Counter-Intuitive Findings from Teacher Education Accreditation Council's Surveys of Candidates and Faculty about Candidate Knowledge and Skill

Frank Murray
University of Delaware

Introduction

In 1996, the Teacher Education Accreditation Council (TEAC) announced an alternative to the prevailing U.S. practice of accrediting programs by their conformity to consensus standards (see Levine, 2006; Meier, 2000; Murray, 2011a; Ohanion, 1999, 2000 for an analysis of the shortfalls of consensus standards). The TEAC proposal addressed, instead, the program's quality control system and the quality of the evidence that the system yields in terms of the accomplishments of the graduates of the teacher education degree programs (see Dill, Massy, Williams, & Cook, 1996; Ewell, 2008; Graham, Lyman, & Trow, 1995; Trow, 1998 for a discussion of this approach). Contrary to long-standing assertions made by critics (e.g., Aldeman, Carey, Dillon, Miller, & Silva, 2011; Conant, 1963; Crowe, 2010; Judge, Lemosse, Pain, & Sedlak, 1994; Kanstoroom & Finn, 1999; Koerner, 1963; Mitchell & Barth, 1999), TEAC found that program faculties seemed to "actually know what they were doing" and that evidence on which the program faculty rely to support its claim that its graduates are competent was persuasive enough to warrant accreditation and public assurance of the program's quality. A complete list of these accredited programs can be found at www.teac.org, but they are, in the

Frank Murray is the H. Rodney Sharp Professor of Educational Leadership in the College of Education and Human Development at the University of Delaware, Newark, Delaware. His e-mail address is fmurray@udel.edu

main, private liberal arts colleges, flagship research universities, and a few state colleges that were formerly normal schools.

To date, the results from this accrediting work support the following conclusions: Despite beliefs to the contrary, teacher education programs are not "cash cows"; teacher education students are as able in their teaching subjects as arts and sciences majors are in the same subjects; the limited evidence readily available to programs (grades, license scores, ratings by alumni, teachers, and employers) shows uniformly high scores, "widget," or ceiling effects; and the results of clinical and academic evaluations (grades and license scores) are invariably not related to each other, i.e., those high on one may be high, low, or neither on the other (Murray, 2011b; Murray, Raths, & Ramineni, 2006).

Accreditors must be wary of how representative the opinions are of those whom they interview while conducting their on-site verification and corroboration visits. Thus, in 2008, TEAC instituted direct online surveys of students, faculty, and cooperating teachers in regard to adequacy of the graduates' knowledge and skill. The surveys are designed to corroborate the evidence that the program submits and provides in its self-study (called the *Inquiry Brief*) that its graduates are competent in subject matter, pedagogy, teaching skill, multicultural understanding, technology, and independent learning. The adequacy of the courses, faculty, facilities, support services, and institutional commitment also are rated (Murray, 2010; 2011b).

This article is a report of the findings from a sample of approximately 2,700 students and 1,000 faculty in the first 50 TEAC-accredited programs for which the online surveys were used. The sample represents nearly all the full-time faculty members surveyed and approximately 30% of the students. On the common questions in the surveys, the findings from all the cooperating teachers' surveys are indistinguishable from those of the sample (Murray, 2011b). The student sample also is representative of all those surveyed, as their grade point averages (GPAs) were insignificantly different from the GPAs of those reported by the programs about all of their students (as verified by the TEAC auditors on site (Murray, 2010).

Method

Shortly before the accreditation site-visit, the TEAC surveys are sent by email in the name of the program head to students, faculty, and cooperating teachers. The survey contains a series of questions about the adequacy of the program's graduates' knowledge and skill and the adequacy aspects of the program (e.g., courses, facilities, resources, support

services). The TEAC surveys are confidential and essentially anonymous, as TEAC sends the survey electronically to email addresses provided by the program, and they are returned directly to TEAC by a third-party vendor (Zarca) and are not seen by the institution. The program head is asked to provide email addresses for all faculty members, cooperating teachers, and students whose records are cited in the *Inquiry Brief,* as these are the students about which the program had the most evidence of accomplishment (i.e., grades, license scores, and ratings).

