



# General Methodology for Evaluation of Traditional Prep Programs

The general methodology for evaluation of non-traditional programs is found [here](#)

## SCOPE

The NCTQ *Teacher Prep Review* evaluates the quality of programs that provide preservice preparation of public school teachers. The Review database includes both traditional and non-traditional programs.<sup>1</sup>

**Traditional programs:** The *Teacher Prep Review* evaluates a total of 2,481 undergraduate and graduate elementary, secondary, and special education programs offered by education schools in 1,117 public and private institutions of higher education.<sup>2</sup>

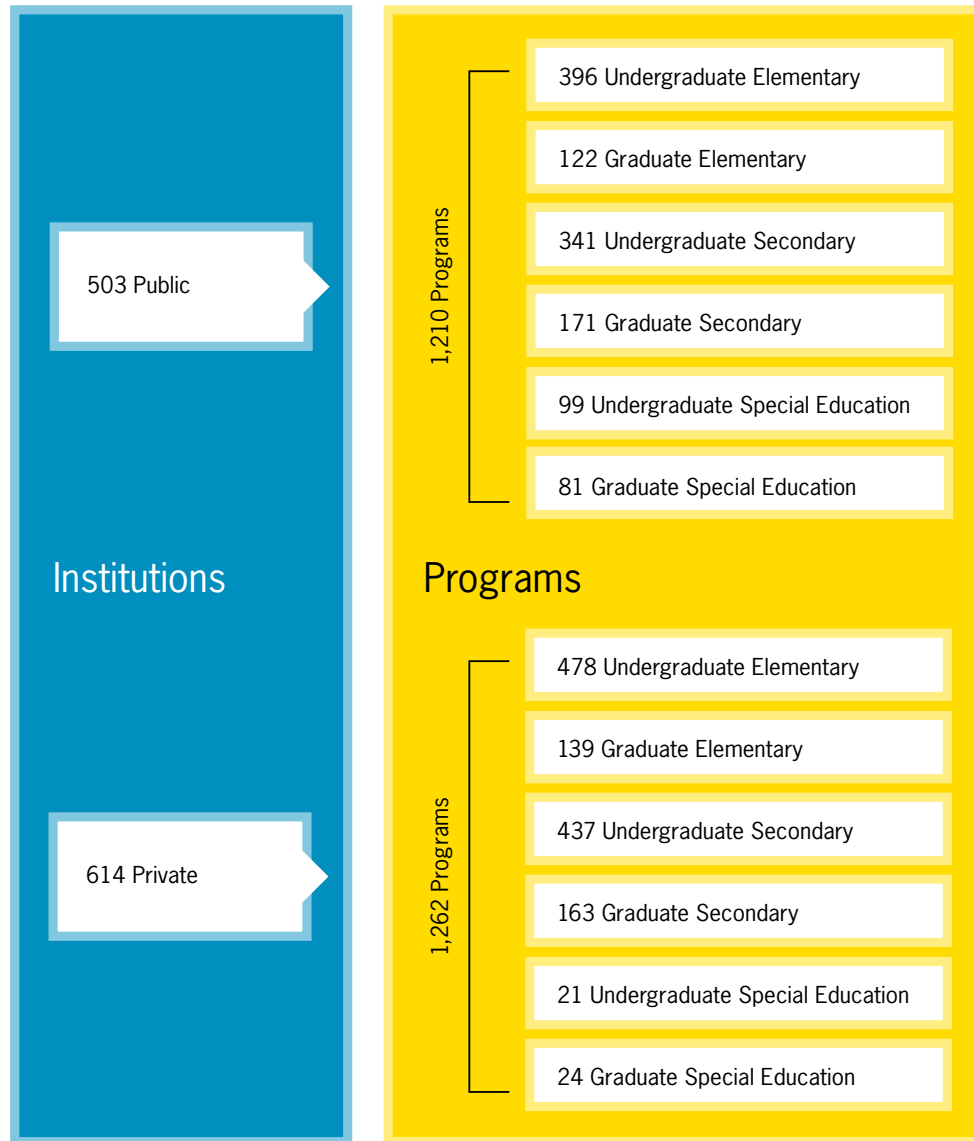
**Non-traditional programs:** Program overviews are provided for 85 non-traditional secondary programs, expanding to 146 non-traditional elementary and secondary programs in future releases.

**Total sample:** In total, over the six releases that comprise this edition of the *Review*, the *Teacher Prep Review* will post evaluations of 2,627 teacher preparation programs. These are the programs referred to as “the sample.” The approximately 240 institutions that produce fewer than 20 teachers annually (and together produce less than 1 percent of the nation’s public school teacher candidates) are not included in the sample. In past years, the *Review* has evaluated prep programs responsible for producing 99 percent of traditionally prepared teachers, and 85 percent of all teachers including those from non-traditional programs. We are expanding our reach to include even more traditional and non-traditional programs, and will update these numbers following the 2018 releases of the *Teacher Prep Review*.

1 For an explanation of the difference between traditional and non-traditional programs, see [here](#).

2 Most programs are “ranked” with a percentile rank. If they instead have a “–” in the place of a rank, it means that not all key standards could be evaluated for the program.

Fig. 1 Who is in the Review



*The Review contains evaluations of at least one program at 1,117 IHEs on at least two standards.  
In most cases, more than one program at an IHE is evaluated.*

The Review also includes ratings for 85 non-traditional secondary programs, and will expand this sample to include more secondary programs and non-traditional elementary programs in future editions. The methodology used for evaluation of non-traditional programs is available [here](#).

Our methodology is largely unchanged since the Review's inception. For example, we were very systematic in selecting traditional programs for evaluation at each of the institutions in the sample at the outset of the Review. Because the sample has remained the same, no new traditional program selection has taken place with the exception of adding additional special education prep programs. More information about program selection is included [here](#).

