

***Results of a Study of a Teaching-Credential
Program's Impact on Recent Graduates:***

SUB-REPORT 1, BEHAVIORAL OBSERVATIONS

*Julie Gainsburg, Suzanne Scheld, Christina von Mayrhauser,
Michael Spagna, and Carrie Rothstein-Fisch*

Version 4/8/08

This report, the first of two sub-reports, presents the study's behavioral observation data and situates this data against the backdrop of national trendlines.

The report is followed by a second sub-report using anthropological triangulation of interview, survey and fieldnote data to make sense of the behavioral observations.

Teachers for a New Era Evidence Committee:

**Michael Spagna, Julie Gainsburg, Suzanne Scheld, Christina von Mayrhauser,
Carrie Rothstein-Fisch, Marian Pasternak, and Peter Marshall**

OVERVIEW

During the 2006-07 school year, the TNE Evidence Committee conducted an evaluative study of the impact of a Department of Secondary Education (SED) teaching-credential program. This report, the first of two sub-reports, presents the study's behavioral observation data and situates this data against the backdrop of national trendlines. Sub-report 1 will be followed by a second sub-report that uses anthropological triangulation of interview, survey and fieldnote data to make sense of the behavioral observations.

The specifics:

Research Questions

- To what degree do new credential-program graduates implement CSUN-identified effective teaching practices in their classrooms? (Sub-report 1 focuses on this question.)
- What factors facilitate or impede implementation of pedagogies taught in the credential program? (Sub-report 2 focuses on this question.)
- Future question: Is the program improving over the years in terms of this implementation?

Participants

- All 1st and 2nd-year fulltime secondary math teachers who earned CSUN Traditional Single-Subject Math credentials from fall 2004 to spring 2006, and who agreed to be observed (ten teachers)

Methods

- Collaborative development with key SED personnel of a structured protocol for classroom observations (the “tool”) to detect the presence/significance of research-based “effective practices” stressed in the single-subject mathematics credential program
- Collection of written background information from each participant prior to observation
- One period-long observation of each participant, using the “tool”
- Audiotaped interview of each participant immediately after each observed lesson (see Sub-report 2 for details on how interview was constructed, administered and analyzed)
- A second observation and interview of some participants as available (six teachers). (See Sub-report 2 for details on this second interview.)

Data Analysis

- Percentage of time spent per teaching mode total (i.e., breakdown of the average lesson)
- Significance of each effective practice in the lessons overall
- Qualitative analysis of interview data to identify factors facilitating and impeding the practices (see Sub-report 2 for details of this analysis)
- Development of case studies to further explore facilitating and impeding factors (see Sub-report 2 for presentation and analysis of these case studies.)

Our vision is that this study will grow into a sustainable research program, encompassing multiple credentialing programs and departments, with results feeding back systematically to improve programs and assessment. Recent graduates will be observed annually—ideally the same teachers in multiple years to detect growth. The program will be part of the greater teacher-education assessment initiative at the college and university levels, and will contribute to (and make use of) the “data warehouse.”

RESULTS SUMMARY FOR SUB-REPORT 1, BEHAVIORAL OBSERVATIONS

Who participated?

- The prospective pool comprised the 31 graduates who earned the traditional, single-subject math credential between fall 2004 and spring 2006.
- We were able to find and communicate with 27.
- 21 of the 27 were fulltime secondary math teachers at the time of the study.
- 12 of the 21 agreed to participate. We eliminated 2 because of distance.
- 10 teachers were observed and interviewed at least once.
- We conduct second observations of 6 of teachers, for a total of 16 observations/interviews.

Where, what, and whom did they teach?

- The 10 participants taught in 10 different schools (5 middle and 5 high schools) in 7 districts.

What does the “average” lesson look like?

Method

We recorded the amount of time in each lesson that was devoted to each of 11 teaching modes:

TI	teacher presentation of behavioral directives or task instructions
TM	teacher presentation of math content (minimal or no pupil input)
SD	pupil demonstration (more significant than verbal answer from seat)
WP	whole-class discussion to review/apply learned procedures (significant pupil input)
WC	whole-class discussion to co-construct new concepts or procedures
GP	pair or small-group work practice of learned procedures
GC	pair or small-group concept development (e.g., discovery activity/lab/project)
IP	individual practice of learned procedures
IC	individual concept development (e.g., discovery activity/lab/project)
T/Q	test/quiz
U	activity or discussion unrelated to math or implementation of the lesson
O	other

We summed these times across all 16 lessons and found the percent of time each mode contributed to the total. In this way, we arrived at a breakdown of an “average” lesson for the 10 teachers (Figure 1).

