiveness. However, a growing body of research has shown that full certification has little to no impact on teacher quality and effective classroom instruction. In a study which analyzed teacher effectiveness through the use of specific types of successful math teaching techniques, Thomas Smith, Laura Desimone, and Koji Ueno concluded that “the certification status of current teachers does not appear to be an effective proxy for

preparedness to teach mathematics content or increased emphasis on conceptual learning and use of conceptual teaching strategies.” Certification is not a clear indicator of a teacher’s use of effective strategies in the classroom. They suggest that although uncertified teachers are more likely to fall back on less successful teaching techniques, certified teachers are not significantly more likely to use the best techniques than their uncertified counterparts. Therefore, having a certified teacher is not a guarantee of the best, most effective classroom practices.

A study by George Palardy and Russell Rumberger also supports these conclusions. In their study of student achievement in math and reading, they found that “although full certification had the largest effect size on reading gains of any teacher measure examined in the study, it is not associated with gains in math achievement.” The study further theorizes that these gains in reading may also be the result of a type of instruction practice stressed in teacher preparation programs. Therefore, relying primarily on certification as an indicator of teacher effectiveness is not a sound practice. Palardy and Rumberger assert that although it is wrong to dismiss the importance of teacher training and certification entirely, “the highly qualified teacher provision of the NCLB act is insufficient for ensuring that classrooms are led by highly effective teachers in first grade.” Although there may be some benefit to having a fully certified teacher leading a classroom, it is, at best, a marginal benefit to student achievement.

---


Id.


Id.

Id.
Furthermore, state teacher examinations, the backbone of the certification system, are only weakly aligned with state standards for student learning and have little value in predicting teacher quality. Most states, due to budget constraints, use teacher tests that are designed without input from individual states by national testing companies.\textsuperscript{10} The content of these tests is not closely aligned with state standards for teaching and learning, and therefore passage of the test does not reflect a teacher’s competency in the subject matter as the state defines it. Additionally, the tests themselves, whether created specifically for the state or an “off the shelf” test, only assess the most basic knowledge. Pamela Harrell asserts that the tests’ lack of rigor is a serious fault that undermines the concept of establishing competency through standardized assessment.\textsuperscript{11} Even though NCLB mandates that every teacher have a bachelor degree, most tests assess only high school level content and many science tests primarily assess content that would have been learned in middle school. For a individual teaching high school physics, the state test would have little predictive power of her competency with the material. Furthermore, the low bar for passage means that a teacher does not even have to demonstrate mastery on the low-level material on the test.\textsuperscript{12} Therefore the test themselves are not an adequate indicator of teacher quality.

NCLB also allows teachers to demonstrate competency in a subject matter through a major or advanced degree in the subject they will be teaching. However, this is also an ineffective measure of teacher quality. Most states, including Illinois, will allow a major in a field to substitute for a state content test. With the indicated weaknesses in state testing regimes, requiring that students major in their content area seems like an excellent way to ensure teacher

\textsuperscript{11} Id.
\textsuperscript{12} Id.
quality. Smith, Desimone, and Ueno’s study asserts that teachers with a degree in their content area did feel more prepared to teach and did use more effective techniques. However, they also found that teachers with a minor, major, and even a graduate degree in math were equally as likely to be prepared to teach content and to use effective teaching strategies. Although NCLB only allows a major in the content area to demonstrate competency, research shows this measure is ineffective at defining effective teaching. Smith, Desimone, and Ueno argue that “classifying teachers with bachelor’s or master’s degrees in the field taught as highly qualified and teachers with minors as less than highly qualified may be too rigid a policy.” A wide variety of teacher educational backgrounds can benefit student achievement. Furthermore, the study established that the content of the degree program has an even more important impact on teacher effectiveness than the level of the degree. Focusing simply on the title of the degree, instead of the content and methods learned is inadequate. This strict requirement limits the number of candidates eligible for highly qualified teacher status and does not reflect the research. Expanding the requirement to a minor or equivalent coursework would give states the benefit of more eligible teachers while still ensuring students have a teacher who is prepared and effective. Therefore, the requirement that teachers have a bachelor degree in their content area is an inadequate and overly narrow means of assessing teacher quality.

