Appendix B: Methodology

The methodology for the alternative certification pilot study is found in Appendix C.

Scope

The *NCTQ Teacher Prep Review* evaluates the quality of programs that provide preservice preparation of public school teachers.

In conjunction with *Teacher Prep Review 2014*, *U.S. News & World Report* posts the *ranking* status of a total of 1,612 undergraduate and graduate elementary and secondary programs offered by education schools in 1,127 public and private institutions of higher education institutions.[1] Combined with additional rankings on NCTQ's website of 55 special education programs (and evaluations of an additional 40 special education programs on some standards), this second edition posts evaluations of at least some standards of 2,400 teacher preparation programs offered in 1,127 institutions. (These are the institutions referred to as "the sample.") The 343 institutions producing fewer than 20 teachers annually (and together producing less than 1 percent of the nation's public school teacher candidates) will not be included in the sample for any edition of the *Review*.[2]

Fig. B1 Two ways of looking at the *Review's* coverage

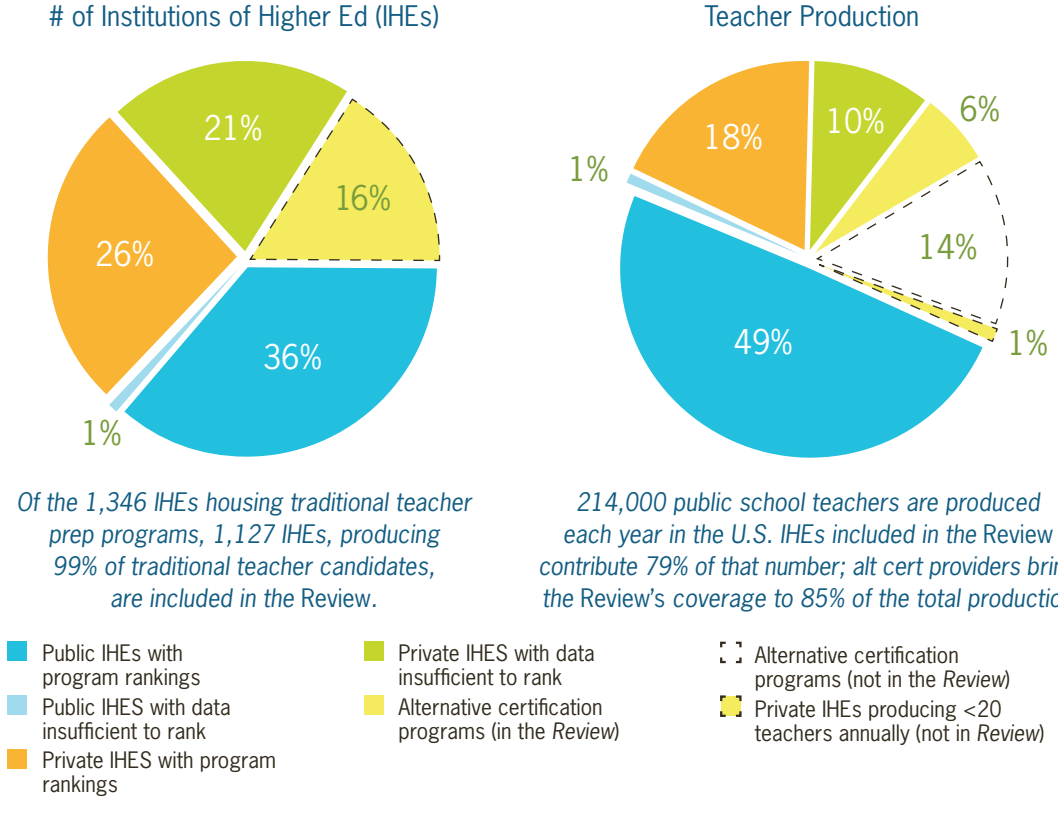
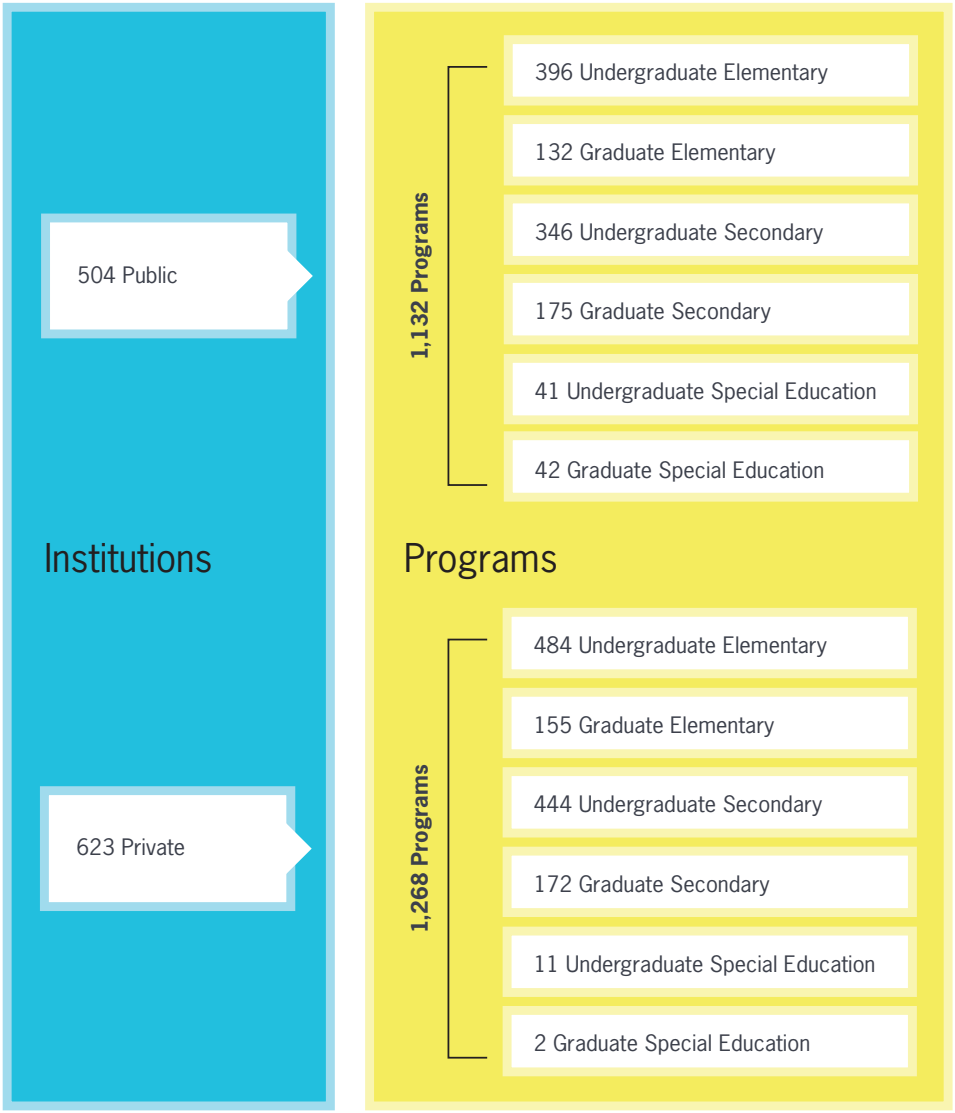