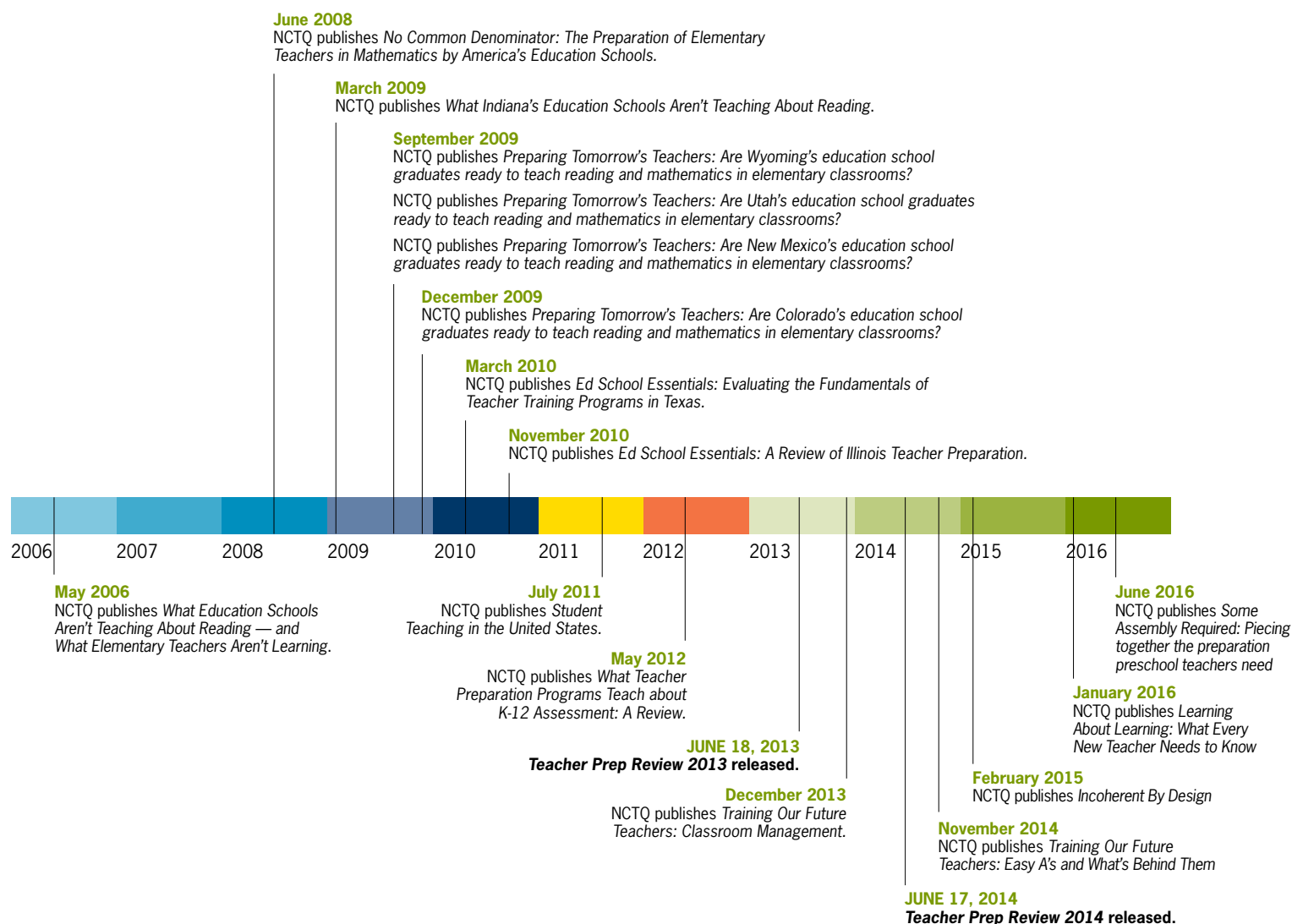
Similarly, for traditional teacher preparation program evaluation, our data collection and verification methods are unchanged, although over the past several editions of the Review, we have used more open records requests of school districts.

## TIMELINE

The development of the NCTQ standards and methodology was accomplished deliberately over a period of nine years, with 13 pilot studies of 1,597 programs in all 50 states and the District of Columbia, and field tests of 41 standards in all.<sup>3</sup> We've written a [primer](#) on traditional teacher preparation to provide some important background information. For definitions of key terms, see our [glossary](#).

An appeal to institutions to provide data for each edition of the Review is sent months in advance of the release date, and data are accepted for several months. All institutions were invited to submit new data for all standards, regardless of whether they had been evaluated on those standards in past editions of the Review.