Although there is an increase in student learning when there is a certified teacher or a teacher with a degree in the content area, these gains are very small and not necessarily linked to the credentials themselves. The benefits to students from having a certified teacher may only be

---

14 Id.
15 Id.
16 Id.
a reflection of the types of classroom practices these individuals regularly use. Therefore, looking at classroom practices themselves would be a far better measure of teacher quality than requiring certain credentials.

**III. A New Evaluative Paradigm**

It is clear from the research that the measures of teacher quality established through NCLB are ineffective and inadequate. Although certain measures, like a degree in the content area, have a statistically significant impact on student achievement, other requirements have no impact and may even serve to deter people from entering the field of teaching. Although research demonstrates that credentials have some limited benefit in indicating teacher quality, they are far from enough to make sure that the goal of having a “highly qualified” teacher in every classroom is achieved. By changing and expanding the measures used to define “highly qualified,” states can increase access to the profession by cutting down on limiting requirements, such as a major in the field, while actually employing indicators which accurately measure teacher effectiveness in the classroom. By combining the best existing measures with new, properly tailored measures of teacher outputs, states can more successfully determine who is a “highly qualified teacher.” However, no measure is perfect and even each new evaluation technique has inherent weaknesses which make it ill-suited to function alone as an absolute measure of quality. Only by combining measures will a clear picture of teacher quality emerge. Therefore, each discussion of a suggested evaluative method will contain an indication of how states can best avoid the traditional pitfalls of these measures when implementing them.
A. Principal or Supervisor Observational Evaluations

Most districts employ some form of observational evaluation for teachers. However, the variety of ways in which this technique is employed is problematic. Some evaluative schemes are very good at provide districts with valuable information about their teachers. Other schemes are the product of untrained or inexperienced evaluators who employ biases or political decisions in their assessment of teacher quality. However, when done correctly, principal evaluations can be an excellent measure of teacher effectiveness.

Currently, nearly 99% of the tenured public school teachers working in large districts receive satisfactory or better evaluations, which are usually based on a biennial visit by one administrator. Yet, it is unlikely that 99% of teachers are actually meeting or exceeding state standards as evidenced by poor educational progress and a large achievement gap. The labeling of the majority of teachers as proficient through principal evaluations may come from an inherent weakness in the evaluation system. In their study on the problems with principal evaluations, Brian Jacob and Lars Lefgren concluded, “principals are generally effective at identifying the very best and worst teachers. On average, however, they are not able to distinguish teachers in the middle of the achievement distribution.” They argued that principals are generally unable to make fine distinctions in identifying effective teachers because they are only asked to make evaluations to reward the best teachers or remove the very worst. Without the practice necessary to perform detailed evaluations or an effective method of doing so, principals will continue to be only a minimally effective measure of high quality teaching. With

19 Id.
this data in mind, it is clear why principal evaluations have generally not been viewed as an effective measure of teacher quality.

However, a growing body of research demonstrates that when employed properly, principal evaluations can be one of the best measures of teacher effectiveness. By choosing the appropriate method of gathering observational data, principals can make evaluations that are reliable reflections of teacher quality and valid measures of potential student outcomes. A successful evaluation system is one where clear goals for effective teaching are defined and observers demonstrate accuracy in measuring the implementation of these goals. First, principal evaluations tend to be weak measures of effective teaching because it is unclear what is actually being evaluated. While principals may be able to recognize truly outstanding teachers or egregious violations of best practices, without a clear set of school or department-wide learning and teaching objectives, evaluators do not know what they are looking for and will continue to make broad, subjective judgments. Thus, the first step to identifying high quality teachers is to clearly define what high quality teaching actually looks like. Based on its research, The Measures for Effective Teaching Project suggests that “teachers and supervisors need a common vision of effective instruction to work towards” for any evaluation to be successful.20 States may choose to use a pre-packaged system with demonstrated accuracy, such as the Framework for Teaching, or may rely on district standards to create their own method of evaluation. Regardless of the tool used or a particular state’s standards, if teachers and principals are in agreement about student learning goals, principal observations become much better measures of teacher quality.