The introduction to and directions for the survey are as follows:

> The teacher education program in which you participate is currently being considered for accreditation by the Teacher Education Accreditation Council (TEAC) in Washington, DC. Your candid opinion about the program is particularly valuable and we hope you will take five minutes to respond to the statements in the survey below. Your responses are confidential to TEAC and will not be made available to the program.
>
> For each item please select the word or phrase [in a Likert-formatted table] that best describes your assessment: (1) inadequate, (2) barely adequate, (3) adequate, (4) more than adequate, (5) excellent, or (0) not applicable. You also may add comments that clarify, explain, or elaborate your answers.

The common survey items for students and [faculty] concern the adequacy of the following six attributes of the students' competence:

> 1. Your [or your students'] understanding of your [their] teaching subject matter.
>
> 2. Your [or your students'] understanding of the methods of teaching.
>
> 3. Your [or your students'] ability to teach in a caring manner.
>
> 4. Your [or your students'] ability to teach students who are very different from yourself [themselves] culturally.
>
> 5. Your [or your students'] ability to use educational technology in your [their] teaching.
>
> 6. Your [or your students'] ability to grow professionally by learning things on your [their] own.

In addition, students were asked to rate the adequacy of their subject matter, methods, clinical courses, and faculty; the classroom equipment and supplies; and student support services. The faculty members were asked, in addition, to rate the adequacy of the institution's commitment to the program, the resources available to support their teaching and scholarship, the facilities (classroom equipment, media, and supplies), and student support services.

Results

The results show that students and faculty rated all aspects of their programs in the more than adequate to excellent range (4.00-5.00). Given the large number of raters, however, nearly any mean difference would be statistically significant by any standard parametric test. With three exceptions, they all are. The three exceptions are the mean differences between students' ratings of the adequacy of subject matter and pedagogical faculty, between their ratings of the adequacy of subject matter and pedagogical courses, and between the faculty mean ratings of the adequacy of resources and facilities.

Both the students and the faculty see the adequacy of the students' teaching skills as superior to the adequacy of the knowledge of the subject matter and pedagogy. The source of the superiority, as the data in Table 1 suggest, do not appear to be due to the clinical courses or the clinical faculty, both of which received relatively lower ratings by the students.

Further, lower correlations are found between the adequacy of teaching skill and the adequacy of the clinical faculty and clinical courses

Table 1

Program Students' Ratings of their Own Understanding and the Quality of Courses, Faculty, Facilities, and Student Support Services

Adequacy Topic	Number Students	Minimum Rating	Maximum Rating	Mean Rating	Standard Deviation
Subject Matter Knowledge	2745	1	5	4.44	.75
Pedagogical Knowledge	2707	1	5	4.29	.79
Teaching Skill	2713	1	5	4.68	.60
Multicultural Understanding	2717	1	5	4.37	.78
Knowledge of Technology	2732	1	5	4.26	.87
Capacity to Learn	2737	1	5	4.52	.71
Subject Matter Courses	2654	1	5	4.27	.87
Pedagogy Courses	2689	1	5	4.28	.85
Clinical Courses	2666	1	5	4.09	.97
Subject Matter Faculty	2656	1	5	4.35	.86
Pedagogical Faculty	2691	1	5	4.33	.87
Clinical Faculty	2662	1	5	4.19	.95
Instructional Facilities	2659	1	5	4.06	.94
Student Support Services	2676	1	5	4.09	.94
Grade Point Average	2509	2.00	4.00	3.67	.37

Note. 1 = inadequate, 2 = barely adequate, 3 = adequate, 4 = more than adequate, 5 = excellent.

(Table 2) than between the adequacy of subject matter knowledge or pedagogy and the relevant faculty and courses.

As the correlations in Table 2 show, the students see their own understanding of their teaching subjects and of pedagogy as well as their ability to teach in a caring and effective manner as independent of their overall high grades in the program (3.67/4.0, SD=.37). Their grades also were weakly related to their ratings of the adequacy of the courses and faculty in these areas (r=.07-.08). However, the 724 students in the sample with GPAs of 4.00 rated every survey item, with the exception of classroom adequacy, which they rated significantly lower, and student support services, which showed no difference in the means, significantly higher than did those with lower grades. That the reported grades were correlated significantly, but weakly, with all but two student survey items, demonstrated the expected individual differences among the students' evaluations. The two understandable exceptions were the ratings of the adequacy of the program's facilities and services, which were not correlated with the differences in students' reported grades.