Fig. 2 NCTQ Teacher Prep Studies timeline 2006-2016



3 All NCTQ teacher preparation studies can be accessed at: <http://www.nctq.org/reports.do?d=Teacher%20Prep>

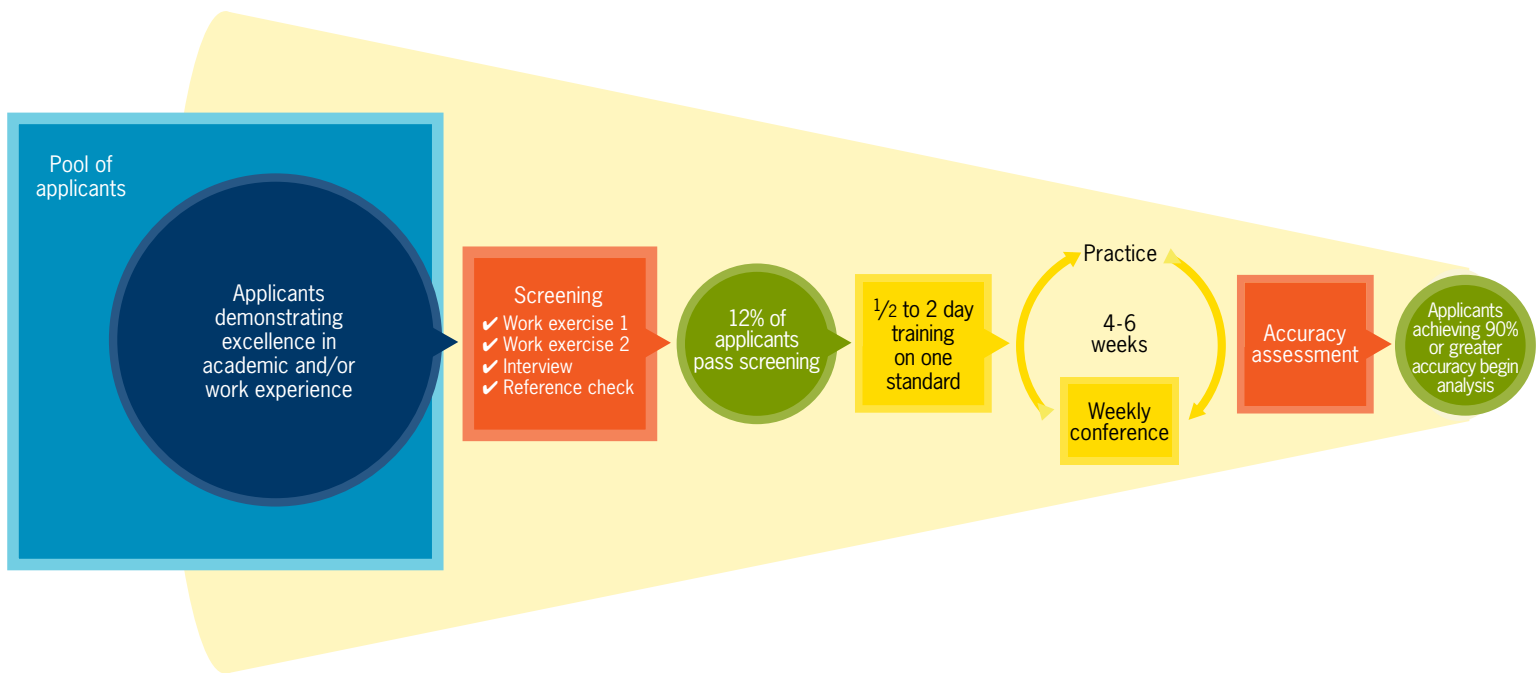
## EXPERT ADVICE AND ANALYSIS

A [Technical Panel](#) comprising teacher educators, PK-12 leaders, and education experts provides ongoing advice and support. Its members receive no compensation. The members of the panel make themselves available for consultation on a wide variety of methodological issues. Panel consensus has been achieved on all issues on which it has provided consultation.

An [Audit Panel](#), whose work will be described shortly, was also formed to advise on the reliability of scoring processes.

Each of the standards of the *Teacher Prep Review* is scored by a specially trained team. In the case of five standards — **Early Reading, Early Reading for English Language Learners, Struggling Readers, Elementary Mathematics, and Instructional Design in Special Education** — the scoring teams comprise subject specialists who participated in rigorous training processes.<sup>4</sup> Teams comprising “general analysts” who undergo both a thorough screening in the hiring process and a rigorous training process rate all other standards. The figure below illustrates how general analysts were selected and trained.

Fig. 3 Qualifications and training of general analysts



*Rigorous screening and training prepares NCTQ's corps of general analysts to accurately evaluate programs on selected standards.*

<sup>4</sup> Biographical information on subject-specialist analysts can be found [here](#).

## STANDARDS

Standards are a crucial governing feature of every institution involved in education, including teacher preparation programs. NCTQ's standards focus on what programs should do to prepare teachers to teach to the high level required by college- and career-readiness standards, and their application is measurable.

NCTQ developed these criteria for policies and practices to raise the level of training of the nation's teacher workforce through a number of different sources.

To the extent that high-quality research can inform how teachers should be prepared, NCTQ uses that research to formulate standards. Unfortunately, research in education that connects preparation practices to teacher effectiveness is both limited and inconsistent in quality. Our standards for the *Teacher Prep Review* are also based on the consensus opinions of internal and external experts, the best practices of other nations, the states with the highest performing students, and, most importantly, what superintendents and principals around the country tell us they look for in the new teachers they hire. The standards have been refined over ten years by 13 national and state studies, and by consultation with experts on NCTQ's Technical Panel. Because many were developed before increasingly rigorous state student learning standards had been implemented, they have also been honed to ensure alignment with those standards.

We continue to develop new standards, including one on **Rigor of Grading** that was applied to undergraduate teacher preparation programs in a separate [report](#) released in fall 2014, and one on **Fundamentals of Instruction**, which was developed in a new [report](#) in winter 2016 and will be used to evaluate secondary programs in the future.

### MORE ON THE RATIONALES FOR OUR STANDARDS AND THE RESEARCH BEHIND THEM

For each of our standards, we have developed a rationale that lays out the support found in research and other sources. These rationales can be found in the "[Understanding Our Standard](#)" documents we have created for every NCTQ standard for teacher preparation used in the *Teacher Prep Review*. All but two of the "Understanding Our Standards" also contain an inventory of research that has any bearing on the type of preparation addressed in the standard. The purpose of the inventory and the means by which it was developed is found in an [introduction](#).

We welcome an ongoing discussion with others — state policymakers, accrediting bodies, prep program leaders, teacher educators, and teachers — about the best way to evaluate teacher preparation program quality.

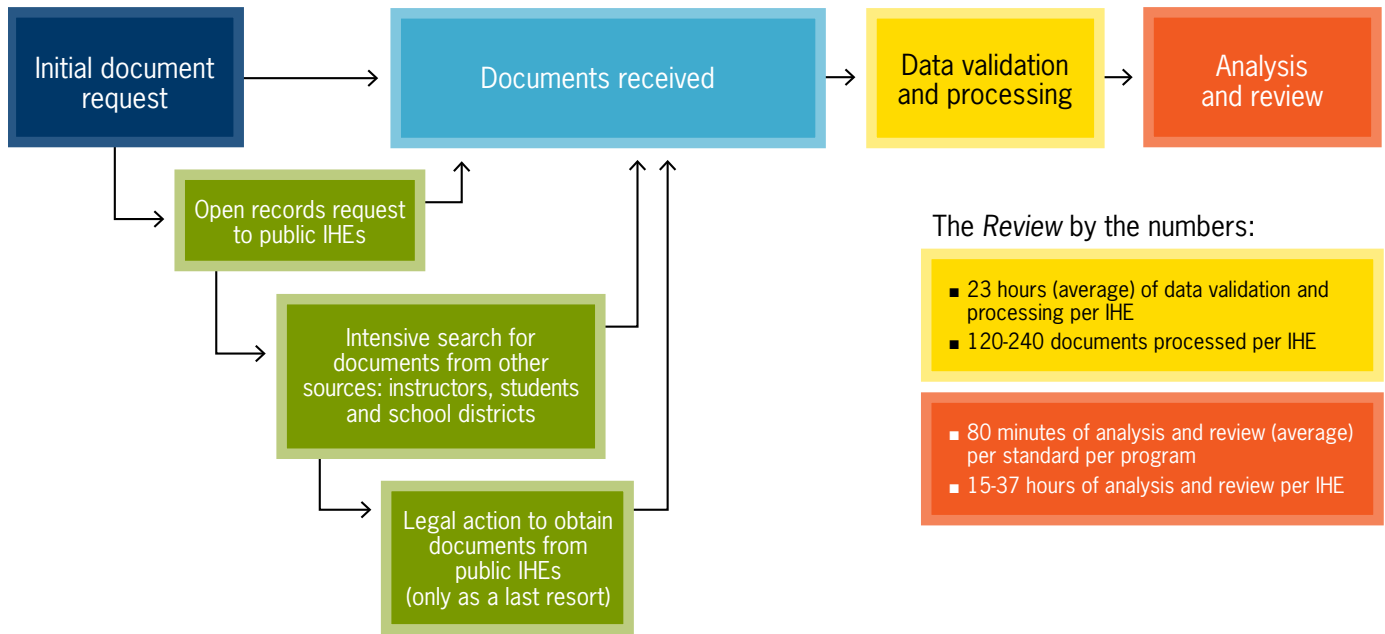
## DATA COLLECTION, VALIDATION, AND ANALYSIS

There's much to say about the process of data collection, validation, and analysis, but the most important fact to keep in mind is that *all* 1,117 institutions housing programs in the sample have been asked to submit new data for this release. Those who cooperated, or for whom we were able to collect data without cooperation, were evaluated on all key standards (and additional standards, if possible). We have seen increasing rates of cooperation with institutions since the first edition of the *Review*. For some institutions that still choose not to provide data, we have obtained much of the necessary data in spite of their lack of cooperation. Through these efforts, we are able to evaluate most programs on key standards.