20 Thomas J. Kane et al., Gathering Feedback for Teaching: Combining High-Quality Observations with Study Surveys and Achievement Gains, Measures of Effective Teaching Project Policy and Practice Brief, 1, 7 (2012). Available at  http://www.metproject.org/reports.php
After a common vision has been established, evaluators then need to be able to accurately measure a teacher’s ability to implement these standards. The MET project has demonstrated that “inaccurate classroom observations lead to mistrust and poor decisions” about who is highly qualified. The Project’s research suggests that evaluators must undergo significant training to properly implement district evaluation standards, minimize evaluator bias, and identify subtle differences in teachers’ ability to implement the goals. Only when evaluators or principals are properly trained to identify good teaching can observations shift from subjective judgments to valuable measures.

When schools use proper evaluation systems, teachers with higher observation scores have students who learn more. A properly created and implemented evaluation system is able to accurately predict which teachers will have a positive influence on student achievement. Furthermore, the MET study suggests that a combined measure of effective teaching which includes observational evaluations is significantly better at predicting student achievement on state tests than a teacher’s experience or credentials. When used properly and in concert with other appropriate methods of evaluating teacher effectiveness, principal observations are much better than the credential-only measures suggested by NCLB.

It is important to note that even in the most well-designed evaluation system, there may still be room for favoritism and social pressures. Although any quality observation system would train evaluators in bias-reducing methods, human error is still possible. Jacob and

\[\text{Id. at 14.}\]
\[\text{Id. at 23.}\]
\[\text{Id. at 26.} \quad \text{(using a combined evaluation measure of the Framework for Teaching observational evaluation, student surveys on teacher performance, and value-added assessments from state tests. Because I have chosen to include two of these three measures in my suggestions for multiple measures of evaluation, I feel these outcomes accurately reflect what they might be under my proposed system.)}\]
Lefgren point out that even when the weaknesses in an evaluation system’s ability to identify teacher quality are corrected for, these personality issues may still play a role.\textsuperscript{25} Therefore, if evaluations are used in hiring and firing or other high-stakes decisions, it is important that other measures of teacher quality be used to lessen the impact of these factors.

\textbf{B. Value-Added Measures}

Value-Added assessments of teachers are an increasingly popular method for identifying highly qualified teachers. Value-added assessments longitudinally measure the learning of each student to identify a school’s or a specific teacher’s impact on that individual’s learning. The data from prior achievement is used to project future gains. When a student performs above those gains, the gain is attributed to the school and teacher. Likewise, a downward deviation from the projected path is also attributed to the learning environment. Proponents argue that these tests improve on the “point-in-time” snapshots of student achievement required under NCLB by better accounting for previous school quality and non-school factors.\textsuperscript{26} The student acts as his own control and is not compared to other students, only to his own previous accomplishments on state tests. Supported by the federal government’s “Race to the Top” program, states are investing billions of dollars to create and implement these measures.\textsuperscript{27} However, there is still a great deal of debate about their validity and effectiveness. Some argue that evaluations of teacher effectiveness should never include data from state tests, while others argue that the current model is valid. Unfortunately, in today’s political climate, it is unlikely proponents of the first view will ever be successful. Evaluations of teacher quality based on

\textsuperscript{27} \textit{Id.}
standardized student tests are a permanent fixture in American education, and the best way to mitigate their weaknesses is to choose a method of employing the data that best corrects for potential problems. Despite considerable doubts as to their validity as an indicator of teacher quality, value-added measures are still the best way to satisfy the drive for objectively based decisions on teacher quality. However, to be successful, value-added measures must be combined with other measures of teacher effectiveness to give the most accurate picture of who is a “highly qualified” teacher.

