

Quality in Student Teaching: Flawed Research Leads to Unsound Recommendations

Susan M. Benner
Director, Graduate School of Education
Associate Dean of Professional Licensure
College of Education, Health, and Human Sciences
University of Tennessee, Knoxville
sbenner@utk.edu

Executive Summary

The National Council on Teacher Quality (NCTQ), formed in 2000 to advocate for reform in education, seeks to impact policies associated with teachers and teacher preparation. NCTQ has generated numerous reports on districts/unions, states, and teacher preparation, including several state level reviews of teacher preparation programs. In conducting these reviews and disseminating their reports, NCTQ has created controversy and debate within the teacher education community and beyond.

While NCTQ reports trigger both criticism and applause, none has stirred as much debate as the national review of teacher education programs in partnership with *U. S. News and World Report* that is scheduled for dissemination in 2012. University administrators note that the NCTQ review is limited to inputs such as course syllabi and curricula, not the performance of program graduates. They highlight NCTQ's overreliance on static inputs, their shifting standards, and the absence of information about the credentials of the

persons conducting the review. Supportive superintendents and education commissioners argue that valuable information could be derived from this review.

Within this hostile climate, NCTQ released a report entitled *Student Teaching in the United States*. It focuses on the elementary grades student teaching component of 134 university-based teacher preparation programs. NCTQ gathered and reviewed documents, conducted interviews and surveys of principals, and made four site visits. NCTQ then assigned each program a rating of “model,” “good,” “poor,” or “weak.” It labeled 101 programs as “weak” or “poor.” The commentary offered here provides a critique of the NCTQ student teaching report, which is limited in design, dated in perspective, and flawed in attention to detail. The critique addresses: 1) the conceptual design, 2) the NCTQ standards and their operationalization, 3) the sampling techniques, and 4) the findings and recommendations.

The NCTQ review of student teaching is based upon the assumption that it is not only possible, but also worthwhile and informative to isolate student teaching from the totality of a teacher preparation program. This notion is in direct conflict with the perspective that effective teacher education programs avoid the isolation of pedagogy and classroom management content, offering such knowledge and skills within a learning environment centered upon a clinical experience.

The sample of programs cannot be characterized as representative based on any statistical standard or recognized sampling technique. The problems include disproportionate samples, artificial restrictions, selection bias toward the

weakest programs within universities, lack of clarity regarding sample size, and unsound selection procedures for the sample-within-sample. The problems with data collection include how the ratings were derived, how site visit destinations were selected and how the site visits were used in the data analysis, and how principals were surveyed and/or interviewed.

Limitations in the development and interpretation of the standards, sampling techniques, methodology, and data analysis unfortunately negate any guidance the work could have offered the field and policy makers. However, the fact that this particular review is ill-conceived and poorly executed does not mean that all is well in teacher education. The education of future teachers can be greatly improved by increased selectivity of the students admitted into teacher preparation programs, strengthened clinical experiences woven into the study of teaching and learning, increased demand for teachers to have strong content knowledge and understanding of content-specific instructional strategies, and stricter enforcement of program approval standards.

Quality in Student Teaching: Flawed Research Leads to Unsound Recommendations

Susan M. Benner
Director, Graduate School of Education
Associate Dean of Professional Licensure
College of Education, Health, and Human Sciences
University of Tennessee, Knoxville
sbenner@utk.edu

The National Council on Teacher Quality (NCTQ) is an organization that was formed in 2000 to advocate for reform in education, particularly interested in impacting federal, state, and local policies associated with teachers and teacher preparation programs. Its intended outcome is increasing the number of effective teachers in U. S. public schools. The organization aspires to provide hard evidence derived from its research that can set a reform agenda for policy makers. The NCTQ Board of Directors and Advisory Board believe that recruitment, preparation, retention, and compensation of teachers are all long overdue for reform. NCTQ has generated numerous reports on districts/unions, states, and teacher preparation. For example, in 2006 it completed a national study of reading education programs that involved 72 teacher preparation programs in 35 states. It has also completed several state-specific reviews of teacher preparation programs, including reviews of Illinois and Texas. Recent reports about districts/unions have focused on Springfield, MA Public Schools, Los Angeles, Unified School District, and the Kansas City, Missouri School

District. All of NCTQ's reports can be accessed at the NCTQ website, <http://nctq.org>.

Critics of NCTQ reports have found fault with their methodologies as well as their analyses and conclusions (e.g., Allington, 2006; Eduventures, 2010). For example, Dudley-Marling, Stevens, and Gurn (2007) critiqued the NCTQ review of reading education programs. They noted that the data collection was limited to the gathering of course syllabi via the internet and review of texts used in reading education courses. NCTQ would not consider contacting faculty or interviewing students as a means of learning about reading education teacher preparation, arguing that such approaches might bias their findings. Dudley-Marling, Stevens, and Gurn further critiqued the absence of sound underlying theories of reading upon which NCTQ constructed its review and the numerous assumptions and beliefs implicitly situated in its methods and findings.