Fig. B2 Who is in the *Review*



The Review *contains evaluations of at least one program at 1,127 IHEs on at least two standards. In most cases, more than one program at an IHE is evaluated.*

The big change in coverage between the first and the second editions is not in the number of institutions in the sample or the number of traditional programs in those institutions that are evaluated; the change lies in the *scope of evaluation* of the sample and the programs it contains. In the first edition, 1,200 elementary and secondary programs were evaluated on the key standards; in the second edition, an additional 412 programs that were previously evaluated on only a few standards were evaluated on all key standards, for a total of 1,612 elementary and secondary programs evaluated on all key standards.

In the *Review*'s third edition, we plan to expand the scope of evaluation again to include evaluation on all key standards for the programs for nearly all or all of the institutions in the sample and expand our ratings of alternative certification programs.

This edition also includes ratings for 85 secondary alternative certification programs.

Our methodology is largely unchanged. For evaluation of traditional teacher preparation programs, the second edition of the *Review* builds on the framework for analysis established in the first edition. For example, we were very systematic in selecting programs for evaluation at each of the institutions in the sample. Because the sample has remained the same, no new program selection has taken place.[3] Thus information on the process of program selection need not be repeated here, but is still accessible here.

Similarly, for traditional teacher preparation program evaluation, our data collection and verification methods are unchanged, although the different mix of public and private institutions on which we are gathering data means that for this second edition we have used open records requests of institutions less, while we have used open records requests of school districts and collection from students and faculty more.

Timeline

The development of the NCTQ standards and methodology was accomplished deliberately over a period of nine years, with 10 pilot studies of 583 program evaluations in all 50 states and the District of Columbia, and field testing of 39 standards in all. We've written a primer on traditional teacher preparation to provide some important background information. For definitions of key terms, see our glossary.

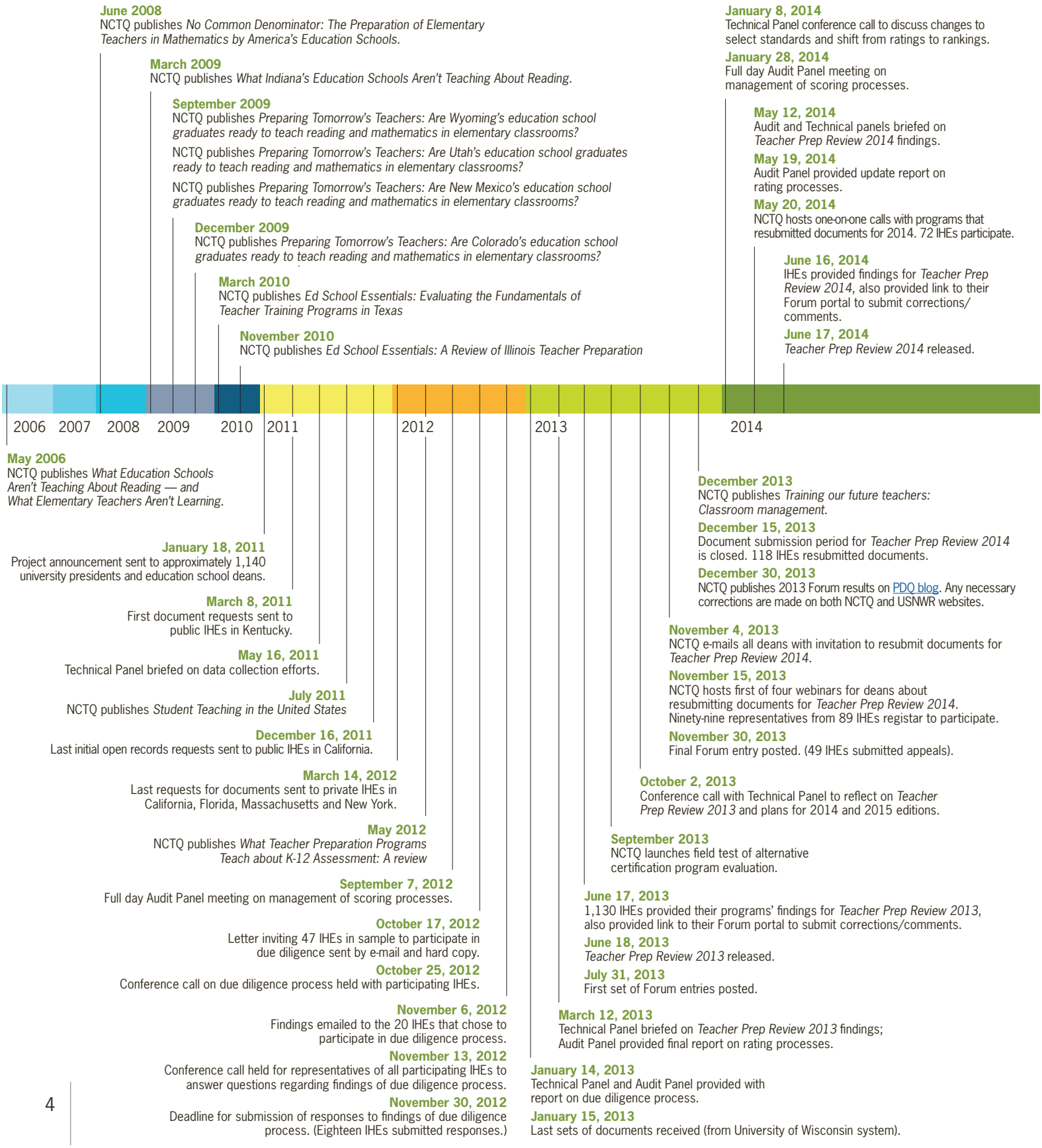
An appeal to institutions to provide data for the second edition of the *Review* was sent on November 4, 2013. Data was accepted until December 15, 2013. For institutions that had already submitted full sets of data for evaluation on all standards, this was an invitation to update the data; for institutions that had never submitted data (or had not submitted sufficient data for evaluation on all key standards), it was an invitation to submit new data.