### Data collection

We have established a wide array of techniques to collect and validate the data we need for the *Teacher Prep Review*. As always, our chief concern is ensuring that we obtain valid data that accurately reflects the training these institutions provide teacher candidates.

Fig. 4 Data collection, processing and analysis



*NCTQ draws upon 12 sources of data to evaluate all standards for elementary, secondary, and special education programs.*

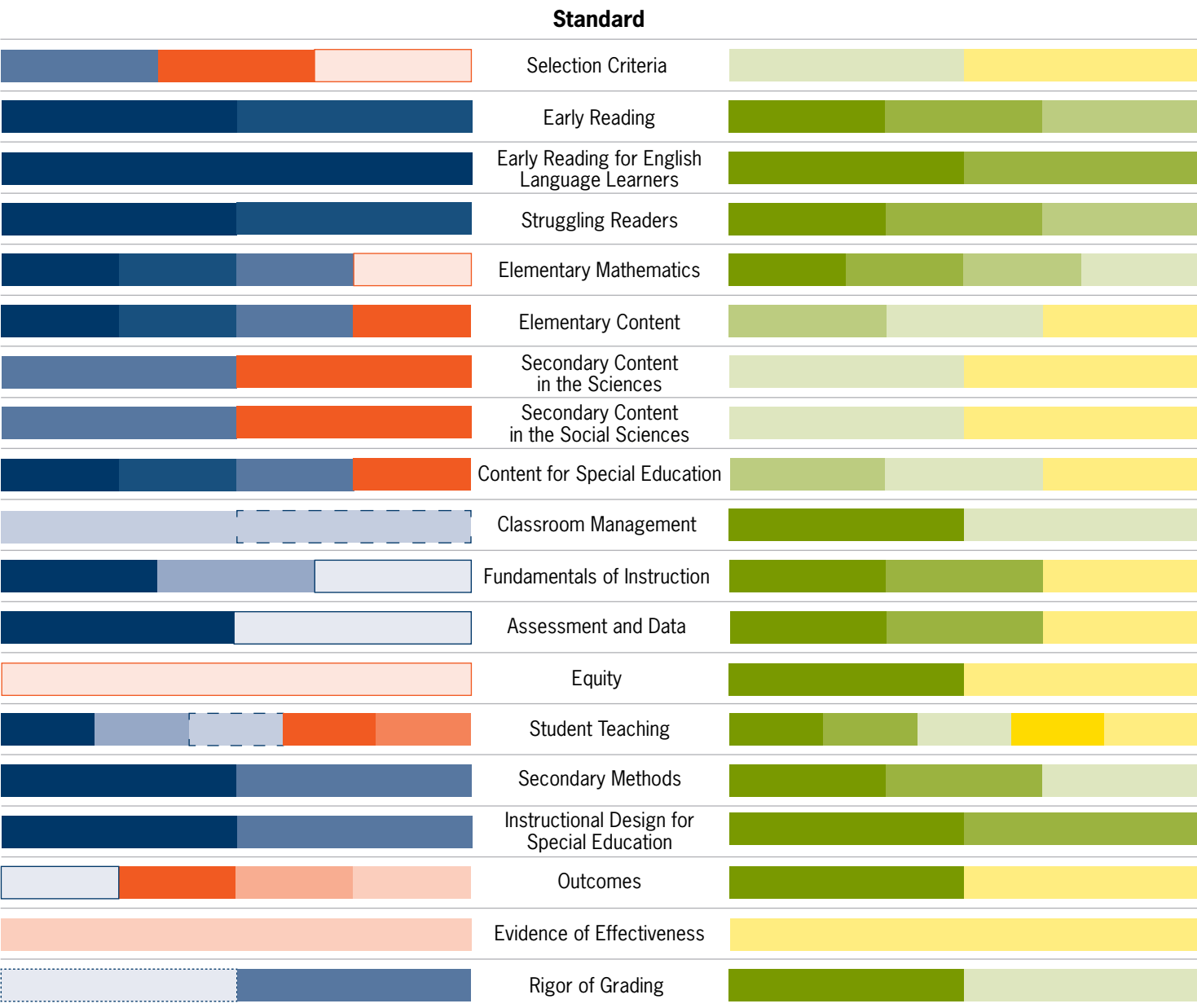
To determine what data we needed from institutions and to gather those data for program evaluation, we begin by analyzing each program, and reviewing university catalogs and other program material posted publicly by the institution. By this means we identify general education and professional course requirements, along with course descriptions.<sup>5</sup>

After a comprehensive review of this publicly posted material, we ask the institutions for materials such as syllabi for particular courses,<sup>6</sup> information on graduate and employer surveys, and material related to student teaching placements.

<sup>5</sup> With the exception of evaluation of coursework requirements for the standard on **Instructional Design in Special Education**, requirements for general education and professional coursework were taken from catalogs. In the case of the **Instructional Design** standard, catalog descriptions of requirements proved so difficult to decipher that degree plans were consulted. In a recent comparison of catalog requirements with those in “degree plans” provided by institutions, we found that there are substantial differences between requirements listed in catalogs and degree plans for the same academic year. To the extent that they conflict, we take catalogs to provide a more authoritative source of requirements.

<sup>6</sup> If multiple sections of the course were offered, the institution could select the section whose syllabus would be sent (providing it was for a specified academic year, not including summer sessions unless only offered in summer).

Fig. 5 Data sources for the Review’s standards for traditional teacher prep



Types of data from

IHEs:

- Syllabi
- Required textbooks
- IHE catalogs
- Student teaching handbooks
- Student teaching evaluation forms
- Capstone project guidelines (including Teacher Performance Assessments)
- Graduation ceremony programs

Sources external to IHEs:

- State regulations
- IHE–District correspondence/contracts
- Graduate and employer surveys
- State data on IHE performance
- IHE demographic data

Data obtained from

IHEs:

- IHEs and/or individual faculty
- Teacher candidates or other students at IHE
- IHE bookstores
- IHE websites

Sources external to IHEs:

- K-12 school districts
- Websites and reports provided by K-12 school districts, state and/or federal agencies

A variety of data, obtained from multiple sources, were used for evaluation.