While NCTQ reports have often triggered an array of criticisms and applause, none has stirred as much debate as the national review of teacher education programs that is currently underway in partnership with *U. S. News and World Report*. While previous reports have been limited in scope to particular states or content, virtually all university-based teacher education programs are targeted for this review. Many deans of education united to resist it, as is documented in letters from universities, university systems, state councils of deans, and other professional organizations. In these letters, which are posted on the NCTQ website (NCTQ, 2012), deans note that the NCTQ review is limited to inputs such as course syllabi and curricula, not the performance of program graduates available through teacher performance assessments and other

outcomes-focused measures. Letters from the deans of states where NCTQ reviews have already been conducted, such as Illinois, direct sharp criticism to NCTQ's overreliance on static inputs rather than student outcomes, their shifting standards and unwillingness to provide information about the operationalization of standards used in the reviews, and the absence of information about the credentials and experience of the persons employed to review and evaluate universities. There are also letters on the site (NCTQ, 2012) from supportive superintendents and education commissioners who value the information that could be derived from a meaningful review of teacher education programs.

NCTQ has found political allies who have urged university administrators to participate in this review. It has also initiated Freedom of Information requests when cooperation has not been forthcoming. These strategies leave most education deans at public universities with little choice but to participate. However, such strategies will not ease their concerns about the review or encourage them to join with NCTQ in its calls for education reform in teacher preparation.

NCTQ Rates Student Teaching

Within this hostile climate, NCTQ released a report entitled *Student Teaching in the United States* in July 2011 that was authored by Julie Greenberg, Laura Pomerance, and Kate Walsh (2011a, 2011b). NCTQ undertook the task of reviewing student teaching in the U. S. because it understands that the stakes are high when it comes to the student teaching experience. They recognize that, "A mediocre student teaching experience, let alone a disastrous one, can never be undone" (p. 1). An effective teacher has the power to bring a novice into the

profession on a positive trajectory, whereas an ineffective teacher or one unwilling to mentor a novice properly can have a profoundly negative impact on an aspiring teacher. NCTQ used standards for this student teaching report that will be included in the national study of teacher education programs described above.

The report focuses on student teaching associated with the preparation of teachers for elementary grades in 134 teacher preparation programs based at institutions of higher education. NCTQ has assigned each university program a ranking of “model,” “good,” “poor,” or “weak.” It determined the rankings by gathering and reviewing documents, such as course syllabi and handbooks related to student teaching; conducting interviews and surveys of principals; and in a four cases conducting site visits. In some instances, university administrators did not wish to be included in the review, but NCTQ was unwilling to remove them from the sample and worked to gather data with or without their cooperation. In addition to the program ratings, the report contains five findings and two recommendations.

It labeled a total of 101 of the 134 university programs as having a “weak” or “poor” student teaching. Such findings would be cause for concern if they were an accurate reflection of the institutions reviewed or an accurate reflection of the quality of all elementary teacher education programs they are intended to represent. However, careful review yields questions about the institutional ratings, findings, and recommendations presented in the report. The framework NCTQ used to conduct its study is limited in design, dated in perspective, and flawed in attention to detail. The critique below addresses each of the following:

1) the conceptual design, 2) the NCTQ standards and their operationalization, 3) sampling techniques, and 4) findings and recommendations.

Conceptual Design of the Review

NCTQ treats student teaching as though it were something that could and should be isolated from other components of a teacher preparation program. Paradoxically, the authors express regret that they cannot offer any comparisons to superior international student teaching experiences due to the integrated and complex nature of such programs. They rationalize that “it is difficult to learn much from international examples of student teaching arrangements without considering the full continuum of pre-service coursework, fieldwork, and in-service development” (Greenberg, Pomerance, & Walsh, 2011a, p. 11). Yet, the NCTQ review of student teaching is based upon the assumption that it is not only possible, but also worthwhile and informative to isolate the U. S. student teaching experience from the totality of teacher preparation.

NCTQ expects universities to sequence their teacher preparation programs so that all course-work precedes student teaching. The authors believe that Student teaching should not be characterized as a component of teacher preparation that can be done at virtually any point in the preparation process nor [sic] should it be done simultaneously with other coursework. Student teaching is a culminating activity for which methods coursework and practica provide preparation and are the necessary antecedents. (Greenberg, Pomerance, & Walsh, 2011b, p. 27)

Contrast that to the praise they give to Finland's teacher preparation process in which,

...teacher candidates (all of whom are graduate students) engage in a full year of clinical experiences in training schools associated with a university (whose staffs [sic] have proved themselves competent to work with student teachers) serving hundreds of teacher candidates at any one time. (Greenberg, Pomerance, & Walsh, 2011a, p. 11)

The NCTQ standard that student teaching be separated from all other components of the program is in direct conflict with the dominant thinking that effective teacher education programs actually need to minimize the isolation of pedagogy and classroom management courses and create such knowledge and skills within a learning environment centered upon the clinical experience (Darling-Hammond, 2005; Darling-Hammond, 2006; NCATE, 2010; National Research Council, 2010). The National Council for Accreditation of Teacher Education (NCATE) commissioned a Blue Ribbon Panel report, *Transforming Teacher Education through Clinical Practice: A National Strategy to Prepare Effective Teachers*, (NCATE, 2010) that involved the development of a series of briefing papers (all of which are available at www.ncate.org). Drawing from the best research and theoretical perspectives available, members of the NCATE panel developed ten design principles for clinically based preparation. The second of these reads, "Clinical preparation is integrated throughout every facet of teacher education in a dynamic way. The core experience in teacher preparation is clinical practice. Content and pedagogy are woven around clinical

experiences throughout preparation, in course work, in laboratory-based experiences, and in school-embedded practice” (p. 5). There is a clear disconnect between the thinking of NCTQ, that student teaching is a stand-alone culmination of an education, and that of NCATE, in which the development of knowledge and skills are woven together throughout a well-designed teacher preparation program.