Staff

In-house staff members' expertise in the preparation necessary to become an effective teacher is broad and deep:

- Julie Greenberg, Senior Policy Analyst (who taught secondary mathematics for 13 years in Maryland's Montgomery County Public Schools), has overseen two of NCTQ's national studies on teacher preparation, six of its state studies, and the first edition of the *Review*.
- Robert Rickenbrode, Director (a former teacher and chief academic officer of a network of charter schools), developed all operational aspects of the current *Teacher Prep Review* as an outgrowth of his work on NCTQ's **Texas** and **Illinois** studies as well as the first edition of the *Review*.
- Of the seven staff members with teaching experience, six received their teacher certification through traditional preparation programs; all but one worked on the first edition of the *Review*.

Fig. B3 *NCTQ Teacher Prep Studies* timeline 2006-2014

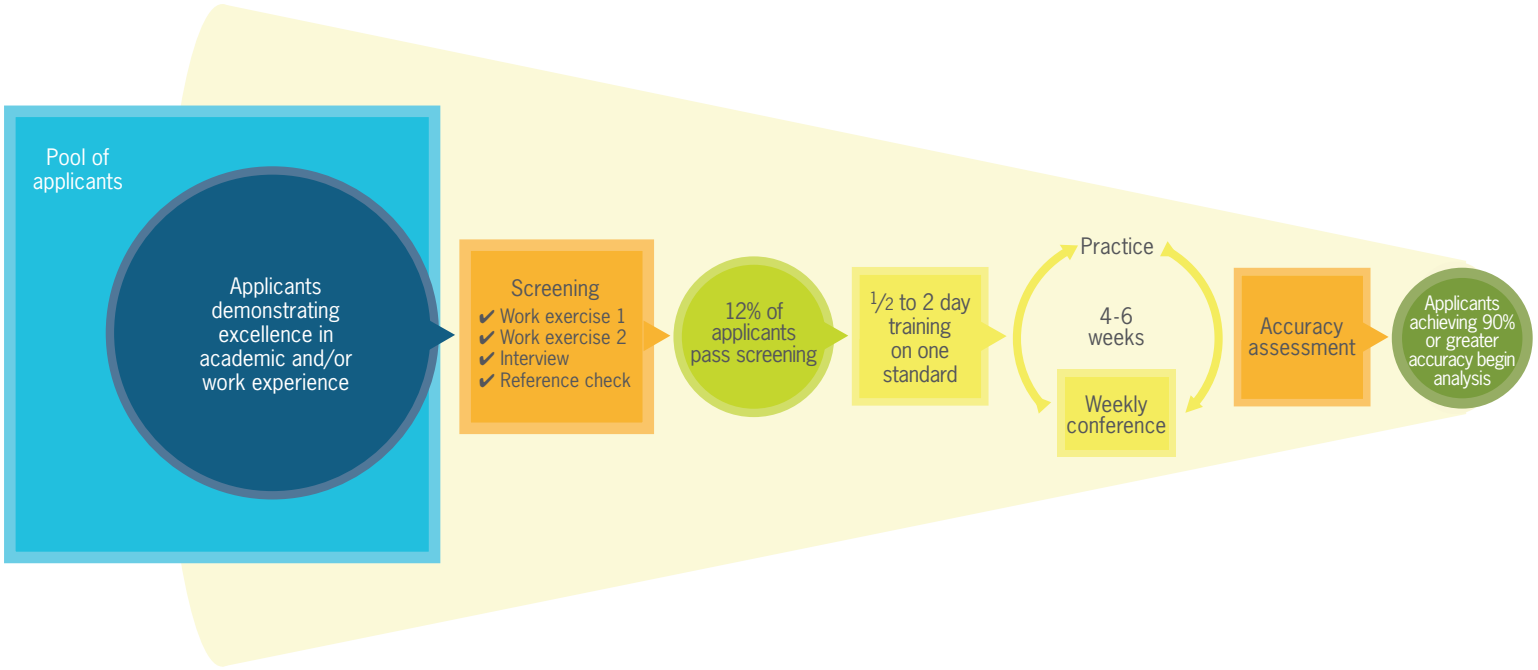


A Technical Panel comprising teacher educators, PK-12 leaders, and education experts provides ongoing advice and support. Its members receive no compensation. The members of the panel make themselves available for consultation on a wide variety of methodological issues. Panel consensus has been achieved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

An Audit Panel, whose work will be described shortly, was also formed to advise on the reliability of scoring processes. The panel has posted a statement on its oversight of our maintenance of the reliability of our ratings processes.

Except for the **Evidence of Effectiveness Standard**, which is evaluated by staff, each of the standards of the *Teacher Prep Review* is scored by a specially trained team. In the case of five standards — **Early Reading**, **English Language Learners**, **Struggling Readers**, **Elementary Mathematics** and **Instructional Design in Special Education**[4] — the scoring teams comprise subject specialists who participated in rigorous training processes.[5] 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. B4 Qualifications and training of general analysts



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

Standards

Standards are a crucial governing feature of every institution involved in education, including teacher preparation programs. What sets NCTQ's standards apart from other standards is that they focus on what programs should do, at minimum, to prepare teachers to teach to the high level required by college- and career-readiness standards. Moreover, we actually use the standards to *measure* programs, as difficult and controversial as the results may be. Our analysis has revealed that some states do not hold teacher preparation programs accountable for meeting the state's own standards.

NCTQ developed its expertise in 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 spotty. 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 nine years by 10 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 have been implemented, they have also been honed to ensure alignment with those standards.

We continue to develop new standards, including one on **Rigor** that will be applied to undergraduate teacher preparation programs in a separate report to be released in fall 2014.

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 "standard book" we have created for every NCTQ standard for traditional teacher preparation used in the *Teacher Prep Review*. All but two of the standard books 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, teacher educators, and teachers — about the best way to evaluate teacher preparation program quality.