Highlights

- For nearly 1/3 of the duration of the average math lesson taught by our graduates, pupils individually practice learned procedures, i.e., do “seatwork.” In this mode, teachers sometimes allow but do not explicitly encourage pupils to consult with peers.
- For nearly 1/4 of the duration of the average math lesson, the teacher presents math content, taking minimal or no input from pupils. This can take the form of lecture, but it more often includes teachers asking questions of pupils that require short answers that recall taught content.

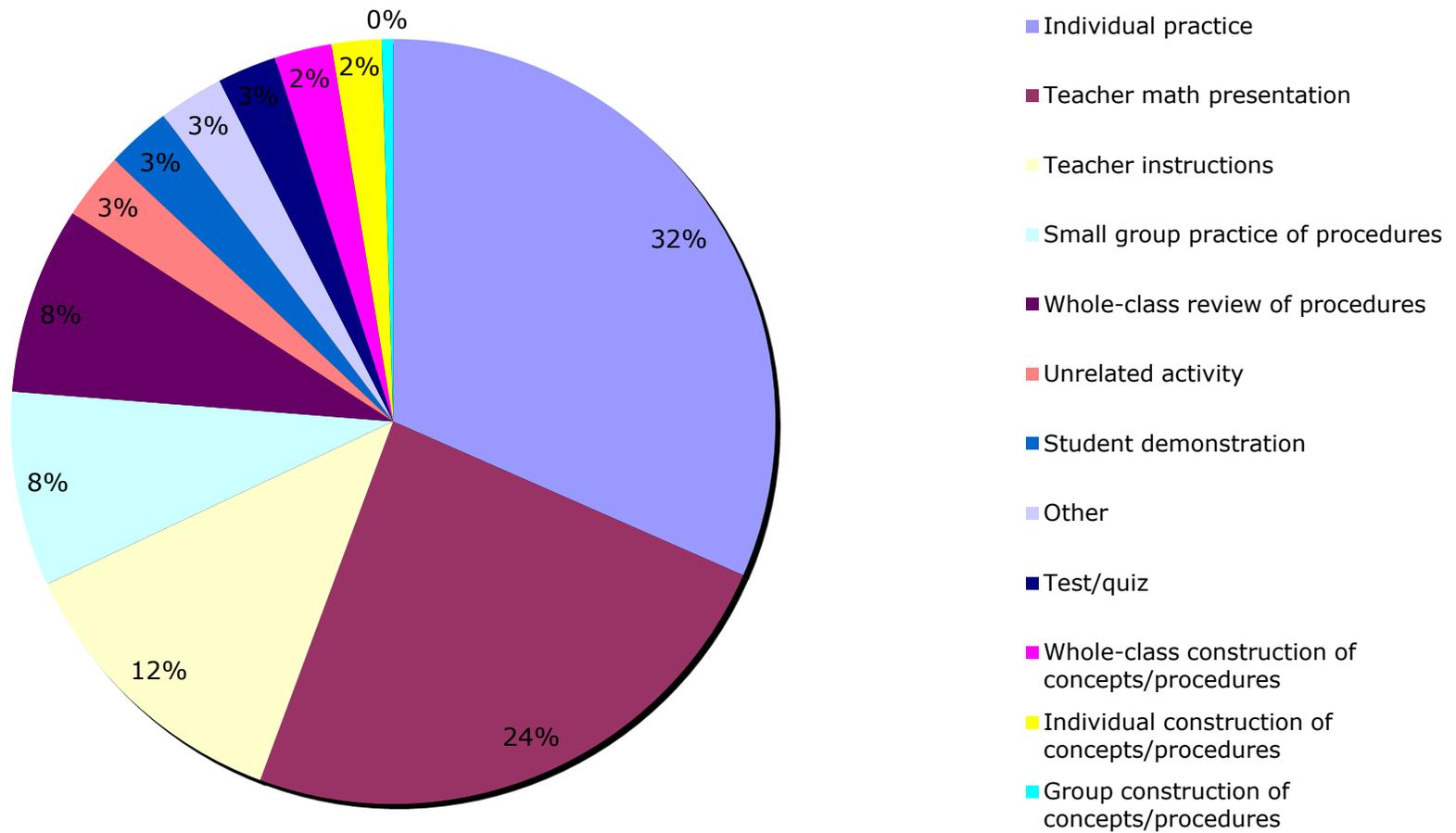
- Other significant modes are the teacher giving behavioral directives or task instructions (12% of lesson time), pair or small-group practice of learned procedures (8%), and whole-class discussion with significant pupil input to review or apply learned procedures (8%).
- Activities that promote pupil construction of new concepts or procedures total only 4% of the average lesson.

