Gainsburg, J., Scheld, S., von Mayrhauser, C., Rothstein-Fisch, C. and Spagna, M.
AMTE Monograph 6
Scholarly Practices and Inquiry in the Preparation of Mathematics Teachers

7

One Credential Program Looks in the Mirror

Julie Gainsburg, Suzanne Scheld, Christina von Mayrhauser, Carrie Rothstein-Fisch, and Michael Spagna
California State University, Northridge

This article describes the work of an interdepartmental research team at California State University, Northridge, to begin to systematically evaluate the impact of its teaching credential program. First- and second-year teachers who had graduated from our secondary mathematics program were observed and interviewed to determine the degree to which they implemented program-emphasized teaching practices in their classrooms and the factors—internal and external to the program—that facilitated or impeded the implementation of those practices.

Introduction: *Teachers for a New Era*

In 2002, the Carnegie Corporation, with the Annenberg and Ford Foundations, launched the *Teachers for a New Era* (TNE) Initiative, with the goals of catalyzing reform in teacher education, consolidating knowledge about excellence in teacher preparation, and strengthening public confidence in university teacher education programs (UTEPs). In 2002, four universities, including California State University, Northridge (CSUN), were invited to join TNE and awarded $5 million each over five years; another seven joined in 2003. Our charges were:

1. To make teacher education a university-wide effort, specifically engaging arts and sciences faculty in improving subject-matter learning for potential teachers;
2. To strengthen the clinical-practice experience, by intensifying partnerships and the theoretical alignment between the university and K–12 schools;
3. To ground continual program improvement in evidence of the impact of programs on the teaching practices of graduates and the learning of their pupils.

This last charge also addresses the goal of public confidence: If UTEPs can establish systematic processes for gathering evidence of their impact on teachers and pupils, they become, for the first time, positioned to demonstrate their value to the enterprise of teacher development.

TNE efforts at CSUN have involved over 200 people through multiple projects addressing the charges. These include faculty members and administrators at CSUN and area community colleges, local K–12 school personnel, preservice teachers, and area business people. This article describes the work of one committee, tasked with the third charge, to systematize the collection of evidence of programmatic impact. In the 2006–2007 school year, we conducted an evaluation study of the secondary-level mathematics-credential program. The main purpose of this article is not to convey our particular results, which may be of limited relevance to other programs, but to explain the study's design and rationale and what we learned, to inform other programs interested in conducting evidence-based self-evaluations.

The Challenges of Evaluating a Program by the Performance of its Graduates

A teacher educator need only consider for a few minutes the prospect of evaluating a credential program on the basis of its graduates' teaching practices before problems become apparent. The few listed here weighed heavily on our team and shaped our study's design.

1) *Scarce resources.* As a public, regional, masters university, time and money for research is limited and we have no doctoral students to assist. TNE provided initial funding, but we hoped to design an ongoing evaluation that could be sustained with normal university resources.
2) *What to look for.* It was obvious that we would need to observe graduates "in the field," teaching in classrooms. It was not obvious what to look for that would convince not only department members but also the broader public of our program's impact.
3) *How to discern program impact.* A host of factors shape what teachers do in the classroom. Some arise on the job, including curricular mandates, professional development and mentoring, characteristics of pupils and their families, and available resources. These can support or compete with factors teachers bring to the job: personal philosophies and school experiences; knowledge, skills, and dispositions gained in UTEPs; and personal or family situations. How would we know whether to attribute certain teaching practices to our credential program?

These challenges make clear why few UTEPs have engaged in this sort of evaluation. In our study, we reached partial resolution to these problems but cannot claim to have solved them.

Study Design

The core study team comprised two faculty members in the D~~d~~epartment of A~~a~~nthropology and three in the College of Education. Only the first author was in S~~s~~econdary E~~e~~ducation, the department that housed the program being evaluated. We also employed two research assistants. The core team designed the study to serve two main purposes: We aimed to provide specific evidence for the D~~d~~epartment of S~~s~~econdary E~~e~~ducation about its graduates' teaching, to inform program improvement; and to demonstrate that an "in-house" evaluation of a credential program, based on evidence from the field, was possible, and to present one method.

Our research questions were:

- To what degree do new credential-program graduates implement program-emphasized teaching practices in their classrooms?
- What factors facilitate or impede the implementation of program-emphasized practices?
- Is the program improving over the years in terms of this implementation?