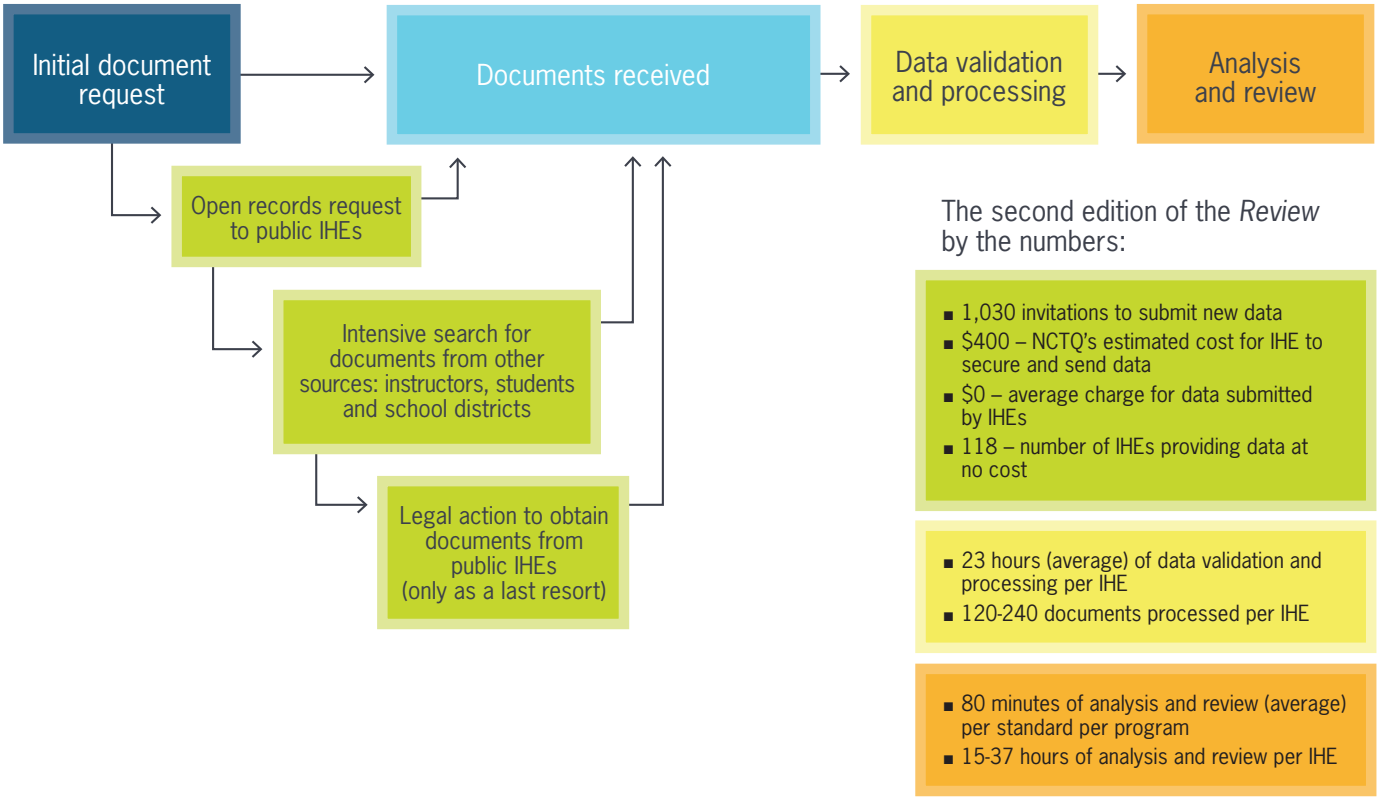
Data collection, validation, and analysis

There's a lot to say about the process of data collection, validation, and analysis, but the most important fact to keep in mind is that *all* 1,127 institutions in the sample housing traditional teacher preparation have been asked to submit data in our first round of collection, in early March 2011. Those that cooperated, or for whom we were able to collect data without cooperation, were evaluated on all key standards (and additional standards, if possible) for the first edition; a share of those institutions that did not originally cooperate now have programs evaluated on key standards by cooperation (a very few) or by use of data we have obtained in spite of their lack of cooperation. We anticipate that most if not all of the last remaining institutions will be evaluated on key standards in the third edition of the *Review*.

Data collection:

The first edition of the *Review* stimulated a nationwide boycott of our effort that has only now begun to subside. We have had to devise a wide array of techniques to collect and validate the data we need for the *Teacher Prep Review*. As always, our chief concern was ensuring that we obtained valid data that accurately reflected the training these institutions provide teacher candidates.

Fig. B5 Data collection, processing and analysis



NCTQ draws upon 12 sources of data to evaluate all standards for elementary, secondary and special education programs (see Fig. B5).

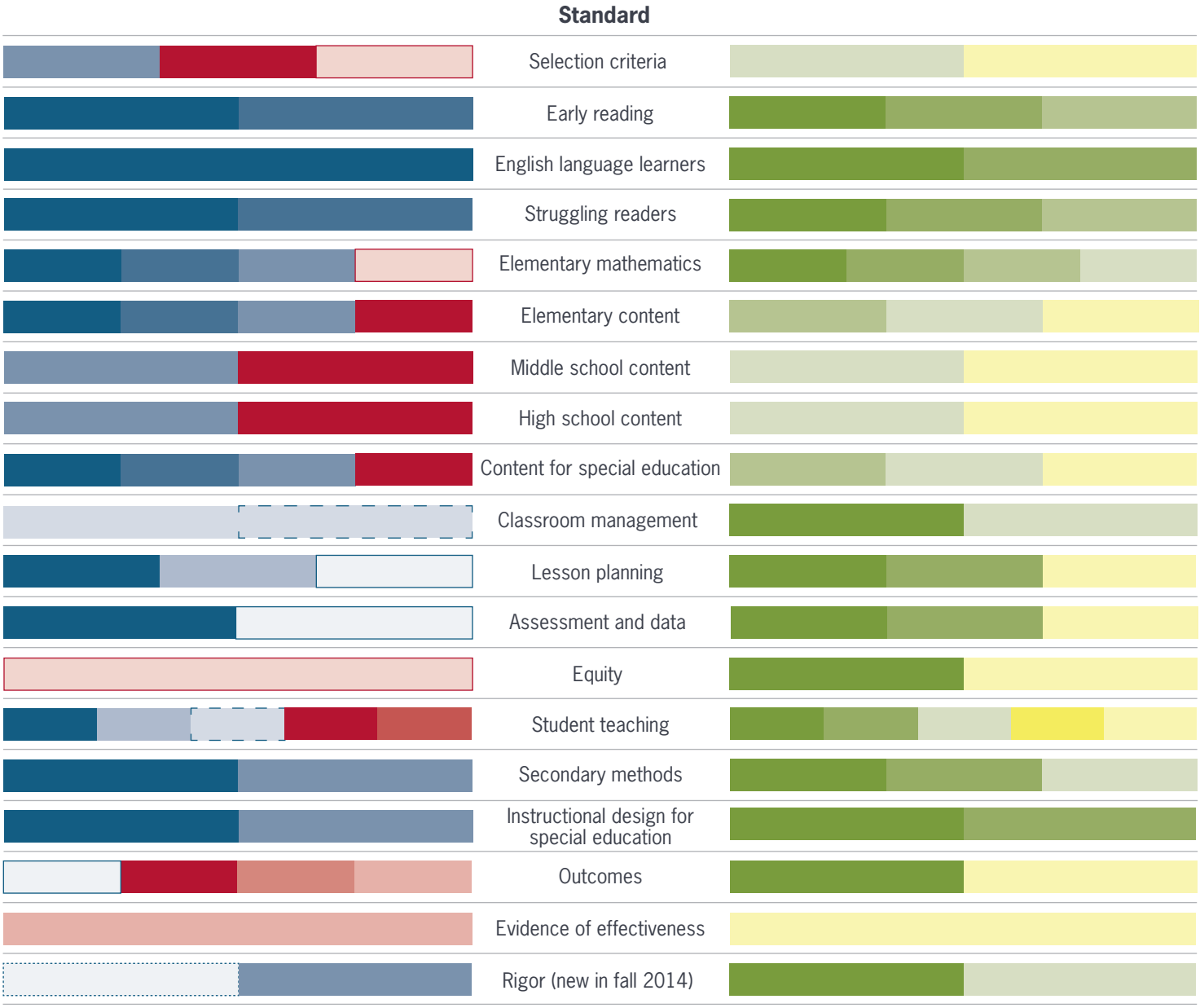
To determine what data we needed from institutions and to gather data for program evaluation, we began by analyzing each program and reviewing university catalogs and other program material posted publicly by the institution. By this means we identified general education and professional course requirements, along with course descriptions.[6]