Caveats

- The 16 lessons ranged in duration from 40 to 96 minutes. We did not normalize these data; thus, a longer lesson had greater influence on the average than a shorter one.
- Six of the 16 lessons were repeat visits. Thus, six teachers each had twice the influence on the average that each of the other four teachers did.
- Teachers helped select the lesson to be observed. Thus, they possibly made greater use of modes they believed would impress our team (e.g., groupwork) than they usually do.
- The time spent on tests or quizzes reported here is probably lower than normal, because we tried to avoid observing lessons that included tests or long quizzes.

Figure 1

Breakdown of "Average" Lesson: Combined Observations



To what degree do new credential-program graduates implement CSUN-identified effective teaching practices in their classrooms?

Method

In conjunction with key SED personnel, we developed this list of research-based “effective practices” that reflected the emphases of our math credential program:

- 1) Teacher asks question or poses task with a **high level** of cognitive demand
- 2) Pupil given **authority to judge** the mathematical soundness of publicly presented solution or method
- 3) Teacher **connects** (or poses task that prompts pupil to connect) the featured math topic to **another math topic**
- 4) Teacher **connects** (or poses task that prompts pupil to connect) the featured math topic to **another academic topic**
- 5) Teacher **connects** (or poses task that prompts pupil to connect) the featured math topic to a **real-life situation or object**
- 6) Pupils allowed or encouraged to **choose** among solving **methods** or present alternative methods
- 7) Teacher or pupils use **technology, manipulatives**, body movement, or other nonverbal support for a math concept
- 8) Specific attention paid to **language**, i.e., writing, reading, or speaking skills
- 9) Teacher uses or encourages pupil to use multiple **forms of representation** for the same problem

We noted each incident of these practices, then rated the overall significance of each practice to the lesson (Significant, Marginal, or None). For each practice, we counted the number of lessons at each significance level. Figure 2 shows, for each practice, the percents of the total set of lessons that displayed each significance level.