In contrast, the students saw the adequacy of the program faculty and courses (also rated highly at 4.0+/5.0) as somewhat related to their own understanding of each area (mean r=.52 for courses and .44 for faculty). Much stronger, however, are the relationships between the adequacy of the faculty and the adequacy of the courses, which by contrast, are more highly related to each other (r=.70+).

The ratings by faculty of the students' understanding align in all key respects with the students' own ratings of their understanding and skill (Table 3). The faculty members see the institutional commitment to the program and the student support services as more than adequate, but they are significantly less positive about the adequacy of the resources available to them and about the facilities available to the program, rating each as simply adequate.

The faculty gave significantly lower ratings to the students than

Table 2

Correlations of Student Ratings of Their Own Knowledge and Teaching Skills with Their Ratings of the Adequacy of their Courses, Faculty, and GPA (N = 568)

Own Knowledge	Own with Course	Own with Faculty	Own with GPA	Courses with Faculty[a]
Subject Matter	.52**	.46**	.13**	.71**
Pedagogy	.64**	.52**	.10**	.73**
Teaching	.41**	.36**	.08**	.73**

Note. [a]Correlations are between student ratings of the adequacy of the courses and their ratings of the adequacy of the faculty in each area; **$p < .001$.

the students gave themselves with regard to the students' understanding of subject matter, multicultural issues, technology and independent learning, and classroom facilities, but not to student support services, which the students rated significantly lower than did the faculty (t=6.88, df=3553, p< .000). The mean differences between faculty and students in their ratings of the adequacy of their understanding of pedagogy and teaching skill were not statistically significant. This pattern was not universal, however; in 30% of the programs, the faculty gave the students higher ratings than the students gave themselves in regard to their understanding and skill. During the on-site interviews, the respective faculties offered two plausible interpretations of the findings. In the case of higher student evaluations, a supposed natural inclination of students to over-value their abilities was posited. In the case of the faculty's higher ratings, the faculty's more realistic and experienced view of what is needed to be a successful beginning teacher was posited.

Differences among the 50 programs were small, but statistically significant, for the mean ratings of each survey item from each program (student surveys, F(49,2230)=3.47, p=.000; and faculty surveys, F(49,767)=2.52, p=.000). One program was rated below 4.0 overall (and in 9 of 14 survey items) by its students, and six programs were rated below 4.0 overall by the faculty (and in 4-5 of 10 survey items), which supports an interpretation of only adequate quality overall for a few programs in the sample. There were significant differences as well in the mean responses between private and public institutional responders (1,604 students and 598 faculty members in private institutions

Table 3
Program Faculty Ratings Graduates' Understanding, Institution's Commitment to the Program, Resources for Teaching, and Student Support Services

Topic of Rating	Number of Raters	Minimum Rating	Maximum Rating	Mean Rating	Standard Deviation
Subject Matter Knowledge	1001	1	5	4.37	.70
Pedagogical Knowledge	981	1	5	4.33	.71
Teaching Skill	955	3	5	4.66	.56
Multicultural Understanding	977	1	5	4.07	.83
Knowledge of Technology	979	1	5	4.16	.84
Capacity to Learn	1019	1	5	4.37	.74
Institutional Commitment	1036	1	5	4.42	.92
Resources for Teaching	1019	1	5	3.84	1.00
Facilities for Teaching	990	1	5	3.86	.99
Student Support Services	1025	1	5	4.34	.77

and 1,141 students and 403 faculty members in public institutions). The students and faculty at private institutions gave higher ratings in significantly more survey items (13/14 survey items, binomial test, $p = .002$, and 10/10 survey items, binomial test, $p=.002$, respectively). The private institution students also reported significantly higher grades (3.73) than did the public institution students (3.59; $t=8.82$, $df=2507$, $p<.000$).

Cronbach's alphas were .92 for the 14-item student survey and .84 for the 10-item faculty survey, which shows acceptable scales of adequacy with regard to the programs' quality. There is evidence of indiscriminant rating, however, as 7.6% of the students and 3.2% of the faculty gave perfect ratings (5.00) on every item in their respective surveys.