Participants

An early and crucial decision was how many graduates to include in the study and how often to observe them. With limited resources, this was a breadth-versus-depth decision. We opted for breadth, observing as many participants as we could, once or twice each, because we were interested in the impact on *all* graduates, and because we were anxious to convey an evaluation of the program, not of individual teachers. While we understood that no single teacher's practices would be fairly captured in one or two observations, we hoped the large number of "snapshots" would paint a reasonably accurate picture of our graduates in aggregate.

Another key decision was which graduates. New teachers can be overwhelmed by the challenges of classroom and time management and becoming familiar with school policies and facilities. These challenges may prevent them from utilizing practices they have learned in a UTEP, would like to use, and may use in later years (Richardson, 1996). Observing teachers with more experience might give a truer picture of how they "really" teach, but it would be harder to credit our program for what we observed, because of intervening professional development experiences in the years since graduation. Our ultimate decision to observe first- and second-year teachers

had two advantages: It is easier to track down more recent graduates, and they are likelier to still be teaching. We invited all 31 first- and second-year graduates of our traditional single-subject mathematics-credential program to participate. Of the 27 we were able to contact, 21 were fulltime mathematics teachers. Twelve agreed to participate, but we eliminated two because of distance. Thus, ten teachers were observed and interviewed at least once; six were observed twice, for a total of 16 observations/interviews. We secured permission from each participating teacher's principal; thus, it was important to disguise vignettes or quotes in reported results. Participants were paid $50 for each observation/interview.

Observations and Interview

We made two kinds of observations in the classroom. We recorded instances of the teacher's use of specific practices that the program emphasized; how these were designated is discussed later. Because these designated practices were exemplary and somewhat advanced, we did not expect to see them used frequently, and by documenting only them we would risk missing most of what the teachers did during the lessons. Therefore, we decided also to keep a running record of the teaching/learning modes, noting the clock time of each mode change. We developed a manageable list of modes to capture critical classroom-environment qualities known to impact learning: individual versus collaborative work, teacher versus pupil talk, and teacher presentation versus pupil discovery or idea generation (e.g., Boaler, 1997; Hiebert et al., 1997). The final list of modes was:

1) Teacher presentation of behavioral directives or task instructions
2) Teacher presentation of math content (minimal or no pupil input)
3) Pupil demonstration (more significant than verbal answer from seat)
4) Whole-class discussion to review/apply learned procedures (significant pupil input)
5) Whole-class discussion to co-construct new concepts or procedures
6) Pair or small-group work practice of learned procedures
7) Pair or small-group concept development (e.g., discovery activity/lab/project)
8) Individual practice of learned procedures
9) Individual concept development (e.g., discovery activity/lab/ project)
10) Activity or discussion unrelated to math or implementation of the lesson.

Designating the specific practices to observe for was more problematic. One option was to derive a list from the research literature about effective mathematics-teaching practices. But this might have clouded our view of our program's impact, because not all of those practices would necessarily have

been emphasized in our program, due to limited program duration and the beginning level of our students. Similarly, existing observation protocols, such as the Reformed Teaching Observation Protocol (Arizona Collaborative for Excellence in the Preparation of Teachers, 2000) and the National Survey of Science and Mathematics Education Mathematics Questionnaire (Horizon Research, Inc., 2000), were far more extensive than our resources would support and assessed more practices than appropriate for new graduates. Instead, we convened a group of faculty members who taught in the math-credential program to generate a list of practices they believed were emphasized in courses and fieldwork. They focused on research-based practices that seemed reasonable to expect from a new teacher and could validly be observed in one lesson; hence, they excluded assessment practices. From this list, the core team developed an observation protocol, combining like practices, eliminating less significant ones, to attain a manageable number. Published protocols were referenced; indeed, all items on our protocol have close counterparts in other protocols. But for the sake of program evaluation and faculty buy-in, it was important that our decision to include each practice was based on faculty agreement that it was emphasized in our program.

Table 1 lists the practices as described in the observation protocol and gives example citations for scholarship linking the practice to enhanced pupil learning. During the lesson, the observer(s) described in field notes any instances of these designated practices. Afterwards, the observer(s) rated each practice as being a significant, a marginal, or no part of the overall lesson. These ratings were holistic, with the marginal rating used for lessons where instances of the practice were very short in total duration (e.g., one practice problem permitting a choice of method) and/or superficial (e.g., a teacher's side remark about a real-world connection).