Standards and Operationalization

With oversight provided by an advisory group, NCTQ developed 19 standards for its review of student teaching. It considers the standards associated with clinical experiences and student teaching established by NCATE and the Association of Teacher Educators (ATE) to offer insufficient guidance needed to determine program efficacy. In addition, it considers the NCATE and ATE standards to be too broad, because these standards apply to the array of clinical experiences involved in educator preparation, not just student teaching. In the NCTQ report, the authors provide a comparison chart that offers the NCTQ perspective on how their standards align with and stand above those of NCATE and ATE.

NCTQ focused its review on 5 critical standards and an additional 14 less critical standards. All 134 programs involved in the review were assessed against the critical standards, and 32 programs were assessed against all of the standards. NCTQ identified the pool of 32 institutions for review on all 19 of their standards by selecting programs that had either “performed quite well against the first five standards or quite poorly” (Greenberg, Pomerance, & Walsh, 2011b, p. 5).

Critical Standards

STANDARD 1. The student teaching experience, which should last no less than 10 weeks, should require no less than five weeks at a single local school site and represent a full-time commitment.

STANDARD 2. The teacher preparation program must select the cooperating teacher for each student teacher placement.

STANDARD 3. The cooperating teacher candidate must have at least three years of teaching experience.

STANDARD 4. The cooperating teacher candidate must have the capacity to have a positive impact on student learning.

STANDARD 5. The cooperating teacher candidate must have the capacity to mentor an adult, with skills in observation, providing feedback, holding professional conversations and working collaboratively.

Additional Standards

STANDARD 6. Student teaching is part of a rational sequence of coursework that ensures that all methods coursework and practica precede student teaching.

STANDARD 7. Written expectations for competencies on which student teachers will be evaluated are clearly communicated to student teachers, cooperating teachers and supervisors.

STANDARD 8. Written expectations for competencies include the student teacher's analysis of student achievement using informal and formal assessments.

STANDARD 9. The university supervisor should observe the student teacher's delivery of instruction at least five times at regular intervals throughout a semester-long experience.

STANDARD 10. Each observation should be followed by time for conferencing with written feedback aligned with identified competencies.

STANDARD 11. The student teaching experience should include a graded, culminating project that explicitly documents the student teacher's gains on the performance expectations that were communicated at the onset of the experience.

STANDARD 12. Particularly for student teaching during the fall academic term, the schedule for student teaching should align with the elementary school calendar, not the calendar of the teacher preparation program.

STANDARD 13. The student teaching experience should include a gradual increase of student teacher responsibilities, with the student teacher first closely shadowing the cooperating teacher in all professional activities and then transitioning to a more independent instructional role with daily monitoring and feedback. This expectation should be laid out explicitly in guidelines provided to the cooperating teacher, the student teacher and the supervisor.

STANDARD 14. The student teacher should be involved in a full range of instructional and professional activities.

STANDARD 15. The process for selection of the university supervisor should consider the supervisor's instructional knowledge.

STANDARD 16. The university supervisor candidate must have the capacity to mentor an adult, with skills in observation, providing feedback, holding professional conversations and working collaboratively.

STANDARD 17. Cooperating teachers' adequacy should be evaluated by student teachers and university supervisors at the end of each semester. Data from these evaluations should be part of an established and regular review process to ensure that multiple perspectives on the student teaching experience are used to refine it and discontinue placements, if necessary.

STANDARD 18. Schools in which student teachers are placed should be evaluated by student teachers and university supervisors at the end of each semester to determine their functionality—that is, whether the school is high-performing, safe, stable, supportive and collegial. Data from this evaluation should be part of an established and regular review process to ensure that multiple perspectives on the student teaching experience are used to refine it and discontinue placements, if necessary.

STANDARD 19. Recognizing possible geographical constraints, the teacher preparation program should have criteria favoring placement of student teachers in elementary schools in which 1) they have an opportunity to teach children from low-income families and 2) there is an orderly learning environment.

Using the Standards to Produce Program Ratings

NCTQ determined the extent to which each of the 134 programs under review adhered to the 5 critical standards and how the subset of 32 programs aligned with all 19 standards. In the report, each institution received an overall rating based on its performance on Standards 1 through 5. Although there are a few isolated examples of how particular programs were evaluated on individual standards, the report does not provide program ratings for each separate standard other than to note in the Appendix that final ratings for Standards 1 and 6-19 were based exclusively on evidence in documents and that ratings for Standards 2-5 were drawn from documents and principal interviews/surveys. The Appendix does indicate that all programs were rated as either meeting or not meeting each of the 5 critical standards unless there was insufficient data to make such a determination. The smaller sample of 32 programs was rated in this fashion on all 19 of the standards, but the outcomes of these ratings are not included in the report itself or the Appendix. The Appendix offers a partial explanation as to how ratings were assigned and then collapsed into a single, final measure of performance, as noted below.