These are the methods NCTQ uses to collect data:

- 1. Open-records requests to institutions.** All 50 states and the District of Columbia have open-records laws (also known as “sunshine,” “freedom of information act” or “FOIA” laws) that require public agencies to turn over documents upon request by an individual or organization. Except in **Pennsylvania** and **Illinois**, public universities are almost universally considered public agencies under these laws.<sup>7</sup> But although these institutions are publicly approved to prepare public school teachers, teacher preparation programs at private institutions are not. To collect data for the *Review*, we send out an individualized request to each of the state’s programs, asking them to work with us. If they decline, or do not respond after 10 days, we follow up with a formal open-records request listing the documents, including the course syllabi, we required. We made open-records requests of 375 public institutions that initially chose not to work with us.
- 2. Open-records requests to school districts.** Teacher preparation programs partner with one or more school districts to arrange for student teaching as the crucial apprenticeship experience candidates need before taking the reins of a classroom. Programs often provide student teaching handbooks to districts and sign formal contracts or memoranda of understanding with districts that set forth the criteria and processes by which mentor teachers are chosen. To capture this material, we send open-records requests to more than 1,000 districts across the country for the *Review*.
- 3. Online searches.** We judiciously search online for information we need for the *Teacher Prep Review*. Professors post syllabi and programs put up student teaching handbooks on institutional websites — all of this material is publicly accessible. We also periodically collect information on each semester’s textbook listings from institutions’ online bookstores. We do not use a syllabus that is posted online unless we can confirm it is valid and current using dates, required reading that matches bookstore records, and other information.
- 4. Campus outreach.** Because we need such an extensive array of documents for our evaluation (see Fig. 5 for a full list of the data needed for each standard) and because of the resistance we face, the methods outlined above are insufficient, particularly for private institutions. So we have reached out to people on campuses, particularly faculty, to ask them to provide us with the documents we need.

## Data validation

Regardless of the source, each and every document we receive has to be carefully checked to determine whether it is valid. The syllabi should clearly list the course number and, where appropriate, section number, as well as the professor’s name. For courses where we analyze textbooks (reading and elementary math), the syllabi also need to have a list of assigned textbooks.

Trained general analysts working under the supervision of our team leaders perform these thorough checks. At times we have to go back to institutions that have supplied us with documents in response to an open-records request to obtain more complete versions of documents we had requested.

## Data analysis

Teacher prep programs must document their standard policies and procedures because they and their institutions need to communicate with their “consumers” (generally their students), and/or because programs are regulated entities that must interact regularly with various institutions (state agencies, accrediting bodies, and local school districts, among others). Our evaluations are largely based on the documents containing policies and procedures. *Descriptions* of policies and procedures provided to us by institutions in lieu of the actual policy statements are not accepted as data that can satisfy any part of a standard.

<sup>7</sup> Four public universities in **Pennsylvania** (**Lincoln University**, **Pennsylvania State University**, **Temple University** and the **University of Pittsburgh**) are specifically exempted from its open-records laws. In Illinois, educational institutions are not required to hand over course materials, including syllabi, in response to an open-records request, apparently for fear that fulfilling such requests would enable cheating.



For example, we often find cover letters from institutions accompanying submitted data to be very helpful in navigating through the many files provided, but statements in the letters are not used in analysis unless they are corroborated by language in official documents.

Our evaluations can be described as “low inference.” Analysts are trained to look only for evidence that teacher preparation programs have particular features related to admissions, content preparation, and professional preparation. For example, when evaluating observation forms that provide feedback to teacher candidates on their use of classroom management techniques in student teaching placements, analysts determine whether the forms contain references to specific techniques. Analysts do not attempt to ascertain whether anything — for example, about the nature of rubrics or instructions to university supervisors conducting observations — will lead to valid and reliable feedback on classroom management. However, it is indisputable that a teacher candidate is more likely to receive feedback on a specific management technique if it is explicitly noted on the feedback form than if it is not noted at all. Our evaluations can therefore distinguish programs that are potentially better-designed to provide great training from those that are less well-designed.

### Scoring processes

Our scoring processes place the full collection of documents relevant for evaluation at the disposal of an analyst after a very methodical and systematic process of coding and sorting. Analysts have been trained to follow a very detailed and systematic standard-specific protocol to make a “yes” or “no” decision about whether each of a standard’s indicators is satisfied.<sup>8</sup> (Scoring methodologies abstracted from these protocols are included in the *Understanding Our Standard* documents, available [here](#).) When an indicator is satisfied, the analyst has to identify the relevant data and document the source. If the indicator is not satisfied but there is information that bears on the indicator, the analyst has to identify the data that are “next closest” to satisfying the indicator and document the source. If there are no data related to the indicator, the analyst has to make an explicit statement to that effect. All data entered in our database are automatically annotated with the date and the analyst’s name. The figure below provides a guide to possible scores by standard.

<sup>8</sup> In very few instances, the analysts make a “yes” or “no” decision on a sub-indicator: Several **Classroom Management**, **Student Teaching**, and **Assessment and Data** indicators are scored by sub-indicators. Due to the structure of the standards for which subject-specialist evaluations are required (**Early Reading**, **Early Reading for English Language Learners**, **Struggling Readers**, **Elementary Math**, **Instructional Design in Special Education**), analyst decisions are not indicator-specific, but focus instead on gathering findings in a manner that is highly structured, detailed, and well documented.

Fig. 6 Possible grades by standard for traditional teacher prep

	A+	A/Pass	B	C	D/Not Pass	F
Selection Criteria	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Early Reading	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Early Reading*	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Early Reading for English Language Learners	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Struggling Readers	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Elementary Mathematics	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Elementary Mathematics*	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Elementary Content	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Secondary Content in the Sciences	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Secondary Content in the Social Sciences	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Content for Special Education	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Classroom Management	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Fundamentals of Instruction	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Assessment and Data	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Equity	Reported only					
Student Teaching	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Secondary Methods	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Instructional Design for Special Education	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Outcomes	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Evidence of Effectiveness	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Rigor of Grading	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Rigor of Grading**	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