Highlights

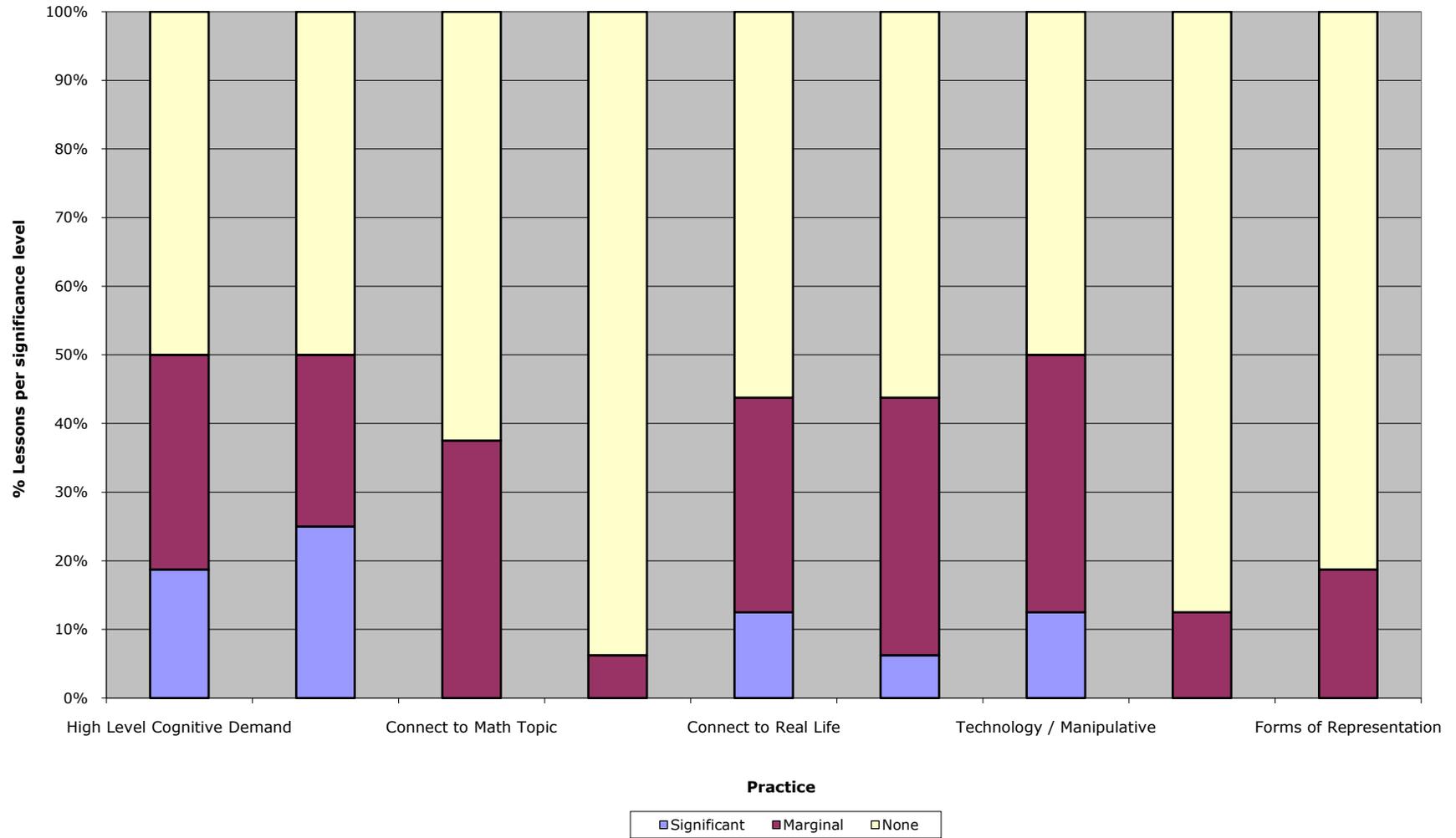
- The most-observed effective practice was giving pupils the **authority to judge** the mathematical soundness of a solution or method. It was a significant part of 25% of the lessons (4 out of 16 lessons, taught by 3 different teachers).
- Other effective practices were significant in very few lessons. Posing tasks with a **high level** of cognitive demand was significant in 19% of the lessons (3 of 16 lessons; from 3 different teachers). Other practices were only significant in 0 to 2 lessons each.
- Tasks with a **high level** of demand, pupil **authority to judge**, and **technology/manipulatives** each played at least a marginal role in half (8) of the lessons. Connections to another **math topic** or to **real life**, and giving pupils a **choice of method** each played at least marginal roles in 35% or 44% (6 or 7) of the lessons.
- Three practices, connections to another **academic topic**, attention to **language**, and multiple **forms of representation**, were rarely observed. They were not significant in any lesson and played marginal roles in 1, 2, and 3 lessons each, respectively.

Caveats:

- 1) Six of the 16 lessons were repeat visits. Thus, six teachers each had twice the influence on these totals that each of the other four teachers did.
- 2) Teachers helped select the lesson to be observed. Thus, they possibly made greater use of practices they believed would impress our team than they usually do.

Figure 2

Effective Practices



Interpreting our data in a broader context:

Prior, larger studies have aimed to evaluate and improve educational programs and initiatives through classroom observation. Two have special relevance for our study: One was conducted by and within the Los Angeles Unified School District (LAUSD) to examine the relationship between teacher performance and pupil achievement and to inform policies aimed at closing the achievement gap. The other examined classroom videos to assess the degree of adherence among U.S. math teachers to reform teaching practices (advocated by the National Council of Teachers of Mathematics, [NCTM]). While these studies differed in significant ways from ours and from each other in terms of methods and samples, they provide context for our study and results by:

- 1) Establishing a precedent for evaluating and informing programs via classroom observational studies
- 2) Demonstrating that even among experienced mathematics teachers, the implementation of research-based and reform practices in the U.S. is low, despite professional development efforts targeting those practices and widespread awareness of them.

Thus, the low level of implementation of CSUN-encouraged practices by recent graduates appears to mirror the typical performance of U.S. teachers at all experience levels, although the reasons for low implementation may differ (neither of the larger studies examines reasons). More specific comparisons between the performances of our recent graduates and the teachers in these larger studies are described below, but they must be interpreted as speculative and problematic because of significant differences in the methods and samples.