An institution's ratings on the five key standards were used to broadly categorize the institutions into four groups ("model program," "good," "weak," and "poor"). These groupings were determined primarily by the proportion of Standards 1 through 5 that each institution passed, although Standards 2, 4 and 5 ...were weighted more heavily than Standard 1. (Greenberg, Pomerance, & Walsh, 2011b, p. 11).

Exactly how the rating (i.e., “meets” or “does not meet”) on each of the five critical standards was then aggregated into a summative rating on the 4-point scale is unknown. The authors indicates that when data were insufficient to determine a particular rating on one or more of the standards, “we reported the average grade for the standards that we were able to measure” (2011b, p. 11). Since the rating was either a “meets” or “does not meet” and some of the standards were weighed more heavily than others, it is unclear how such an average was calculated and then translated into a rating on a 4-point scale. Ten programs had missing data that required this additional data manipulation to generate an overall rating.

Appendix F does include a chart that gives the percentage of programs that met each standard. However, the label on this chart indicates that all 134 programs were given a rating on all 19 of the standards. Since only 32 programs were reviewed on all 19 standards, it is impossible to determine how these percentages were generated. Is the label on the chart a typographical error or were ratings given to all programs on the 19 standards?

Operationalizing the Standards

The operationalization of the NCTQ standards is critical to understanding their meaning and value. Standard 1 addresses the minimum length of a student teaching experience. Other than the absence of a research basis for the specific requirement that student teaching last a minimum of 10 weeks, there is nothing unsuitable about this standard. Likewise, Standards 3-5 are reasonable enough and suitable for consideration. However, Standard 2, which is focused on the selection process for cooperating teachers, is ill-conceived and impractical as

worded. If the standard means exactly what it says, it is not to be a collaborative partnership process between the teacher educators and school districts. It is to be at the sole discretion of the teacher educators where student teachers are to be placed. NCTQ distinguishes its cooperating teacher selection standard from those of NCATE and ATE, both of which expect a collaborative, partnership-based approach to the selection of cooperating teachers. Clinical placements require two-way collaboration, mutual respect, and an ongoing willingness to make adjustments. It is critical that teacher educators reject requests for student teachers when contacted by school administrators who “need” student teachers. Perhaps that is the spirit behind the standard, but as it is written, it is not an appropriate standard.

As has been noted in previous critiques of their work, NCTQ shifts its standards around without acknowledging the limitations or flaws in its own work or the possible value found in work done by well-established accrediting agencies inside the education industry. There was difficulty operationalizing Standard 2, necessitating a footnote that reads as follows, “We note that in our evaluation of an institution against Standard 2, we considered whether it plays an active and *informed* role in the selection of every cooperating teacher, basing its selection decision on *substantive information* on the qualifications of teachers” (Greenberg, Pomerance, & Walsh, 2011a, p. 3). A few pages later the authors paraphrased the standard yet another way, stipulating that “programs must actively participate in the selection of cooperating teachers” (Greenberg, Pomerance, & Walsh, 2011a, p. 5). In Appendix F they offer the following description of what it took to meet the standard:

...institutions must play an active and *informed* role in the selection of every cooperating teacher, but this standard does not require that they in any way actively recruit teachers or have any discussions with teachers, independent of their principal, regarding student teaching arrangements. To meet the standard it was only necessary that institutions base their selection decision on *substantive information* about each possible cooperating teacher's qualifications, information beyond the teacher's years of experience or the fact that the teacher had his or her principal's approval for unstated reasons." (2011b, p. 21).

The wording of their own standard is challenging to operationalize and as stated lacks any collaborative intention when compared to those already available from NCATEⁱ or ATE.ⁱⁱ

There is also confusion regarding NCTQ's standards associated with the qualifications of a cooperating teacher. According to the Appendix, programs that had written requirements that stipulated that cooperating teachers possess exemplary instructional skills met their fourth standard. Any requirement that was less specific about instructional capacity or that was not in written documents was deemed unsatisfactory. For example, the requirement that a teacher be a "successful teacher" was not sufficient to meet the standard.

Inaccuracies in Standards Comparisons

In the report, the authors state that NCATE offers no standard for the characteristics for cooperating teachers, declaring "Most notably, NCATE does not indicate any qualifications that the cooperating teacher should possess"

(Greenberg, Pomerance, & Walsh, 2011a, p. 16). Yet, in their chart on page 13, they quote the NCATE standard regarding qualifications of cooperating teachers. It reads, “Clinical faculty (higher education and school faculty) are licensed in the fields that they teach or supervise and are master teachers or well recognized for their competence in their field” (Greenberg, Pomerance, & Walsh, 2011a, p. 13). Although NCTQ might consider these qualifications as insufficient or too vague, it is not accurate to claim that NCATE has no qualifications for cooperating teachers.

Sampling Techniques

The sample of programs cannot be characterized as representative based on any statistical standard or recognized sampling technique. The problems with the sampling techniques include disproportionate samples by state, artificial restrictions placed on the sample, selection bias toward the weakest programs within targeted universities, lack of clarity regarding sample size, and unsound selection procedures for the sample-within-sample that was reviewed on all 19 standards. As is documented below, the sample was one of convenience, with no consistent rationale or pattern of selection.