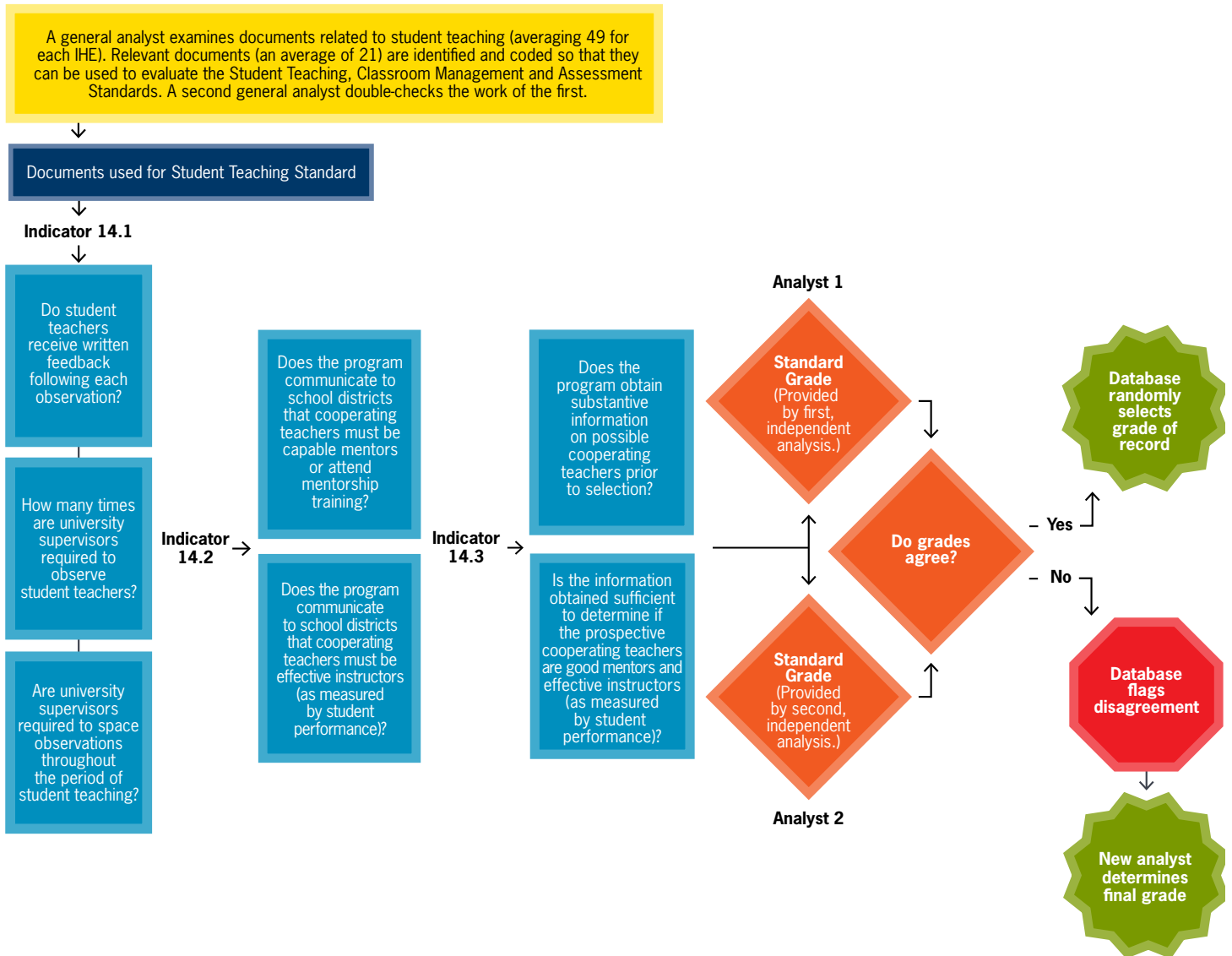
\* Scoring process using imputation.

\*\* Scoring process using more specific data.

For most standards, scores are provided on a 5-part scale, with some standards also offering commendation for Strong Design (A+).

Programs are evaluated on standards by one general analyst. However, for each standard, we randomly select a sample of programs (generally twenty percent of programs evaluated on the standard)<sup>9</sup> for evaluation by a second analyst. The figure below provides a graphic depiction of this process for the **Student Teaching Standard**.

Fig. 7 Steps in scoring a standard with two analysts, using the Student Teaching Standard as an example



*Each standard's grading process involves multiple indicator-related determinations which, for twenty percent of programs on each standard, are made independently by two analysts. Analysis for the other eighty percent of programs is completed by a single analyst.*

In each case, based on the indicator evaluations, a standard grade between "A" and "F" corresponding to a range of scores from "meets standard" to "does not meet standard," is automatically generated. For selected standards, programs that not only meet but go above and beyond the requirements of the standard earn "strong design" on that standard, represented as an "A+."

9 The Early Reading, Early Reading for English Language Learners, Struggling Readers, and Elementary Math Standards are evaluated by only one subject-specialist, with 10 percent of programs evaluated by two analysts to monitor scoring variances. The Evidence of Effectiveness Standard is evaluated sequentially by two in-house analysts.

For the sample of programs that are analyzed by two analysts, when the score produced by both analysts is identical, the analysis of one is chosen randomly by the database to represent the final score. As is explained in greater depth in the description of the RevStat management system posted [here](#), any difference of one grade in program scores based on evaluations by two analysts (for example, one evaluation leading to a score of “B” and one leading to a score of “A”) leads to “coding up,” an automatic awarding of the higher of the two scores. Any difference of two or more levels in scores triggers an “exceeds variance” signal that requires team leader investigation and resolution.<sup>10</sup> Instances of excessive variances are monitored through the RevStat process; whenever variances approach 10 percent, action is taken to improve fidelity to scoring protocols or to modify the scoring process as necessary.

## STATE CONTEXT

States regulate teacher preparation programs extensively. A teacher preparation program must show that it meets its state's standards to earn approval to train and recommend candidates for licensure and must undergo reapproval every five to seven years thereafter.

We therefore thoroughly examine all relevant state regulations as part of our scoring processes for every standard. We begin with the findings of our comprehensive [State Teacher Policy Yearbook](#) and investigate further when necessary. In considering state regulations, we follow these general principles:

- **Give credit for building on strong regulations:** We give credit to programs explicitly affirming state regulations that improve program quality. In **Texas**, for example, programs that affirm that they only admit applicants who achieve scores on the Texas Higher Education Assessment (THEA) that exceed by any amount the state's initial thresholds meet the **Selection Criteria Standard**.<sup>11</sup>
- **Hold programs responsible for ensuring candidates are prepared:** The ambiguity and complexity of state regulations do not relieve programs of doing what is necessary to make sure that their graduates are well equipped to help students learn. For example, 21 states offer only K-12 certification for special education teachers. Programs in those states have an obligation to make sure that their special education candidates have adequate content knowledge, so we evaluate programs for content preparation for both the elementary and secondary grades.