After a comprehensive review of this publicly posted material, we asked the institutions for materials such as syllabi for particular courses,[7] information on graduate and employer surveys, and material related to student teaching placements.

Even if an institution's programs could not be evaluated on key standards for the first edition of the *Review* because it had not provided data in response to our request, we sent a renewed request for data to evaluate the programs on key standards in the second edition. In a small handful of cases, institutions that had previously been unresponsive sent in data. We also followed up with open records requests to public institutions in spite of a lack of response to our first round of requests — again, in a small handful of cases, we found the institution more cooperative.

NCTQ continues to litigate to obtain access to data, but we have made relatively little use of open records requests to institutions in order to obtain data for the second edition of the *Review*. We had fully rated at least one program at the vast majority of public institutions in the first edition; 17 institutions continue to charge us too much to provide documents or refuse to turn over syllabi because of copyright concerns.[8] We are more focused on obtaining information

Fig. B6 Data sources for the *Review*'s standards for traditional teacher prep



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

from school districts on student teaching placement using a variety of data collection means, including open records requests.

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.[9] But although private institutions are publicly approved to prepare public school teachers, teacher preparation programs at private institutions are not required to respond to open records requests. To collect data for used in the initial edition of the *Review*, we made open records requests of 475 public institutions that initially chose not to work with us.[10] We made an additional 20 open records requests to collect data for the evaluations in the second edition.

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 sent out open records requests to more than 1,000 districts across the country for the first edition of the *Review* and to 1,150 districts for the second edition.

3. Online searches.

   We exhaustively 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 generally accessible. We also periodically collect information on each semester's textbooks listings from institutions' online bookstore. 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. B6 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 to ask them to provide us with the documents we needed.

Data validation

Regardless of the source, each and every document we receive has to be carefully checked to determine whether it is valid. Documents need to be clearly dated; we do not rate components of programs that were in place before 2009.

We can only accept syllabi that were distributed to students in an actual course. The syllabi therefore have to 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.

In the first edition of the *Review*, our auditing focused on whether programs had provided us with "counterfeit" syllabi that they thought would do better on our standards than the syllabi distributed to students that actually reflect the training candidates receive.[11] (Conversely, we also checked on whether syllabi provided to us only by students were genuine. The number of fake syllabi that students tried to pass off to us was negligible.) In the second edition, we have conducted an audit of a sample of programs that posted significant score increases in each standard. The audit is still ongoing, but in no case to date have we found invalid scores.

Data analysis

Standard policies and procedures of teacher preparation programs must be documented because 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 never accepted as data that can satisfy any part of a standard.

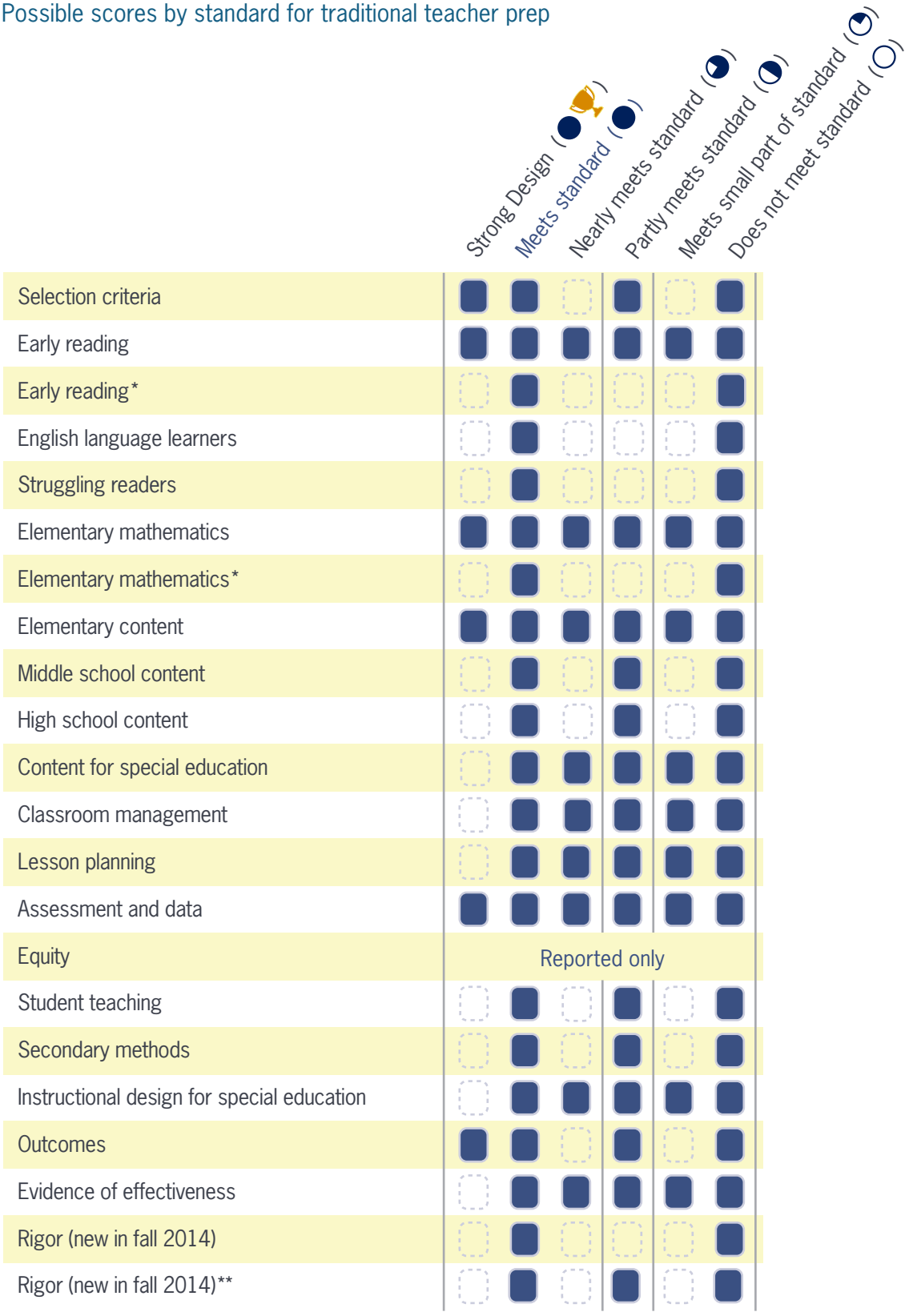
For example, we often find cover letters from institutions accompanying submitted data to be very helpful in navigating through the many files provided, but statements in the letters are not used in analysis unless they are corroborated by language in official documents.

