Standard 18: Evidence of Effectiveness

The program's graduates have a positive impact on student learning.

Why this standard?

The ultimate purpose of teacher preparation programs is to train teachers to be successful in their own classrooms from day one. This standard examines whether the students of teacher preparation program graduates are achieving academically.

What is the focus of the standard?

This standard examines state reports, where available, on the effectiveness of graduates of individual teacher preparation programs. To be used for evaluation, report data must be specific to particular grade spans (i.e., elementary or secondary) and be available for at least two consecutive years.

Standard applies to elementary and secondary programs in institutions in states with adequate data models.

The rationale summarizes research about this standard. The rationale also describes practices in the United States and other countries related to this standard, as well as support for this standard from school leaders, superintendents and others education personnel.

The methodology describes the process NCTQ uses to score institutions of higher education on this standard. It explains the data sources, analysis process, and how the standard and indicators are operationalized in scoring.

Standard and Indicator

Standard 18: Evidence of Effectiveness

The program's graduates have a positive impact on student learning.
Standard applies to: Elementary and Secondary programs in institutions in states with adequate data models.

Teacher preparation student performance data models that allow evaluation of teacher preparation programs now exist in only a small number of states. And even in those states, a first generation of models may soon be supplanted by a second generation. This standard and its indicator will develop in parallel with the development of teacher preparation student performance data models nationwide.

The only programs that will be evaluated are those located in states in which teacher preparation student performance data models allow association with graduates in their first year or two years of teaching with either:

- A determination of individual student growth.

     OR

- A determination of the teachers' contribution to individual student growth.

Any institution located in a state whose teacher preparation student performance data models do not meet either of these two criteria will receive a rating indicating that the standard is "not applicable."

NCTQ's own preference would be that only programs whose graduates on average are effective (i.e., at least half of a program's graduates produce student learning gains in their first year) would meet this standard. But the theory and practice of using student growth data to assess teacher preparation quality are still in their infancy, so we cannot at this point impose such a threshold. As more states build adequate models, we will adjust the standard and indicator accordingly.

Indicator that the program meets the standard:

18.1 The state's own criteria for evaluating and rating teacher preparation programs and identifying those that meet or exceed state standards will determine a program's rating under this standard.

Rationale

Standard 18: Evidence of Effectiveness

The program's graduates have a positive impact on student learning.

Standard applies to elementary and secondary programs in institutions in states with adequate data models.

Why this standard?

The ultimate purpose of teacher preparation programs is to train teachers to be successful in their own classrooms from day one. This standard examines whether the students of teacher preparation program graduates are achieving academically.

What is the focus of the standard?

This standard examines state reports, where available, on the effectiveness of graduates of individual teacher preparation programs. To be used for evaluation, report data must be specific to particular grade spans (i.e., elementary or secondary) and be available for at least two consecutive years.

Rationale

Research base for this standard

No "strong research"[1] exists on this topic.

Other support for this standard

The ultimate goal of teacher preparation programs should be to produce teachers who effectively educate their students. This standard evaluates programs by using direct evidence of effectiveness, rather than by inferring a connection between a program's practices and teacher candidates' future performance based on research findings.[2]

While this goal may have been hard to assess a few years ago, that is no longer the case. Across the country, states and school districts are increasingly measuring student learning as a way to evaluate the efficacy of teachers. In the past few years, 95 different states and school districts have received Teacher Incentive Fund (TIF) federal grants, which support their efforts to gather data on students' academic growth that can be attributed in part to teachers' efforts.[3] Concurrently, value-added measures of teachers' contributions to student learning are becoming more refined.

Given the wealth of information that is available on teacher efficacy, it seems reasonable to expect programs to

---

[1] NCTQ has created "research inventories" that describe research conducted within the last decade or so that has *general* relevance to aspects of teacher preparation also addressed by one or more of its standards (with the exceptions of the Outcomes and Evidence of Effectiveness standards). These inventories categorize research along two dimensions: design methodology and use of student performance data. Research that satisfies our standards on both is designated as "strong research" and will be identified as such. That research is cited here if it is *directly* relevant to the standard; strong research is distinguished from other research that is not included in the inventory or is not designated as "strong" in the inventory. Refer to the introduction to the research inventories for more discussion of our approach to categorizing research. If a research inventory has been developed to describe research that generally relates to the same aspect of teacher prep as addressed by a standard, the inventory can be found in the back of this standard book.

[2] Although this standard does not rely on research for its justification, the data on which the standard is based (data taken from state reports providing the results of value-added models for teacher preparation) apply research in the field of value-added models.