Disproportionate Samples

NCTQ set a target of three institutions per state and Washington DC for this review with no regard for the size and population within the state or the number of institutions within each state that prepare teachers. Does it make sense to have the same number of participating institutions from Rhode Island and Texas or New York and Montana? In a footnote they note that “Illinois is the only state in which there was an inadvertent oversampling: three public

universities were included” (Greenberg, Pomerance, & Walsh, 2011a, p. 7). Other states, such as Tennessee, ended up with only two institutions involved in the review. Wyoming has one. There is no clarification or explanation of the target number per state or a rationale as to why it was not achieved in all states other than the “inadvertent” oversampling that occurred in Illinois. They also acknowledge that the sample does not represent the distribution of public and private institutions involved in teacher preparation. Whether NCTQ considers this work a review or a research study, their lack of attention to proper sampling techniques certainly calls to question the meaningfulness of any data analysis and generalization of their findings.

Artificial Restrictions Placed on the Sample

NCTQ used a stratified random sampling of institutions for the review, but limited the pool to those offering undergraduate student teaching. Thus, programs designed around newer clinical residency models that are more consistent with international models of teacher preparation were excluded from the sample. While the sample might have 10% of the nation’s institutions offering traditional programs, it does not accurately reflect the state of clinical practice in teacher education in 2011. Having excluded graduate programs from consideration and restricted their focus to elementary education, the authors claim that their findings and recommendations could “extend to both undergraduate and graduate preparation of all classroom teachers” (Greenberg, Pomerance, & Walsh, 2011a, p. 2). Since many graduate-level programs include a residency or internship rather than a student teaching requirement, NCTQ has no basis upon which to make such claims of generalizability of this review.

Selection Biased Toward Weakest Programs in Targeted Universities

NCTQ focused exclusively on programs within each institution that were closest to its minimum standards rather than the largest or the strongest. Tucked into the Appendix, they note, “For all cases, ratings were based on the institutions’ minimum standards, identifying the easiest, fastest or cheapest process allowed by the institution” (Greenberg, Pomerance, & Walsh, 2011b, p. 11). Of all the sampling errors and limitations, this one seems the most telling. Why did NCTQ restrict the sample to the easiest, fastest, cheapest undergraduate programs if it wants to claim its findings to be representative of all undergraduate and graduate teacher education programs based in institutions of higher education? It seems that they were seeking weakness rather than representative state-of-the-art findings.

Lack of Clarity Regarding Sample Size

The authors indicate that 134 schools were included in the review. However, only a fraction of those (32) were actually reviewed on all 19 of their standards. Footnote 43 indicates that even this number of institutions was not always used in the data collection process. It reads, “Evaluation by both student teachers and supervisors of cooperating teachers were only conducted by one-third of the 32 institutions evaluated on the relevant standards” (Greenberg, Pomerance, & Walsh, 2011a, p. 28). Although it is difficult to determine exactly what is meant by this footnote, it is clear that the sample size relevant to this standard is approximately 10; that is,, one-third of 32. Other footnotes throughout the report and Appendix mark inconsistencies in the quality of the sample. For example, footnote 5 reads, “While this proportion of public and

private institutions does not match the distribution in the population of all institutions offering teacher preparation ..., it does not appear to bias results since the average ratings for public and private programs on the five critical standards ... do not differ” (Greenberg, Pomerance, & Walsh, 2011a, p. 3). In the Appendix, the authors acknowledge that they opted to include a greater proportion of public universities because they anticipated needing “to use open records requests as a means to obtain documents necessary for evaluation” (Greenberg, Pomerance, & Walsh, 2011b, p. 7).

Unsound Selection Procedures for the Sample-within-Sample

As noted previously, the authors narrowed the pool of universities for review on all of their standards by selecting some that had been rated as low performing and others that were deemed strong on the first five standards. Did they pick an equal number from each end of this continuum or select a preponderance of programs from those they deemed weak or strong? They do not provide this information. However, in a footnote, they offer an explanation of this decision to draw from the ends of the continuum rather than select programs at varying levels.

We chose institutions at both ends of the continuum on the theory that if a relationship existed between the ratings on Standards 1 through 5 and Standards 6 through 19, we would be most likely to discern it using this stratified sampling approach. In fact, changes in the ratings on Standards 1 through 5 in the course of further evaluation, coupled with a great variation in ratings on Standards 6 through 19, demonstrated that no such relationship existed. In fact,

there was so little evidence of any pattern to ratings that it is safe to presume that we would have obtained the same findings on Standards 6 through 19 by either rating the full sample or by rating any other subset of the full sample. As with the full sample, with the exception of the relative proportions of public and private institutions, this subset of 32 institutions is representative of the population of teacher preparation programs as a whole and can be presumed to paint a representative picture of practices in that population. (Greenberg, Pomerance, & Walsh, 2011b, p. 5)

It is simply illogical to reach the conclusion without any statistical analysis that the absence of a pattern means that the sample-within-sample of 32 was representative of the entire sample of 134 universities. However, they claim that it was not only representative of the entire sample, but also is representative of all university-based teacher preparation programs.