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, in 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 than if it is not noted at all. Our evaluations can therefore distinguish stronger programs from weaker ones.

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.[12] (Scoring methodologies abstracted from these protocols can be accessed 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 is automatically annotated with the date and the analyst's name. The figure below provides a guide to possible scores by standard.

Fig. B7 Possible scores by standard for traditional teacher prep

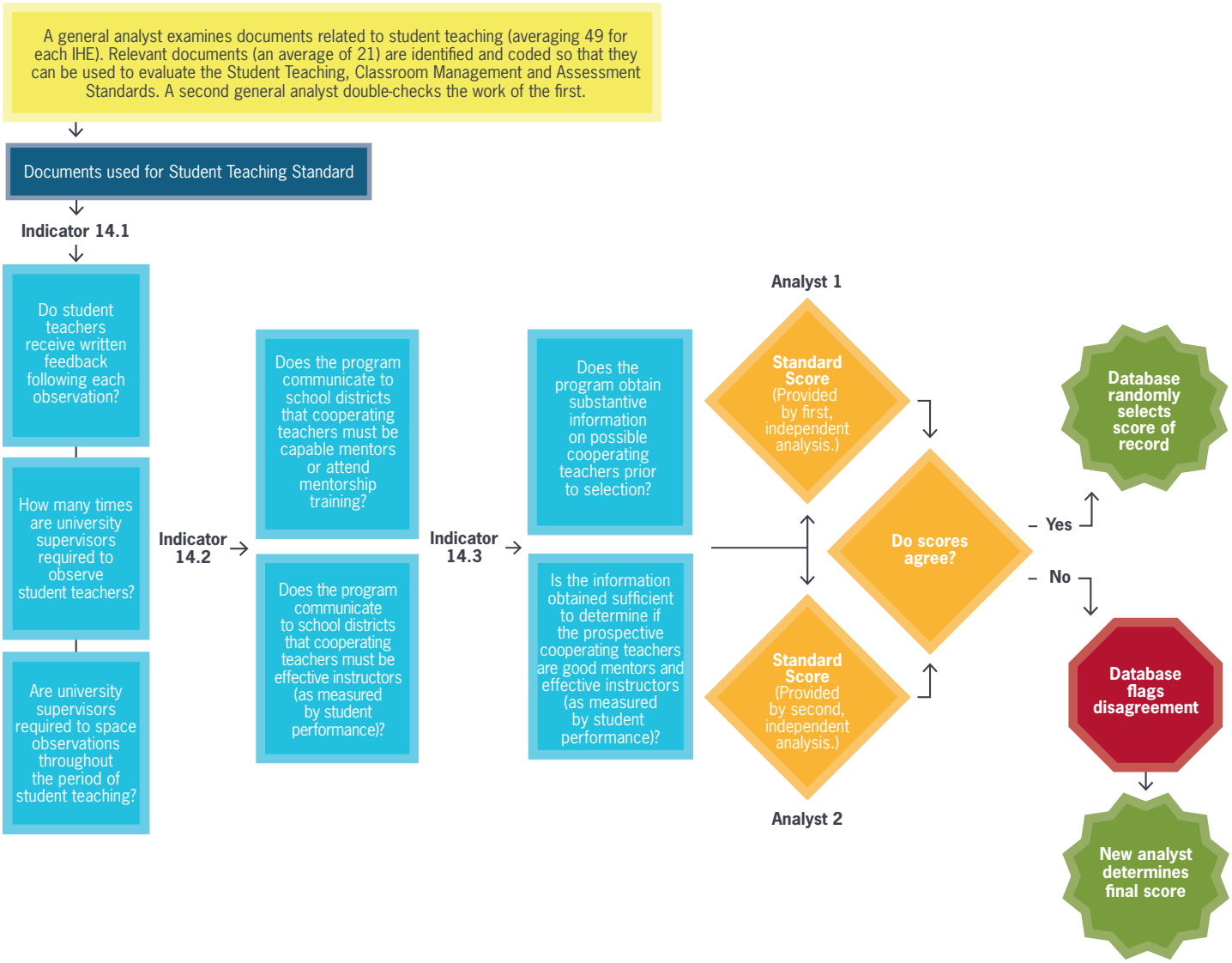


* Scoring process using imputation.
** Scoring process using more specific data.

*For most standards, scores are provided on a 5-part scale, with some standards also offering a special gold trophy commendation for Strong Design. For two standards, scores may be imputed; imputed scores are represented by ●*or ○*.*

For most of our scoring processes,[13] two *general analysts* make independent evaluations of relevant evidence to ascertain if it demonstrates that the program satisfies individual indicators for a given standard. The figure below provides a graphic depiction of this process for the **Student Teaching Standard**.

Fig. B8 Steps in scoring a standard, using the Student Teaching Standard as an example



Each standard's scoring process involves multiple indicator-related determinations which, for the majority of standards, are made independently by two analysts.

In each case, based on the indicator evaluations, a whole number standard score between "0" and "4," corresponding to a range of scores from "does not meet standard" to "meets standard," is automatically generated.