The primary criticism of value-added measures is that they do not actually correct for many of the weaknesses of traditional test-based evaluations. First, many question whether student achievement can even be properly measured through standardized testing. Inherent cultural biases in tests and their inability to really measure higher-order thinking have led many to question what student gains these tests are actually measuring. However, if we accept the premise of testing as a valid measure of achievement, value-added measures still have a significant fault. Thomas Misco argues that these measures should not be used as the sole measure of quality in making personnel decisions because of the difficulty in isolating teacher effect in any increase or decrease in student achievement. He argues that there is “extreme difficulty involved in probabilistically connecting student achievement gains and regressions to teachers and isolating their effects on students.” Additional factors could mean that an increase or decrease in predicted student performance is not the result of a teacher effect. For example, research shows that extraordinarily good teaching has effects that last long after the student has

29 Id.
changed teachers.\textsuperscript{30} The unexpected increase in performance may be the result of the student’s previous teacher. Furthermore, students are not perfect controls. Outside variables such as home life or student health do not stay constant from year to year. A decrease in expected performance may not take into account a student’s new homelessness or significant health problem which affects his ability to learn. Therefore, it can be extremely difficult to specifically link student performance to teacher effects. Without additional evidence to demonstrate quality teaching, these measures alone are not adequate measures of teacher effectiveness.

Despite these concerns, when combined with other measures of teacher quality, value-added measures are good indicators of teacher quality. In fact, Douglas Harris and Tim Sass argue that, “when value-added measures are constructed from multiple years of test score data, past value-added does a much better job at predicting teacher performance than do principal evaluations.”\textsuperscript{31} Palardy and Rumberger support this assertion and also argue that, “these models are particularly suitable for identifying effectiveness of teachers in raising achievement when the curriculum is uniform across all classrooms.”\textsuperscript{32} Both studies agree that value-added assessments are strongest when multiple measures are combined. Harris and Sass explain that even when there are multiple years of data for value-added assessments, “principal ratings still add information that significantly improves the ability to predict future teacher performance in reading.”\textsuperscript{33} Palardy and Rumberger agree, and argue that value-added assessments are really only valuable measures of teacher quality when they can be used to identify ineffective teachers

\textsuperscript{30} Id.
and qualitative evaluations are used to identify the fine details of the teacher’s weaknesses.\textsuperscript{34} Although these measures have their inherent weaknesses, they are still valuable tools when added to other valid measures of teacher effectiveness. If politics demand quantitative analysis of highly qualified teaching, then studies support using value-added measures as the preferred measure of providing this aspect of teacher evaluation.

**IV. Conclusions.**

No measure taken alone can properly identify “highly qualified” teachers. However, the current credentials-only measures are inadequate. Although NCLB attempts to use multiple types of credentials to establish quality, only one, a degree in the content area taught, has an impact on student performance. Minimum teacher credentials may be a cost-effective way of putting some requirements on teacher quality, but they do little to achieve NCLB’s goal of putting a “highly qualified” teacher in every classroom. Achieving NCLB’s goal will require states to completely redefine how they will evaluate teacher quality. Douglas Harris argues that, “although coming at some financial cost, collecting and combining measures reduces bias and imprecision and provides more useful information to guide teachers and school leaders.”\textsuperscript{35} Using teacher credentials as a single method of evaluation is not a valid measure of teacher effectiveness. As Thomas Misco points out, “single-point appraisals are intrinsically fallible…any assessment needs to be responsive to the needs and goals of the individual


communities and be part of a larger grouping of multiple measures, including qualitatively oriented assessments that account for gains in creativity, exploration, and curiosity.\textsuperscript{36}

Neither value-added assessments nor principal evaluations are a perfect measure of teacher quality. However, when combined they compensate for each other’s weaknesses. The objective data of the value-added assessments will combat favoritism and subjectivism in the principal evaluations. Similarly, the principal evaluations can help pinpoint teacher effects in the value-added assessments and provide a clearer picture of the student gains that are not testable. Each state must evaluate its unique needs and the strengths and weaknesses of its teachers, however these basic guidelines will help improve on the weaknesses of NCLB’s teacher quality paradigm and help fulfill the legislation’s initial goal of a quality teacher for each child.