LAUSD

The Program Evaluation and Research Branch (PERB) of the LAUSD conducted a study (Daley & Valdés, 2006) that involved classroom observations of LAUSD teachers across subjects and levels, including 34 secondary math teachers (mean years of experience unreported). The PERB delineated 8 dimensions of teaching, and observers rated the overall quality of each dimension for each lesson. These dimensions were general, not subject specific, but 3 of the 8 corresponded well to elements investigated in our study:

- *Teacher-pupil engagement:* The teacher organizes the class and the learning so that instructional conversations take place.
- *Cognitive challenge:* The materials and activities challenge pupils cognitively.
- *Contextualization:* The teacher makes the content of the material meaningful to her/his pupils and draws out pupils' backgrounds/home culture to make the material relevant.

The LAUSD dimension of teacher-pupil engagement appears to subsume four of our modes:

- whole-class discussion to review/apply learned procedures (significant pupil input) (WP),
- whole-class discussion to co-construct new concepts or procedures (WC)
- pair or small-group work practice of learned procedures (GP)
- pair or small-group concept development (e.g., discovery activity/lab/project) (GC)

The LAUSD dimensions of cognitive challenge and contextualization correspond, respectively, to our practices of asking a question or posing task with a high level of cognitive demand and connecting the math to a real-life topic. Unfortunately, our rating methods were incongruous: LAUSD only reported a

rating (on a 5-point scale) for each teacher (based on multiple observed lessons) to represent the quality of each dimension in her lessons overall. In our study, we rated lessons, not teachers. This makes direct comparison difficult, but some inferences can be drawn:

- *Teacher-pupil engagement*: An examination of the LAUSD rubric suggests that teaching that earned a 1 or 2 on this dimension would not have been “counted” in our study. That is, for teaching to be counted in one of the whole-class or small group discussion modes in our study, it would have to have earned a 3 or higher in the LAUSD study. In our study, 5 teachers spent a significant part of the lesson (28%-40% of the lesson time) in one or more of the four related modes. Thus, arguably half of our sample of graduates would have been rated at a 3 or higher on this dimension in the LAUSD study. To the degree that this is true, our graduates compare very favorably in the area of teacher-pupil engagement to LAUSD secondary math teachers, only about 12% of whom (4 of the 34 teachers observed) earned a 3, and none earned a 4 or 5. This is especially noteworthy in light of the fact that LAUSD studied teachers with a range of experience levels, unlike our sample of beginners.
- *Cognitive challenge*: Here, too, it appears that for a lesson in our study to be rated at the Significant level for the practice of asking a question or posing task with a high level of cognitive demand, it would have earned a 3 or higher in the LAUSD study. In our study, 3 lessons were rated Significant on this practice but were taught by only 2 teachers, or 20% of our sample. In comparison, 8% of the LAUSD sample of secondary math teachers (3 of 34) earned a 3, and none earned higher. While these numbers are disappointingly small in both samples, at least it is arguable that our recent graduates pose cognitively high-level tasks or questions no less often than the average, more experienced LAUSD teacher, and possibly more often.
- *Contextualization*: As before, for a lesson to be rated at the Significant level in our study for the practice of connecting math to the real world, it seems it would have earned a 3 or higher in the LAUSD study. In our study, only one teacher (10% of the sample) was rated as making Significant use of this practice (in both lessons observed), as compared to ~8% (3 teachers) rated at a 3 in LAUSD, and none higher. Indeed, of the 8 dimensions, LAUSD teachers scored the lowest on contextualization. Looking down a level, only 16% of the LAUSD sample (~6 of 34 teachers) earned a 2 for contextualization, whereas 5 teachers (50%) in our study taught a lesson that we rated Marginal, seemingly the comparable rating. However, the two teachers in this set who were observed twice received the Marginal rating for one lesson and a None for the other.

To the degree that comparisons can be made, and there are many obstacles to doing so, there is support for the conclusion that our recent math-credential graduates make use of effective practices at least as often as the LAUSD secondary-math-teacher sample, and arguably more.