Table 1
Practices as Listed on the Observation Protocol, with Examples of Supporting Scholarship

Practice	Example of supporting scholarship
1) Teacher asks question or poses task with a **high level** of cognitive demand	Kilpatrick, Swafford, and Findell (2001)
2) Pupil given **authority to judge** the mathematical soundness of publicly presented solution or method	Hiebert et al. (1997)
3) Teacher **connects** (or poses task that prompts pupil to connect) the featured math topic to **another math** topic	Hiebert and Carpenter (1992)

Practice	Example of supporting scholarship
4) Teacher **connects** (or poses task that prompts pupil to connect) the featured math topic to **another academic** topic	Lehrer, Schauble, Strom, and Pligge (2001)
5) Teacher **connects** (or poses task that prompts pupil to connect) the featured math topic to a **real-life** situation or object	Boaler (1997)
6) Pupils allowed or encouraged to **choose** among solving **methods** or present alternative methods	Bransford, Brown, and Cocking (1999)
7) Teacher or pupils use **technology, manipulatives**, body movement, or other nonverbal support for a math concept	Kilpatrick et al. (2001)
8) Specific attention paid to **language**, i.e., writing, reading, or speaking skills	Borasi and Siegel (2000).
9) Teacher uses or encourages pupil to use multiple **forms of representation** for the same problem	Bransford et al. (1999)

The designated practices are not basic teaching methods but reform-oriented, nontraditional strategies that are difficult to carry out effectively (Tarr, Reys, Reys, & Chavez, 2008) and probably require more intensive training than can be accomplished in a one year post-baccalaureate credential program. Therefore, we decided not to try to assess the quality of their implementation or their impact on pupil learning, even though the latter was part of the third TNE charge. We felt any level of implementation of these practices by new teachers would be notable, but that it would be unreasonable to expect that implementation to result in observable pupil outcomes.

Observations were scheduled with the teacher. We requested a first-year algebra class or the closest course the teacher taught, to attain some commensurability across participants. Also, Algebra I is of great concern in California: required for a high school diploma and with a high failure rate, it has become the “bottleneck” course. We asked only for a typical lesson—nothing special—and that there not be a test that period. Surprise visits were not an option; the schools would not have permitted them, and we wanted to offer participants some agency over the process. We were unconcerned about teachers “putting on a show.” Indeed, if a teacher were to perform atypically, it would not entirely represent a “false positive.” If a recent graduate could anticipate what practices CSUN researchers would hope to see (we did not tell participants we were targeting specific practices), understood how these differed from her usual methods, and were able to incorporate them into a lesson, we could justifiably interpret this as a form of programmatic impact.

It would at least indicate that: a) the teacher had learned what CSUN considered good teaching practices, b) she was able to implement them, c) the stretch to incorporate them into her normal classroom environment was not so far as to be impossible, and d) she was disposed to trying new practices. On the other hand, with the literature on teacher change showing strong resistance even among veteran teachers to adopt new practices, despite significant professional development and explicit encouragement from administrators (Clarke, 1994), it seems unlikely that a new teacher would be willing or able to drastically alter her performance for a researcher who has not explicitly specified a "good performance."

Prior to the observation, each teacher filled out a form that asked for background information (education, experience, and languages spoken) and about the lesson and class to be observed (demographics, lesson objectives, and any special features or conditions). After the observation, on the same day, we conducted a semi-structured interview with each participant to try to understand why we did or did not see the designated practices implemented during the lesson. Grossman, Smagorinsky, and Valencia (1999), in their framework for studying teaching, laid out multiple variables to consider in explaining new teachers' practice appropriation—more than a single study could attend to, especially a necessarily resource-lean program evaluation. Yet we knew it was crucial to examine contextual variables and not presume that whether and how our graduates appropriated program-emphasized practices were solely functions of the quality of our instruction. In particular, we inquired about the impact of teaching settings (collegial and administrative support, demographics, professional development, material resources, etc.), personal background (motivations for teaching and prior experiences), and personal interpretations and affective impressions of our UTEP. Examining these from the teachers' perspective aligns with the stance of Grossman et al. that settings are personally construed and that a teacher's interpretation of these variables determines her motives and impacts practice.

Training

Training our team to use the observation protocol was accomplished with videotaped lessons, adapting calibration methods from large-scale classroom-observation studies such as the TIMSS Video Study (National Center for Education Statistics, 2003). We all viewed the same portion of a video and noted instances of the designated practices and teaching modes. Then we discussed our observations until we reached consensus about whether each instance indeed exemplified the practice or mode. We repeated this exercise several times, also using the first few sessions to refine the protocol and our list of examples of each practice and mode, until we were consistently reaching a high level of agreement (about 6 video segments). Calibration was further achieved by sending observers in pairs for their first two observations;

afterwards, the pair reached consensus about modes and significance ratings. During the fall and spring that observations and interviews occurred, the team met periodically to discuss uncertain ratings, refine the observation protocol's example bank, and discuss emerging trends in our findings.