Discussion

These results demonstrate that students and faculty, in contrast to prevailing narratives critical of teacher education (e.g., Levine, 2006; Teacher's College, 2009; University of Virginia, 2009), rate nearly all aspects of their programs in the more than adequate to excellent range (4.0+/5.00). While the ratings suffer from the so-called "widget" effect (Weisberg, Sexton, Mulhern, & Keeling, 2009), namely, uniformly high ratings, they are not simply undifferentiated ceiling effects, as there are also statistically significant differences among most components in the programs. The results, in fact, reveal that faculty and students have a consistent, logical, nuanced, and coherent, albeit inflated, view of the students and the program. Given the inflated means, however, the meaningful information in these results resides to the right of the decimal point because only there is there any variation in the assessments that can be linked to the presumed true variations in student accomplishment.

The high mean ratings in these surveys are consistent with high ratings in similar surveys (e.g., Weisberg et al., 2009). These ceiling effects or rating inflations also may reflect a positive bias to support the accreditation of the program in which the raters participate. Given also that there were small, but statistically significant, differences in the mean ratings across all the institutions and some evidence of indiscriminate rating, it is plausible that there was widespread bias for favorable evaluations in these surveys, as they are associated with a high-stakes accreditation decision. Unlike the faculty and students, however, the cooperating teachers have much less at stake in whether the program is accredited, but they also gave the same high ratings (Murray, 2011b). It was the better-trained cooperating teachers, in fact, and those who understood the program better, who were more satisfied

with the competence of their student teachers and with the program's potential for ensuring the student teachers' successful teaching career (Murray, 2011b). It could have been the other way around, as those who were more aware and better trained could have been expected to have downgraded their ratings of the preparation that students received, had the programs been truly weak. They did just the opposite, however.

The findings cannot be solely explained by rating inflation, either, as there are genuine differences in the ratings of the various survey items. Teaching skill is always rated significantly higher than any other student attribute by all raters and within all programs. Some raters gave minimum ratings (inadequate) for survey items; the standard deviations for the mean ratings are approximately one rating unit. Approximately 2% of the faculty and 3% of the students gave ratings of inadequate or barely adequate overall to their program, with a range of 1% to 11% of these ratings for various survey items. These percentage differences are reflected in the means in Tables 1 and 3 and corroborate the relative weaknesses in the students' multicultural understanding and technology as well as in the clinical courses and faculty. The larger percentages of inadequate or barely adequate ratings are associated with commitment, facilities, resources, and student support services. Obviously, given the variation in the individual ratings to the right of the decimal point, the findings cannot be chalked up solely to ceiling effects or indiscriminate rating.

A puzzling finding, also seen in the results from the cooperating teacher surveys (Murray, 2011b), is that teaching skill was rated significantly higher than the students' knowledge of the subject matter that is being taught and the pedagogical knowledge that presumably undergirds teaching practice. While this seems to suggest that the whole of teaching is greater than the sum of its parts, it also may mean that the indicators of superior teaching skill are not closely linked to subject matter and pedagogical understanding. Still, it is puzzling that all raters, across the board and within each program, feel that the students' teaching ability is superior to their knowledge of the teaching subject or pedagogy, a rare instance in which performance exceeds pre-requisite knowledge for it (not unlike, perhaps, a concert pianist's performance being acclaimed despite missing some notes).

Even if it is conceded that rating-inflation operated in these findings, there were still meaningful differences in the ratings that indicated that, while the students in these accredited programs are not equally competent, the programs do not have all the problems that are commonly alleged. Overall, an overwhelming number of students, faculty, and teachers who participate in these accredited programs expressed

high levels of satisfaction with the quality of the students' knowledge and skill and with the program.

These results contrast with those of the prevailing narratives that teacher education is broken and that today's new teachers are unprepared for their roles (Greenberg, Pomerance, & Walsh, 2011; Teacher's College, 2009; University of Virginia, 2009). While the students in the sample see that their courses and faculty are highly similar in adequacy, it is curious that the adequacy of their own knowledge and skill is relatively less related to the grades that they have earned or to their ratings of the adequacy of their courses or faculty, particularly with regard to the clinical courses and faculty.

