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 prinicples* (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.