Data Analysis

Analysis of the quantitative data was relatively simple: We totaled the time spent in each teaching mode and calculated each as a percent of the total time, to produce an "average" lesson profile for our graduates. For each designated practice, we reported the percent of observed lessons that employed it at each significance level. The qualitative data were analyzed three ways. First, interview data were read for factors that supported or constrained teaching, some of which had been anticipated and built into the interview prompts, others of which emerged, following a grounded theory approach (Creswell, 1998). When multiple readers had agreed on a coding scheme of factors, the data were reread and coded, and numbers of teachers citing each factor were tallied. Second, case studies (Creswell, 1998) were developed for 6 of the teachers, selected to represent a range of use of the designated practices. Drawing from the interviews, observations, and information forms, these interpretive narratives were intended to provide holistic illustrations of how the various factors interacted to influence teaching. Finally, the 10 teachers were categorized by level of practice implementation, and all data were reanalyzed for differential patterns of influential factors across (relatively) low, medium, or high implementers.

Findings

In general, we found our graduates implementing the designated teaching practices to a low degree. The most frequently observed of these practices was giving pupils the authority to judge the mathematical soundness of a solution or method—a significant part of 25% of the lessons. Posing tasks with a high level of cognitive demand was significant in 19% of the lessons. These two practices played at least a marginal role in half of the lessons, as did the use of technology or manipulatives. All other designated practices were observed less often.

The modes employed in our graduates' "average" mathematics lesson reflected relatively traditional teaching. For nearly 1/3 of the duration of this average lesson, pupils practiced learned procedures ("seatwork"), sometimes being allowed but not explicitly encouraged to consult with peers. For nearly 1/4 of the lesson, teachers presented mathematical content, taking minimal or no pupil input. Teachers gave behavioral directives or task instructions for 12% of lesson time. All other modes were each used less than 10% of the time.

We had anticipated this picture of our graduates' first and second years of teaching. Other, larger-scale studies (e.g., Jacobs et al., 2006; Weiss, Pasley, Smith, Banilower, & Heck, 2003) have documented the low degree of implementation of reform practices in mathematics classroom by teachers at *all* levels of experience. New teachers may be particularly apt to fall back on traditional methods, with preservice teacher education constituting too weak an intervention to overcome teachers' experiences as students in traditional classrooms (Ball & Cohen, 1999) or the pressure to align with the culture of their initial job placement (Grossman et al., 1999). We view these as baseline data, with the real value of the study coming in its continuation over the years, as the results, ideally, feed into and reflect program improvement. Further, if specific program improvements are followed by observations of increased frequency of related practices, it will arguably be evidence that UTEPs impact teaching. Of special interest to us is the effect of a significant program change: Since 2006-07, we have instituted a capstone performance assessment that requires credential candidates to reflect extensively on a unit they plan and teach, including on a videotaped lesson, pupil work samples, and their efforts to develop pupils' academic language.

Also important findings are the factors that participants cited to explain how they teach. We gained specific feedback about aspects of our program and how our graduates, now one or two years into teaching, saw their value. What participants perceived as supports (mainly collegial and administrative support) and constraints (mainly managerial tasks and limited time) at the school site echoed existing research (e.g., Gold, 1996; Darling-Hammond, 1990) about the experience of new teachers.

Repeating the Study

This study demonstrated that an evidence-based self-evaluation of a credential program is possible and can be carried out with minimal resources. In 2008–09, we are repeating the study with a few changes. First, because second observations/interviews of the same teacher appeared to have little effect on the overall results, we will only visit each teacher once but double the pool of potential participants, now targeting graduates in their first through *fourth* years of teaching, thereby revisiting the 2006–07 participants. We have also modified the interviews to yield information more easily interpreted in terms of programmatic improvement. In 2006–07, teachers were asked about constraints and supports for their teaching in general, leaving us unsure whether time pressure, for example, specifically constrained the use of any designated practices or was just an overall impediment. Questions for 2008–09 more directly target factors that influence particularly the implementation of the practices. Other changes reflect the need to reduce the cost of the study to make it feasible on an ongoing basis. The observation

protocol and calibration training procedures are unchanged, but only two researchers will conduct the study. Data analysis will also remain the same except that no further case studies will be developed. Our focus will continue to be on influential factors in the teachers' UTEP preparation and in their current teaching situation, to help the department understand how to more effectively promote good teaching practices and prepare new teachers to navigate constraints in the field.