## THE IMPACT OF STATE REGULATIONS ON OUR ANALYSIS

To provide a more detailed sense of how state regulations impact our analysis, we provide examples below of two standards where context is crucial, and two standards where it has no impact whatsoever.

### State regulations on expectations for secondary teacher subject knowledge

Ratings for two of our traditional teacher preparation standards—Secondary Content in the Sciences and Secondary Content in the Social Sciences (as well as the analogs of these standards that we apply to non-traditional programs) are deeply informed by the state regulatory context in which programs are embedded. The starting point of our analysis is the state's licensing test regime: Does it test all subject matter that a secondary science or social sciences teacher will need to know for all the subjects he or she could be assigned to teach? The more comprehensive a state's testing regime, the less possibility that a secondary teacher will be assigned to teach a course without knowing his or her subject. Where there are gaps in testing, we scrutinize the content of coursework that programs require of their candidates.

For “unitary” subjects such as math, tests are generally an adequate guide to content preparation: Math teacher candidates who are tested only in math can generally only teach math classes. For the social sciences and the sciences, however, state licensing regimes are generally not robust enough. In some states, teachers earning a license in “general science” can teach

<sup>10</sup> When necessary, the “exceeds variance” trigger is adjusted to be more sensitive and provide additional oversight.

<sup>11</sup> While Texas initially set a minimum passing score for the THEA, it has since dropped that threshold. However, the Selection Criteria Standard credits programs that adhere to the threshold initially set by the state.

high school physics without ever having to demonstrate that they know physics. In other states, a person who majored in anthropology could teach U.S. history classes without ever taking more than one or two courses in the subject. In these cases, we take a closer look at whether programs in these states are doing what they should to prepare teachers for the courses to which they could be assigned to teach.

A general consequence of our approach for these standards is that a state's licensing regime provides a ratings backstop for its programs: Programs generally can do no worse than the strength of their state's licensing test system, and can take steps to do better.

(To learn more about how state context impacts these standards, see this [infographic](#) and the scoring methodologies for the [Secondary Content in the Sciences](#) and [Secondary Content in the Social Science Standards](#).)

## State expectations for elementary teacher preparation in early reading and elementary mathematics

State context plays virtually no role in our analysis for these two standards. States do generally articulate expectations for what elementary teachers need to know in these subjects, and some states have good tests for them. Nonetheless, we decided to carefully examine the preparation that programs provide candidates without regard to the regulatory framework in which programs were embedded.

The logic behind taking an approach so different from the one taken with regard to secondary content is simple: Preparation in these subjects is a core responsibility of teacher preparation programs themselves. No liberal arts faculty members can deliver courses in how to teach children how to read. And although elementary math courses can and should be delivered by math faculty, these courses have to be specifically designed with the needs of elementary teachers in mind. A math department at an institution without an elementary teacher preparation program would not offer any courses like the ones elementary teacher candidates need to take.

## CONNECTING STANDARDS TO PROGRAMS

Because of the limited cooperation from institutions, not all standards factor into the scores for programs. This section explains which standards are applied to what programs and how standard scores and program rankings are reported. Scores on “key standards” are used to develop the base for program rankings; scores on a “booster standard” can move a program up in rankings from this base.

Overall elementary, secondary, and special education program rankings are based only on “key” and “booster” standards, even for the programs for which we were able to score on more standards. We made this decision so that the rankings for any given type of program would be based on scores on the same standards.

Program rankings include weighted scores on individual key standards.<sup>12</sup> In elementary program rankings, the weights of scores on the **Selection Criteria Standard** are heaviest, with scores on the **Student Teaching Standard** next heaviest, and scores on the **Early Reading, Elementary Math**, and **Elementary Content** weighted least but equally.<sup>13</sup> In secondary program rankings, the weights of scores on the **Selection Criteria Standard** are heaviest, followed by **Secondary Content in the Sciences** and **Secondary Content in the Social Sciences**, both equally weighted, and then by **Student Teaching**, and finally by the subject-specific methods course component of the **Secondary Methods Standard** weighted

<sup>12</sup> For programs that earn “strong design” in a key standard, the weight of the score associated with the strong design is enhanced beyond the weight of a score of “meeting the standard.”

<sup>13</sup> Program rankings for special education programs are weighted in essentially the same way, except that the weight of scores on the **Instructional Design for Special Education Standard** is weighted slightly less than the Student Teaching Standard, with scores on **Early Reading, Elementary Math**, and **Elementary Content** then least heavily (and all equally) weighted. Special education program rankings are boosted (in order of weight) by scores on the **Classroom Management Standard**.

least.<sup>14</sup> In special education program rankings, the **Selection Criteria Standard** is weighted most heavily, followed by **Student Teaching** and then **Instructional Design; Early Reading, Elementary Math**, and **Content for Special Education** are weighted least but equally.

Program rankings can be increased, or “boosted,” by scores on the **Classroom Management Standard** (for all programs) and the subject-specific practice component of the **Secondary Methods Standard** (for secondary programs).

When we lack the adequate clear data we need to evaluate a program on a particular standard — in most instances, because the program fails to provide it — we do not score it on the standard. There are, however, instances in which the program does supply the material we requested but a score cannot be determined because the materials are not clear. In both cases, the program is removed from the set of programs evaluated on the standard and the score is left blank. *In no instance is a program given a score on the basis of whether it does or does not provide data.*

For two standards, **Early Reading** and **Elementary Mathematics**, a method of imputing scores was developed after extensive fieldwork to ensure that a lack of complete sets of syllabi due to resistance to NCTQ’s requests would not preclude a score on these critical standards.<sup>15</sup>

## QUALITY CONTROL

NCTQ’s priority in all of its studies of teacher preparation has been to conduct its evaluations with integrity and to produce reliable results. Because of the scale of the *Teacher Prep Review* and the vast number of decision points involved in data collection, processing, and analysis, continuing to produce reliable results demanded new mechanisms and safeguards. With the development of a scoring management system component in our database, we have been able to make quality control an integral, ongoing feature of our evaluation.

### RevStat

RevStat, a scoring management system that is designed to be an integral part of NCTQ’s teacher preparation database, manages a variety of aspects of analysis reliability. Using RevStat, the *Teacher Prep Review* team tracks each standard’s reliability of scores across pairs and teams of analysts at any given time and across various time periods. If reliability issues emerge, the scoring protocols and training are recalibrated appropriately.

To develop RevStat, NCTQ partnered with **UPD Consulting**, a national expert on education management. NCTQ and UPD modeled RevStat on the same principles as the **Baltimore CitiStat** and the **New York City CompStat** processes, which have proven effective in managing institutional performance.