Data Collection and Analysis

As noted above, previous NCTQ reports have been criticized for flawed methodology, particularly data collection techniques. In the student teaching report the problems with data collection include the absence of information about the following: (1) how the ratings were derived , (2) how site visit destinations were selected and how the site visits were used in the data analysis, and (3) how principals were surveyed and/or interviewed.

Deriving the Ratings from the Data

The primary means of data gathering from these institutions involved review of specifically requested documents, including materials such as student

teaching handbooks, course syllabi, and textbooks used in reading education courses by NCTQ employees. Exactly how the NCTQ employees reviewed these documents and assigned ratings to the programs is not fully explained in the report. Appendix F does provide the rationale, the list of criteria associated with the standard (which is labeled methodology), examples of how the standard was applied, and the findings associated with each standard. Further, how much weight is given to documents versus other data sources is unknown, but fluctuates across standards. The authors do indicate in the Appendix that when there were inconsistencies across documents, they opted to use the “more authoritative” (2011b, p. 10) of the documents. They provide an example, noting that a contract between the university and a school district would hold a higher weight than would a student teaching handbook. We do not know if multiple reviewers considered the same documents or if any means of establishing inter-rater reliability was used.

Site Visit Selection and Use in Data Analysis

In addition to gathering documents and conducting interviews and surveys, NCTQ did conduct five site visits as a part of their review of student teaching. One of these site visits was to a program that was not included in the 134 institutions under review, which means that only four of the 134 institutions received site visits. The site visits involved interviews of student teachers, university supervisors, cooperating teachers, and field-placement coordinators. Although the universities that had site visits are named, no information is provided regarding how the four sites were selected, or how any data generated from these visits were integrated into their ratings. The report does indicate that

site visits were useful in affirming the accuracy of information obtained through the review of documents. Some anecdotal stories are included in the Findings and Recommendations sections of the report about the site visits.

Principal Interviews and Surveys

A copy of an online survey that was distributed to principals who were identified by the programs is provided in Appendix B. It is also noted that plans to interview principals were changed when it became too difficult to contact principals by phone. The planned interviews were converted to another, longer online survey. On page 9 of the Appendix, the authors claim that they surveyed 166 principals. It is not clear if this is the number who completed the initial survey, or the longer one that was originally intended to be an interview. Nor is it clear with which universities these principals were connected. The report also notes that 127 principals provided feedback regarding the assignment of cooperating teachers. Throughout the report and Appendix there is no explanation of the discrepancy between these two figures. It is impossible to determine either the source of or representativeness of the results of the principal surveys.

The number of schools/principals needed varies by the size of a teacher education program. Even a small program with only 20 elementary student teacher candidates per year would need between 3 and 5 schools for placements if 4 to 6 students are placed in a single building. For a program with 80 elementary student teachers that range would expand to 13 to 20 principals. With 166 principals drawn from 134 programs, approximately 1.2 principals per program would have been surveyed, assuming equal representation across programs.

Based on the demographics of the target institutions provided on page 6 of the Appendix, we know that 42% produced fewer than 50 graduates per year, 30% produced between 51-150, 16% produced between 151 and 300 graduates, and 5% produced between 301 and 850 graduates per year. While it is impossible to ascertain the precise representativeness of a sample of 166 principals, it is evident that an extremely small percentage of the principals with whom teacher educators from these 134 programs worked participated in the survey. A footnote in the Appendix provides the following detail:

We obtained two or more surveys for 38 of the institutions (28 percent) and these proved to be very helpful to verify ratings on Standards 2 through 5. Many of the interviews that could not be used to verify ratings because they were the only interview available for an institution or because the information provided was inconsistent nonetheless provided valuable information on aspects of student teaching arrangements not directly relevant to ratings.

(Greenberg, Pomerance, & Walsh, 2011b, p. 10)

It can be inferred from this quote that no principals responded to the survey for 96 of the 134 universities under review. The number of participating principals and the percentage of programs to which they are connected falls far short of any measure of representativeness of the programs with which they work, much less the entire field.

Findings and Recommendations

The report contains five findings. In relation to the first finding, the report includes several pages articulating a hypothetical projection of the number of available worthy and willing cooperating teachers. In their scenario NCTQ projects one such teacher in each building of 25 teachers as viable to serve as a cooperating teacher. The logic behind these hypothetical numbers does not match with the reality of securing qualified teachers prepared to mentor and supervise pre-service teachers. It ignores the concept of professional development schools, learning communities, and teacher leadership. It is missing any awareness of newer models of staffing with novice teachers working under the guidance and tutelage of master teachers and curriculum specialists. It ignores models of co-teaching currently being implemented in teacher residency programs and internships. It fails to acknowledge the recent challenge to schools of education to partner with the lowest performing schools to achieve better outcomes for the students in these schools. In essence, this section, although presented as findings, is completely hypothetical and disconnected from the data gathering.