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 later in greater depth in the description of the RevStat management system posted here, any difference of one level in program scores based on evaluations by two analysts (for example, one

evaluation leading to a score of "nearly meets standard" and one leading to a score of "meets standard") 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.[14] Instances in which there are 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.[15]

State context

States regulate teacher preparation programs extensively, if not always effectively. 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. Despite these regulations, states' actual track record in holding the line on teacher preparation quality is dismal: In 2011, the last year for which data are available, only nine programs among the many thousands of teacher preparation programs housed in more than 1,400 institutions were deemed "low performing," a category that implies censure but not, generally speaking, action.[16]

Even though the states' track record of enforcement of their standards is not good, state standards nonetheless limit what programs can and cannot do. 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 three general principles:

- *Hold programs harmless:* We do not penalize programs for following their states' regulations where they run counter to our standards. For example, in **Connecticut**, local school boards are granted sole authority to choose cooperating teachers, so we do not downgrade programs on the **Student Teaching Standard** for not taking an active role in selecting them for their student teachers. This Connecticut regulation governing selection of cooperating teachers is one of very few instances where the standards of the *Teacher Prep Review* conflict with state regulations.
- *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 thresholds meet the **Selection Criteria Standard**.
- *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, 28 states offer only PK-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 — the **Middle School Content** and **High School Content** standards (as well as the analogs of the latter standard that we apply to alternative certification programs) are deeply informed by the state regulatory context in which programs are embedded. The starting point of our anal-

ysis is the state's licensing test regime: Does it test all subject matter that any given secondary 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 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.

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 middle school and high school content standards.)

State expectations for elementary preparation in early reading and elementary math

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.

Standard/program connections

Because of the limited cooperation from institutions, there is a complicated landscape of scores and program rankings for traditional programs. See Fig. 5 of the *2014 Teacher Prep Review* report, for a guide to what standards were 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 "booster standards" can move a program up in rankings from this base.

Overall elementary and secondary program rankings that we report to *U.S. News & World Report* are based only on "key" and "booster" elementary and secondary 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.[17] 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.[18] In secondary program rankings, the weights of scores on the relevant content standard(s) is heaviest,[19] with the weights of scores on the **Selection Criteria Standard** next heaviest and scores on the **Student Teaching Standard** weighted least.

Elementary program rankings can be increased, or "boosted," by scores (in order of weight) on the **Classroom Management**, **Outcomes**, **Struggling Readers** and **English Language Learners Standards**; secondary program rankings can be boosted (in order of weight) by scores on the **Classroom Management**, **Outcomes** and **Secondary Methods Standards**.

When we lacked the adequate data we need to evaluate a program on a particular standard — in most instances, because the program failed to provide it — we did not score it on the standard. There are, however, instances in which the program *did* supply the material we requested but a score could not be determined because the materials are not clear, the program is removed from the set of programs evaluated on the standard, and the score is given as "not rated" or "NR." *In no instances is a program given a score on the basis of whether it did or did not provide data.*

In addition, we scored large sets of programs on the **Lesson Planning** and **Assessment and Data** standards**,** but the sets did not include all of the programs whose submitted data included data relevant to these standards. The fact that a program may not have received a score on one or more of these standards does not imply that there was either a lack of cooperation on the part of its institution or that there was a lack of clarity in materials; the program may simply be one that was not included in the set of programs evaluated on the standard.[20] We report that these standards are "not rated" for those programs that are not in the limited evaluation sets.

For two standards, **Early Reading** and **Elementary Mathematics**, a method of imputing scores was developed after extensive fieldwork to ensure that a lack of data would not preclude a score. Because elementary preparation is critical to ensuring that elementary and special education teacher candidates are competent to enter the classroom, NCTQ could not allow the lack of cooperation by institutions to place them out of the reach of evaluations on these standards.[21]

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 demand 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 underlying causes are identified, the scoring protocols and training are recalibrated appropriately.

In development of RevStat, NCTQ partnered with **UPD Consulting**, a national expert on education management. NCTQ and UPD modeled RevStat on the same principals 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 education researchers to join an Audit Panel to provide technical assistance, critique our evaluation processes to date, and recommend improvements for subsequent editions of the *Teacher Prep Review*. 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. It also has pointed us toward measures we intend to implement in subsequent editions of the *Teacher Prep Review* that will allow us to better understand any sources of variance in scoring processes — and thereby use RevStat even more productively. The panel has signed a summary statement on the reliability of our current scoring processes.