It is unusual, as well, and also worth further investigation, that those most familiar with and knowledgeable about a particular teacher education program consistently come to conclusions about the graduates' competence that differ markedly from the conclusions of those who view these same programs at a distance (e.g., Aldeman, et al., 2011; Conant, 1963; Crowe, 2010: Judge et al., 1994; Kanstoroom & Finn, 1999; Koerner, 1963; Mitchell & Barth, 1999). It is perhaps not unlike the predictable annual polls that show that the respondents' own schools, teachers, doctors, congressional representatives, and so forth receive higher grades than all schools, teachers, doctors, representatives, and so forth in the state or nation. This leaves open the question of which is the more accurate picture: the global assessment, given by the critics, or a global assessment comprised only of the sum of the local assessments, such as the one given by this national sample of accredited programs? Accreditation is perhaps a reasonable way to bridge this gap, as it provides both an up-close visit and a review of the evidence with a measure of detached objectivity.

References

Aldeman, C., Carey, K., Dillon, E., Miller, B., & Silva, E. (2011). *A measured approach to improving teacher preparation.* Washington, DC: Education Sector.

Conant, J. B. (1963). *Education of American teachers*. New York: McGraw Hill.

Crowe, E. (2010). *Measuring what matters: A stronger accountability model for teacher education*. Washington, DC: Center for American Progress.

Dill, D., Massy, W., Williams, P., & Cook, C. (1996, September/October). Accreditation & academic quality assurance: Can we get there from here? *Change Magazine*, 17-24.

Ewell, P. (2008). *U.S. accreditation and the future of quality assurance.* Washington, DC: Council of Higher Education Accreditation.

Graham, P., Lyman, R., & Trow, M. (1995). *Accountability of colleges and uni-*

versities: An essay. New York: Columbia University.
Greenberg, J., Pomerance, L., & Walsh, K. (2011). *Student teaching in the United States*. Washington, DC: National Center on Teaching Quality.
Judge, H., Lemosse, M., Paine, M., & Sedlak, M. (1994). The university and the teachers. *Oxford Studies in Comparative Education, 4*(1 & 2).
Kanstoroom, M., & Finn, C. (1999). *Better teachers, better schools.* Washington, DC: The Thomas Fordham Foundation.
Koerner, J. D. (1963). *The miseducation of American teachers*. Boston, MA: Houghton Mifflin.
Levine, A. (2006) *Educating school teachers.* Washington, DC: The Education Schools Project.
Meier, D. (2000). *Will standards save public education?* Boston, MA: Beacon Press.
Mitchell, R., & Barth, P. (1999). How teacher licensing tests fall short [Newsletter]. *Thinking K-16*, 3(1).
Murray, F. (2010). Lessons from ten years of TEAC's accrediting activity. *Issues in Teacher Education, 19*(1), 7-19.
Murray, F., Raths, J., & Ramineni, C. (2006, April). The clinical disconnect in teacher education. Paper presented at the American Educational Research Association Annual Meeting, San Francisco, CA.
Murray, F. (2011a). Evidentiary standards for consensus standards in teacher education. *Teacher Education & Practice, 24*(4), 19-23.
Murray F. (2011b). Findings from ten years of accrediting activity by the Teacher Education Accreditation Council in the United States. In J. O'Meara (Ed.), *International yearbook on teacher education: Teachers and teacher education* [DVD]. Glasgow, Scotland: International Council on Education for Teaching.
Ohanian, S. (1999). *One size fits few: The folly of educational standards.* New York: Heinemann.
Ohanian, S. (2000, January). Goals 2000: What's in a name? *Phi Delta Kappan*, 345-355.
Teachers' College. (2009). *Teacher preparation: Reforming the uncertain profession.* Retrieved from http://www.edgovblogs/duncan/2009/10/a-call-to-teach
Trow, M. (1998). On the accountability of higher education in the United States. In W. Bowen & H. Shapiro (Eds.), *Universities and their leadership* (pp. 15-61). Princeton, NJ: Princeton University Press.
University of Virginia. (2009). *A call to teach.* Retrieved from http://www.tc.columbia.edu/news/article.htm?id=7179
Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness.* Washington. DC: New Teachers Project.