[3] Center for Educator Compensation Reform. *List of TIF Grantees.* Retrieved February 5, 2013, from <http://www.cecr.ed.gov/TIFgrantees/list.cfm>

first gather these data on the teachers they produce (as measured by Standard 17) and then use that information to fine-tune the program to ensure that it yields positive results. While people may disagree about the measures used to evaluate effectiveness, no one can dispute that teacher preparation programs should produce teachers who have a positive impact on student learning.

This standard also received support from school district superintendents.

Scoring Methodology

How NCTQ scores the Evidence of Effectiveness Standard

Standard and indicators

Data used to score this standard

Evaluation of elementary and secondary programs on Standard 18: Evidence of Effectiveness uses the following source of data:

- State reports on the findings of teacher preparation program student performance data models

Who analyzes the data

Two in-house staff analysts independently evaluate findings from state reports.

Scope of analysis

While the analysis of **undergraduate** and **graduate** programs uses the states' own criteria for evaluating and scoring teacher preparation programs and identifying those that meet or exceed state standards (Indicator 18.1), the following conditions are imposed:

- Findings from the state data model must pertain to a specific teacher preparation program (e.g., undergraduate elementary) rather than to the graduates from a combination of programs.

- Findings of the state data model on a specific program must be available for two or more consecutive years

- Findings of the state data model with respect to a specific program must be statistically significant.[1]

All state data models use student performance data in mathematics and reading/English/language arts. Two years of findings for programs are evaluated for an annual scoring based on whether, in considering one or both academic areas, the programs' graduates yield statistically significant positive or negative results when compared to the average performance of all new teachers in the state.[2]

To achieve the top score in the annual scoring, program graduates as a group must generate statistically significant positive results in *both* math and reading and generate no statistically significant negative results.

To achieve the middle score in the annual scoring, program graduates as a group must generate statistically significant positive results in *either* math or reading and generate no statistically negative results.

---

[1] We used state designations of statistical significance if available; if not available, we calculated whether the result was statistically significant by ensuring that the interval created by the addition and subtraction of the standard error to the result did not include the target standard for novice teachers.

[2] For **North Carolina**, the standard for comparison is the average performance of new teachers who did not graduate from the **University of North Carolina** system.

Programs achieve the low score in the annual scoring if program graduates as a group generate statistically significant negative results in *either* math or reading.

The table below shows how annual scores are evaluated (in either order) to produce a final score on the standard.

The sample of programs that could be evaluated on this standard is very small because of the following circumstances:

- Only four states have publicly released reports from their teacher preparation student performance data models and have done so for two or more consecutive years.[4]

- In two of the three states that have issued reports, some IHEs are not included because their production is below the threshold set by the state for inclusion in the model; results for the ones that are included are not given for the type of programs (e.g. undergraduate elementary, undergraduate middle school, and so on) in the *Review*.[5]

- In the case of **North Carolina**, the one state whose report includes all public IHEs and which also reports by the type of program in the *Review*, there are many programs for which statistically significant results for two consecutive years are not available.

A common misconception about how analysts evaluate the Evidence of Effectiveness Standard:

- *Findings from teacher preparation student performance data models can and should be used to evaluate every teacher preparation program.* Even if findings were program-specific rather than including graduates from several different preparation programs as they most often do, at most they would allow one to distinguish those programs about which one can be fairly confident that graduates are very effective or very ineffective relative to any given standard. Since most programs produce graduates whose performance, on average, is difficult to distinguish statistically from the overall state average to which they are being compared, findings on those programs will be ambiguous.[3] NCTQ will continue to use for the *Teacher Prep Reviews* whatever findings from data models are available, but we do not anticipate having a large sample for which evaluation under this standard will be possible.

The sample for evaluation under this standard is the one **North Carolina** undergraduate elementary program that both meets the standard's conditions and is also in the sample of programs included in the *Review*.

Annual score	Annual score	Final score on standard
High score	High score	●
High score	Middle score	◕
Middle score	Middle score	◑
High score	Low score	◑
Middle score	Low score	◔
Low score	Low score	○

---

[3] For a full discussion of the limitations of these models and design principles, see *NCTQ's Teacher preparation program student performance data models: Six core design principles* (access at: http://www.nctq.org/statePolicy/statePolicyHomeNew.do).

[4] **Louisiana**, **North Carolina**, **Ohio**, and **Tennessee** have issued reports.

[5] **Louisiana** and **Tennessee**. **Louisiana** aggregates data from graduates of elementary and middle school preparation programs; **Tennessee** aggregates data from graduates of elementary and middle school programs, as well as aggregating data from candidates who were prepared in undergraduate and graduate programs.