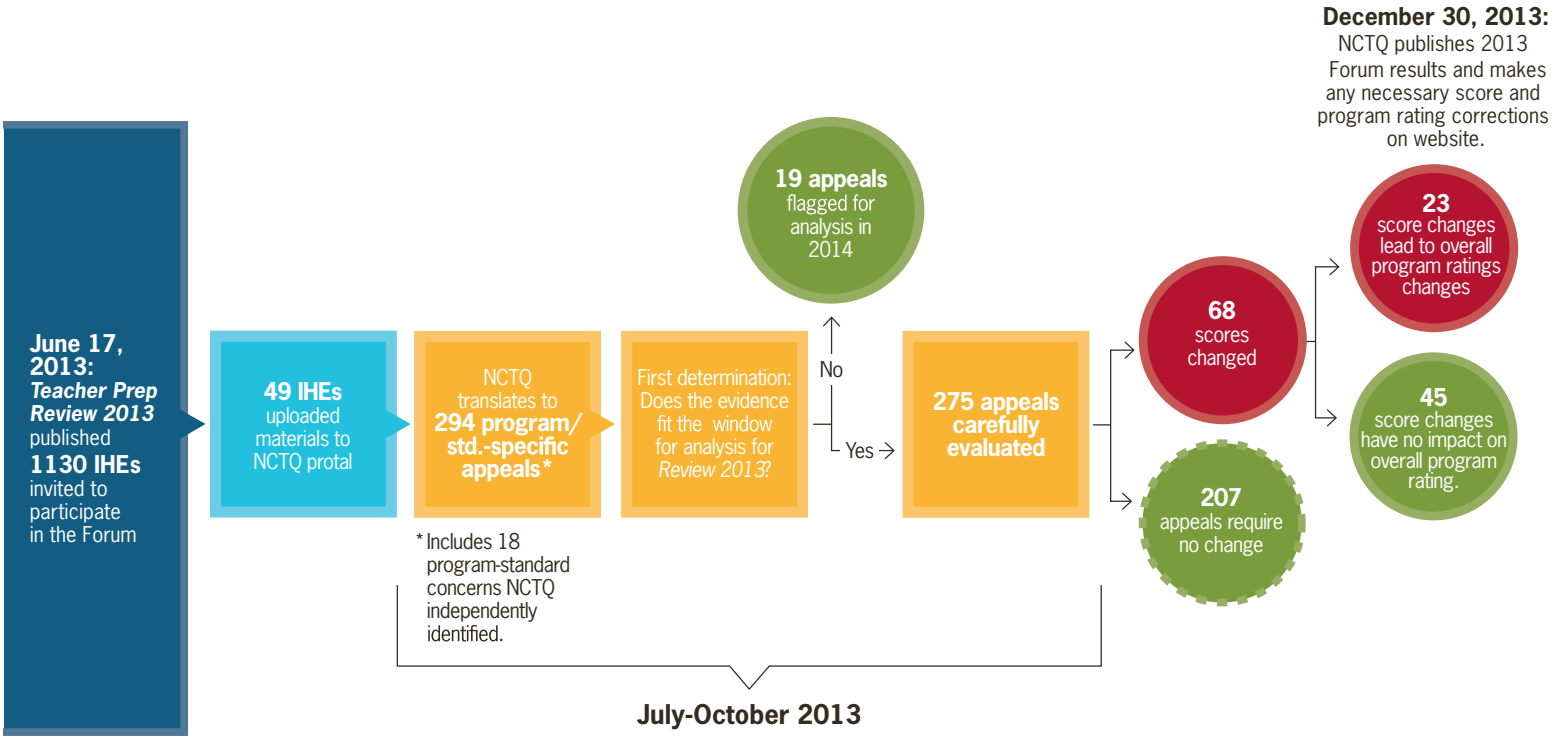
Forum for appeals of scores

After the publication of the first edition of the *Review* in June 2013, we opened up an appeals process that extended to mid-August 2013 through which programs could contest our findings. We also undertook due diligence on our own that revealed sources of error in a small fraction of scores on our **Early Reading** and **Elementary Math** standards. The same appeals process will be available with the publication of the *Review's* second edition.

In the appeals process programs are invited to send in objections to our findings along with supporting evidence.[22] Programs must agree to let us post these materials on our Forum webpage alongside our responses so that any interested third parties can assess for themselves whether our assessments are accurate.

In the first 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 (AACTE).

Fig. B9 *Teacher Prep Review 2013* Forum process



The Forum process provides additional transparency to NCTQ's scoring processes. The Forum will re-open shortly after release of Teacher Prep Review 2014.

In the course of responding to objections to some scores on our **Early Reading Standard**, we discovered that the scoring algorithm was flawed.[23] A careful re-check revealed that the flaw impacted the scores of 18 programs. We also realized that we had not accounted for the fact that Oklahoma elementary certification in math covers grades K-6 rather than K-8, as it does for other subjects.[24] As a result, we re-examined the scores of 12 Oklahoma programs on this standard, which resulted in five score changes.

All told, the Forum process led to changes to 68 scores on individual standards, which resulted in changes to 23 program ratings. We also determined that the *Review* had incorrectly included two programs that do not lead to initial certification.

Discussion on a due diligence process on all scoring processes conducted prior to the completion of the first edition and discussion of limitations can be found here.

Endnotes

1 Programs are designated as "ranked" (with a numeric ranking), "rank not reported" (bottom half of the ranked program), and "not ranked" (evaluations could not be completed on a sufficient number of standards to rank).

2 All production information is based on federal Title II reports. There were 239 small producers in the 2011 Title II report.

3 The only program selection that remains for future editions is to enlarge the selection of special education programs for evaluation.

4 On the **Instructional Design in Special Education Standard**, to address potential conflicts of interest for analysts evaluating programs who are familiar with instructors through professional networks, all documents used in evaluating for this standard were redacted to eliminate identifying references. Because of the limited number of cases in which material relevant to this standard was submitted for evaluation for the second edition, it was evaluated in-house rather than by these experts.

5 Biographical information on subject-specialist analysts can be found here.

6 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.

7 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).

For reading courses, we asked to be provided with syllabi from all sections.

8 During the development of the first edition of the *Review*, 57 institutions in 12 states claimed that course syllabi are not subject to open-records requests because they are the intellectual property of the faculty who wrote them. This conflicts with the near-universal interpretation that syllabi can be used by any entity, including NCTQ, under the "fair use" provisions of federal copyright law, providing that the use does not in any way infringe on the rights of the faculty who created them. NCTQ's use would not infringe on those rights. We litigated these claims in nine states. On October 31, 2012, a county court in **Minnesota** delivered a ruling in our suit against the **Minnesota State College and University System** indicating that "[a]ny way this case is analyzed, NCTQ is entitled to the copies of the syllabi it seeks." The System has chosen to appeal the ruling (though the **University of Minnesota** system was persuaded to provide us with the syllabi we had asked for). On May 20, 2014, oral arguments were heard by the Western District of the Missouri Court of Appeals in our case against the **University of Missouri** on these same issues. A judgment is pending.

9 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.

10 Beginning in the summer of 2011, we would first collect course information about programs at all public institutions in a given state. We would then send out an individualized request to each of the state's programs, asking them again to work with us. If they declined, or did not respond after 10 days, we would follow up with a formal open-records request listing the documents, including the course syllabi, we required.

11 In comparing copies of syllabi that we obtained via campus outreach with those we received directly from programs, we found no instances of counterfeit syllabi.

12 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**, **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.

13 The **Early Reading**, **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.

14 When necessary, the "exceeds variance" trigger was adjusted to be more sensitive and provide additional oversight.

15 For the standards for which only one subject specialist conducted an evaluation (**Early Reading**, **English Language Learners**, **Struggling Readers**, **Elementary Math**), 10 percent of programs were evaluated by two subject specialists to determine the variance rate.

16 *Preparing and Credentialing the Nation's Teachers: The secretary's ninth report on teacher quality*, (2013) p. 37 accessed May 24, 2014 at https://title2.ed.gov/Public/SecReport.aspx

17 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." This is not the case for the weight of the scores associated with strong design in booster standards.

18 Program rankings for special education programs (reported only on NCTQ's website to institutions and not to *U.S. News & World Report*) 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** and **Outcomes Standards**.

19 The relevant content standard may be the **High School Standard** or both the **High School** and **Middle School Standards**. If the latter, the weighting of scores is divided between the two standards, with the **High School Standard** score weighted most heavily.

20 There was also a limited set of programs for evaluation of the **Classroom Management Standard** in the first edition, but all programs for which we had data available were rated on this standard for the second edition.

21 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 "fail" (1 on a 0-4 scale).

22 The one source of evidence that can be submitted in the appeals process that was not used in the original evaluation of standards was end-of-course examinations.

23 Specifically, the algorithm unfairly penalized courses with unclear lectures but moderately strong accountability elements when compared to those with clearly irrelevant lectures and similar accountability.

24 The grade span of certification informs our analysis of whether required math content coursework for elementary teachers sufficiently focused on topics they must know. Courses for teachers getting certified in grades K-8 are presumed to combine topics appropriate for elementary and middle school teachers, rather than simply elementary school teachers.