Further, it is in direct conflict with the data obtained from the principals that were surveyed. As just quoted above, the authors claim

Findings
Finding 1: Institutions are routinely exceeding the capacity of school districts to provide a high-quality student teaching experience—and exceeding the demand for new hires.
Finding 2: While the basic structure of many student teaching programs is in place, too many elements are left to chance.
Finding 3: Institutions lack clear, rigorous criteria for the selection of cooperating teachers—either on paper or in practice.
Finding 4: Institutions convey a strong sense of powerlessness in their dealings with school districts.
Finding 5: Institutions do not take advantage of important opportunities to provide guidance and feedback to student teachers.

to have found the principal surveys very helpful in setting ratings for Standards

2-5, related to the selection and qualifications of cooperating teachers. However, they interject into the hypothetical projections the fact that principals surveyed as a part of the student teaching review reported that they accepted an average of six student teachers per building. Using their hypothetical model, the authors then assume that 50% of the cooperating teachers being used are not qualified to serve in such a role and others would rather not be doing it. In a footnote they note that this average was derived from reports provided by 127 principals. That number is less than one principal per institution and again calls into question the adequacy of their sampling and data gathering techniques. Follow their logic as presented in the footnote of the Appendix:

Survey responses indicate that principals, who naturally think well of the capabilities of their staff, are not apt to screen teachers with sufficient rigor when considering their qualifications to be cooperating teachers. Only a tiny fraction of principals (6 percent) estimated that *28 percent or less* [sic] of their staffs [sic] are likely qualified on all measures (experience, effectiveness, and mentoring skills), though our own estimate shows it to be unlikely that there are more than 12 percent in any given school. In contrast, a large minority of principals (41 percent) indicate that *60 percent or more* of their teachers are likely qualified, confirming that teachers who are likely to be less qualified than average are routinely considered qualified to be a cooperating teacher. (Greenberg, Pomerance, & Walsh, 2011b, p. 8)

They acknowledge in this footnote that they have created their own estimates of the availability of qualified cooperating teachers rather than actually drawing this information from the survey data associated with their own work. In essence they discount the opinions of principals whose opinions they sought and accepted on other matters.

The first finding is connected to the hypothetical model that leads to the conclusion that one suitable cooperating teacher can be found per elementary building. It includes the “finding” that teacher education programs are producing teachers at a rate that exceeds the demand for new hires. Although there are national and state workforce reports that support this statement, the authors do not have evidence or data to connect this “finding” to this particular review. There is no indication that they monitored employment patterns of graduates of these programs.

The second finding is abstract and vague, as it reads that “too many elements are left to chance” (Greenberg, Pomerance, & Walsh, 2011a, p. 22). However, the discussion of this finding does include some results that were derived directly from the NCTQ review. For example, the authors note that virtually all of the programs reviewed include full-time student teaching and that most prohibit other coursework during the experience. They also reported that all institutions require a minimum of 10 weeks of student teaching and are aligned with school calendars. They note that one-third of the reviewed programs do not align with the school calendar, resulting in student teachers starting well after the start of the school year. They reported that one-fourth of the 134 institutions allow student teachers to complete student teaching “abroad or in

distant urban areas” (Greenberg, Pomerance, & Walsh, 2011a, p. 24). There is no discussion or clarification of the overall finding as it is stated, or how these results justify the finding as stated.

The third finding, focused on the absence of rigorous criteria for the selection of cooperating teachers, is also directly connected to the review. The authors report that most of the programs had some criteria regarding the selection of cooperating teachers, but these criteria were missing the expectation that these teachers be effective instructors or capable of serving as mentors in sufficient detail to be acceptable.

In the fourth finding, the authors indicate that teacher preparation program faculty have a sense of powerlessness when working with local school districts. The only example drawn from the review of student teaching that is offered to support this finding has to do with the selection of cooperating teachers.

As was the case with the first finding, the fifth finding, that Institutions do not take advantage of important opportunities to provide guidance and feedback to student teachers, is disconnected from the results reported in the review. The discussion of this finding focuses on research cited from a Hamilton Project Discussion Paper associated with the Brookings Institute (Gordon, Kane, & Staiger, 2006), not on data from the 134 institutions under review. The narrative in this section does refer back to the review of student teaching programs and includes some non-examples as well as examples of standards being met.

However, the interpretation of the data must be scrutinized carefully. The report reads, “A significant proportion of institutions (30 percent) fail to require that the

supervisor conduct a conference with the student teacher after each visit and provide written feedback” (Greenberg, Pomerance, & Walsh, 2011a, p. 30). The possibility that a supervisor might make visits for purposes other than conducting a formal or even informal observation/evaluation is missing. Are their conclusions drawn from document review or principal surveys? We have

Recommendations

Recommendation 1: Shrink the pipeline of elementary teachers into the profession.
Recommendation 2: Institutions must make the role of cooperating teacher a more attractive proposition to classroom teachers.

already seen that the principal surveys are far too few to use for such conclusions and the pursuit of such policies in written materials is unlikely since supervisors do not conduct observations for every visit made. We have also already seen that NCTQ accepts or rejects data from the principal surveys to meet its own biases.

The report contains two major recommendations, with a number of corollary recommendations associated with each. The first recommendation is connected to other studies and national statistics, but again has no connection to this review of student teaching. The worthiness of the recommendation can and should be discussed; it simply does not seem to be drawn from the data collected as a part of this review. In the discussion of this recommendation, the report does include a statement that reads, “The low to nonexistent academic bar for entry into all too many teacher preparation programs means that students are accepted who have no serious interest in becoming a teacher and/or who meet no academic standard” (Greenberg, Pomerance, & Walsh, 2011a, p. 36). By design, the review looked exclusively at student teaching. Neither admission criteria nor student demographics are mentioned in the 19 standards that served as the guide

for the review. They have no basis upon which to connect this general statement to the work at hand.