### Audit Panel

Although RevStat provides invaluable data on scoring processes, we wanted to ensure that we had the advice of experts who could have the broadest possible vantage point on the reliability of our work. For that reason, we invited a group of eminent

14 The Secondary Methods standard was previously a booster standard. To give proper weight to the importance of this standard, the first indicator of the standard, which evaluates whether programs require a subject-specific methods course and which can be evaluated for virtually all programs based on publicly available data, is now a key standard. The second indicator, which considers whether coursework gives the opportunity for subject-specific practice teaching and which necessitates access to course syllabi, remains a booster standard.

15 We estimate that in 80 percent of programs, this scoring approach produces the same program scores in the **Elementary Math Standard** as evaluation with complete data. We estimate that in 70 percent of programs, this scoring approach produces the same program scores in the **Early Reading Standard** as evaluation with complete data. The Pearson correlation coefficient of scores in the elementary math content sample when scores are calculated using all available data and when scores are imputed is  $r=.6$ . The analogous correlation in reading is  $r=.87$ . In elementary math the correlation increases to  $r=.81$  when we compare the scores for programs in which there is no subject overlap between or among the elementary math content courses for each program, 60 percent of the sample. Program ratings for programs evaluated by these alternate processes are reported as “Pass” (3.5 on a 0-4 scale) or “Not Pass” (1 on a 0-4 scale).

education researchers to join an Audit Panel to provide technical assistance, critique our evaluation processes to date, and recommend improvements in subsequent *Teacher Prep Reviews*. Discussion with the panel has reassured us regarding the utility of the steps we have taken to date to ensure reliability and suggested some refinements we have adopted.

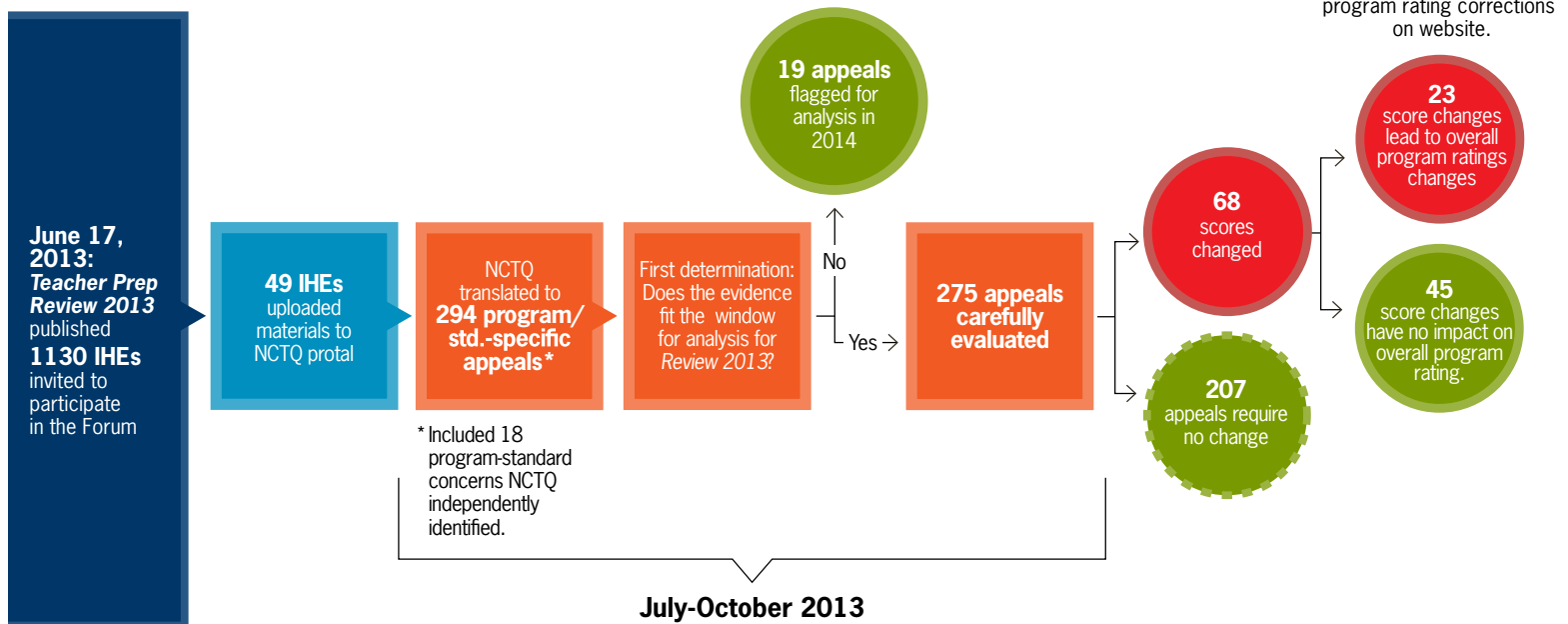
### Forum for appeals of scores

While we rigorously ensure accuracy throughout our ratings process, given the thousands of rating decisions analysts make and the frequency of change by teacher prep programs, some rare errors may occur. When programs believe a rating is in error or does not reflect their current practice, they can appeal that rating through our forum process. Starting with the 2016 Review, we provide programs with an overview sheet that includes their scores and scoring comments for each standard in advance of the public release of new rankings. This advance notice provides programs with an opportunity to submit any corrections or ask for clarification around their program grades. After the publication of each edition of the Review, we open up an appeals process that extends for several months, through which programs can discuss or contest our findings.

In the appeals process, programs are invited to send in objections to our findings along with supporting evidence.<sup>16</sup> We will update programs' standard scores and overall rankings on a regular schedule, and programs' Overview sheets will indicate when new material has been submitted in order to revise a score.

In the first *Teacher Prep Review* Forum, 49 institutions sent in direct appeals. We also addressed the objections to our findings posted by five institutions on the [website](#) of the American Association of Colleges for Teacher Education. In the second Forum, 62 teacher prep programs were involved in the appeals process. The majority of changes, summarized [here](#), were due to programs submitting new material.

Fig. 8 Example Forum process from *Teacher Prep Review* 2013



The Forum process provides additional transparency to NCTQ's scoring processes and is available to every IHE with a program in the Teacher Prep Review.

<sup>16</sup> The only source of evidence that can be submitted in the appeals process that were not used in the original evaluation of standards are end-of-course examinations and average SAT/ACT scores for the program.



National Council on Teacher Quality

1120 G Street, NW, Suite 800

Washington, D.C. 20005

Tel: 202 393-0020 Fax: 202 393-0095

Web: [www.nctq.org](http://www.nctq.org)