References

Arizona Collaborative for Excellence in the Preparation of Teachers. (2000). Reformed Teaching Observation Protocol. Retrieved ~~December 30, 2008~~August 21, 2009, from http://cresmet.asu.edu/instruments/RTOP/index.shtml ~~http://cresmet.asu.edu/prods/rtop. shtml~~

Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 3–32). San Francisco: Jossey-Bass.

Boaler, J. (1997). *Experiencing school mathematics: Teaching styles, sex, and setting.* Philadelphia: Open University Press.

Borasi, R., & Siegel, M. (2000). *Reading counts: Expanding the role of reading in the mathematics classroom.* New York: Teachers College Press.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.) (1999). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

Clarke, D. (1994). Ten key principles from research for the professional development of mathematics teachers. In D. B. Aichele & A. F. Coxford (Eds.), *Professional development for teachers of mathematics: 1994 Yearbook* (pp. 37–48). Reston, VA: National Council of Teachers of Mathematics.

Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions.* Thousand Oaks, CA: Sage Publications.

Darling-Hammond, L. (1990). Teachers and teaching: Signs of a changing profession. In W. R. Houston (Ed.), *Handbook of research on teacher education* (pp. 267–290). New York: Macmillan.

Gold, Y. (1996). Beginning teacher support: Attrition, mentoring, and induction. In J. Sikula, T. J. Buttery & E. Guyton (Eds.), *Handbook of research on teacher education* (2nd ed.; pp. 548–594). New York: Simon & Schuster Macmillan.

Grossman, P. L., Smagorinski, P., & Valencia, S. W. (1999). Appropriating

tools for teaching English: A theoretical framework for research on learning to teach. *American Journal of Education, 108*, 1–29.

Hiebert, J., & Carpenter, T. P. (1992) Learning and teaching with understanding. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 65–97). New York: Macmillan.

Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K. C., Wearne, D., Murray, H., Olivier, A., & Human, P. (1997). *Making sense: Teaching and learning mathematics with understanding.* Portsmouth, NH: Heinemann.

Horizon Research, Inc. (2000). 2000 National Survey of Science and Mathematics Education Mathematics Questionnaire. Retrieved December 30, 2008, from http://2000survey.horizon-research.com.

Jacobs, J., Hiebert, J., Givvin, K., Hollingsworth, H., Garnier, H., & Wearne, D. (2006). Does eighth-grade mathematics teaching in the United States align with the NCTM Standards? Results from the TIMSS 1999 video studies. *Journal for Research in Mathematics Education, 37*(1), 5–32.

Kilpatrick, J., Swafford, J., & Findell. B. (2001). *Adding it up: Helping children learn mathematics.* Washington, DC: National Academy Press.

Lehrer, R., Schauble, L., Strom, D, & Pligge, M. (2001). Similarity of form and substance: From inscriptions to models. In D. Klahr & S. Carver (Eds.), *Cognition and instruction: 25 years of progress* (pp. 39–74). Mahwah, NJ: Lawrence Erlbaum Associates.

National Center for Education Statistics. (2003). *Teaching mathematics in seven countries: Results from the TIMSS 1999 Video Study,* (NCES 2003–013 Revised). Washington, DC: U.S. Department of Education.

Richardson, V. (1996). The role of attitudes and beliefs in learning to teach. In J. Sikula, T. J. Buttery & E. Guyton (Eds.), *Handbook of research on teacher education* (2nd ed.; pp. 102–119). New York: Simon & Schuster Macmillan.

Tarr, J. E., Reys, R. E., Reys, B. J., & Chavez, O. (2008). The impact of middle-grades mathematics curricula and the classroom learning environment of student achievement. *Journal for Research in Mathematics Education, 39*(3), 247–280.

Weiss, I. R., Pasley, J. D., Smith, S. P., Banilower, E. R., & Heck, D. J. (2003). *Looking inside the classroom: A study of K–12 mathematics and science education in the United States.* Chapel Hill, NC: Horizon Research, Inc.

Julie Gainsburg

Carrie Rothstein-Fisch

Suzanne Scheld

Michael Spagna

Christina von Mayrhauser