In the report, the authors express dissatisfaction with teacher preparation programs that have extremely high rates of students passing assessments. Although there is no reasonable connection between that sentiment and the present study, there is a real need to give consideration to this matter. State education policy makers are already moving to raise admission standards for teacher education programs and scrutinizing licensure standards as well. Unfortunately, this review provides no meaningful data to guide us in the reform of standards for entry into the teaching profession. However, we can look to the work associated with the Teacher Performance Assessment Consortium and other similar efforts to help raise the entry requirements into the teaching profession.

Conclusion

Quality may be in the NCTQ name; unfortunately it is not a characteristic of its report on teacher training programs. The NCTQ report and the accompanying Executive Summary are full of generalizations purportedly derived from an in-depth review and analysis of 134 student teaching programs that generated thousands of documents. The reality is that they have self-published a report that fails to meet the necessary standards for any such review. Whether NCTQ considers the work research or something less than research, it presents the work as though it met the standards of a research study. The flaws that start at the conceptual level and spread throughout it are overwhelming. Limitations in the development and interpretation of the standards, sampling techniques,

methodology, and data analysis unfortunately negate any possible value the work could have offered the field.

NCTQ has created a four-point scale (weak, poor, good, and model) and identified 67 of the programs reviewed (50%) as weak. Another 34 (25%) were rated as poor. This means that a total of 75% of the programs reviewed fell into the categories of weak or poor, so the review offers no meaningful distinction between 75% of the programs that they claim to have studied in depth.

The failure of the NCTQ report to offer the field and the nation usable information about the quality of our teacher preparation programs is unfortunate. However, the fact that this particular review is ill-conceived and poorly executed does not mean that all is well in teacher education. As has been articulated clearly in *Transforming Teacher Education through Clinical Practice: A National Strategy to Prepare Effective Teachers* (NCATE, 2010), the clinical component of teacher preparation is of great consequence and should be a central element of our programs. We must heed this call to action and ensure that the recommendations put forth in the NCATE report and similarly well-documented calls for higher standards and greater accountability in teacher education become our guide for reform of teacher preparation. Educators must hold one another to the highest possible standards and accept the challenge to strengthen and improve the quality of public education. The reality is that some teacher preparation programs are weak and must either enact radical change or be closed. Unfortunately the NCTQ report on student teaching offers no guidance to the field as to where or how to begin the process. We must respond to calls for reform in teacher preparation with increased selectivity of our students,

strengthened clinical experiences woven into the study of teaching and learning, increased demand for teachers to have strong content knowledge and understanding of content-specific instructional strategies, and stricter enforcement of program approval standards. However, as we engage in policy making and reform initiatives, we must give great caution to the selection of reports upon which we rely to inform this process.

References

- Allington, R. (August 2006). Richard Allington comments on NCTQ. Retrieved December 5, 2011 from <http://www.intersivity.org/lists/epata/archives/Aug2006/msg00013.html>.
- Darling-Hammond, L., (2005). *Powerful Teacher Education: Lessons from Exemplary Programs*. San Francisco: Jossey-Bass.
- Darling-Hammond, L., (2005). *Preparing Teachers for a Changing World: What Teachers Should Learn and Be Able to Do*. San Francisco: Jossey-Bass.
- Dudley-Marling, C., Stevens, L. P., & Gurn, A. (April 2007). A critical policy analysis and response to the report of the National Council on Teacher Quality (NCTQ). Retrieved December 5, 2011 from <http://www.ncte.org/magazine/perspectives/nctqcritique>.
- Eduventures. (2010 September). Review and Critique of NCTQ & Advance Illinois Study on Illinois Teacher Preparation Programs. Retrieved December 5, 2011 from <http://eduventures.com>.

- Gordon, R., Kane, T. J., & Staiger, D. O. (April 2006). *Identifying Effective Teachers Using Performance on the Job (Hamilton Project Discussion Paper)*. Washington, D.C.: Brookings Institution.
- Greenberg, J., Pomerance, L., & Walsh, K. (2011a). *Student Teaching in the United States*. Washington, D. C.: NCTQ. Retrieved October 21, 2011 from <http://www.nctq.org/edschoolreports/studentteaching/>.
- Greenberg, J., Pomerance, L., & Walsh, K. (2011b). *Student Teaching in the United States: Appendix*. Washington, D. C.: NCTQ. Retrieved October 21, 2011 from <http://www.nctq.org/edschoolreports/studentteaching/appendix.jsp>.
- NCATE. (2010). *Transforming Teacher Education through Clinical Practice: A National Strategy to Prepare Effective Teachers*. Washington, D. C.: NCATE. Retrieved October 21, 2011 from <http://ncate.org/>.
- National Research Council (2010). *Preparing Teachers: Building Evidence for Sound Policy*. Washington, D.C.
- National Council for Teacher Quality. (2012). *Building Better Teachers: National Review of Teacher Preparation Programs*. Transparency Central. Retrieved January 4, 2012 from <http://www.nctq.org/transparency.do>.

Footnotes

ⁱ The institution and school partners jointly determine the specific placements of student teachers.

ⁱⁱ School-based teacher educators are collaboratively chosen by campus-based educators and school administrators.