TIMSS

Another study offers similar data on a larger scale. Jacobs, et al. (2006) reanalyzed data from the 1995 and 1999 video studies from the Third International Math and Science Study (TIMSS) to investigate the degree to which U.S. 8th-grade math teachers adhered to practices recommended by the NCTM. The TIMSS had already identified a large set of teaching practices from the 164 U.S. video lessons in their sample; Jacobs et al. examined a subset of those practices that best reflected the NCTM *Standards* (1989, 2000). Within their subset are practices with clear counterparts in our study. Here, only the 1999 data are used for comparison because they are more recent, although they did not

generally differ significantly from the 1995 data. The 83 U.S. teachers in the 1999 video sample had a mean of 14 years of teaching experience (Hiebert et al., 2003).

- *Configuration:* Jacobs et al. found that about 6% of the time in an average U.S. 8th-grade math lesson was spent in groupwork; for our graduates the comparable value was about 9%. About 26% of the average U.S. lesson was spent in individual work, as compared to about 34% of our graduates' "average" lesson. Thus, our graduates seem on a par with the average U.S. 8th-grade math teacher in the use of groupwork, but spend comparatively more time on individual work and less on whole-class instruction.
- *Real-life connections:* Jacobs et al. found that about 22% of the problems posed in the average U.S. lesson had a real-life context, but 66% of the lessons had at least one such problem, and about 10% incorporated a real object. These data are difficult to compare to ours, because we rated lessons as a whole, not individual problems. Only 13% (2) of our graduates' lessons made significant use of real-life connections, but 35% (6 lessons) made at least marginal use. These data suggest that our graduates' use of real-world contexts or objects lags behind the average U.S. 8th-grade teacher.
- *Multiple methods:* Jacobs et al. found that in 45% of the U.S. lessons, teachers explicitly allowed pupils to choose a solution method for at least one problem; on average this happened for about 10% of the problems per lesson. Our graduates made this practice at least a marginal part of 44% (7) of the lessons, but it was only a significant part of one of those lessons. This suggests that our graduates may make marginal use of this practice with similar frequency to the average U.S. math teacher, but how the two samples compare in terms of significant use is impossible to tell.
- *Technology and Manipulatives:* 6% of the U.S. 8th-grade lessons used graphing calculators, 1% used computers, and 39% used non-graphing calculators. Assuming one technology per lesson, this means 46% of the U.S. lessons involving some sort of technology, usually non-graphing calculators.

Overall, these comparisons, while as tentative as those to the LAUSD study, suggest that our recent graduates may teach similarly to the average 8th-grade U.S. math teacher of 1999 in the degree of use of groupwork and encouragement of multiple methods but make less use of real-life connections.

CONCLUSION

As the LAUSD and Jacobs et al. studies amply illustrate, American secondary math teachers as a whole make disappointingly little use of math-teaching practices that research tells us are most effective. This is true even when the sample includes teachers at all experience levels in all regions of the country. Thus, it is no surprise that recent graduates of the CSUN secondary math-teaching credential program, in their first and second years of teaching, also make fairly minimal use of research-based practices, even when those practices have been stressed in the program. Reasons for the inertia of practice—teachers’ allegiance to traditional methods—are complex and have inspired a large research literature. Credential programs are but one of myriad factors that influence how teachers teach. Our graduates attributed some of their practice to their learning at CSUN but also cited collegial support, professional development, textbooks and other available resources, time and logistical constraints, external curricular mandates, and pupil behavior as crucial to shaping their practice.

Nevertheless, the credential program is the variable we at CSUN can control, and this study aimed to detect the impact of that variable and guide the improvement of our program so that its impact is amplified. The knowledge that we can and should observe and measure our graduates’ teaching practice are the primary outcomes of this first study, and they represent a major step for teacher education at CSUN. Our recent graduates appear to make as much (or possibly more) use of some effective practices—engaging pupils in instructional conversations, high-level tasks, and contextualization—as the average veteran LAUSD math teacher. Though they may lag behind the average U.S. 8th-grade math teacher (of 1999) in their use of contextualization, our graduates appear on a par with this national group in their use of collaborative groupwork and encouraging multiple approaches. That our graduates compare thusly in their first two years of teaching to samples spanning all experience levels may be evidence of the strength of our program. But any conclusions drawn from this first set of data about the impact of our credential program can only be speculative. Repeated cycles of study followed by data-driven program improvement are necessary before we can come to understand the impact of our credential program relative to other influences on our graduates’ teaching practice.

In addition to examining trends in practice over the coming years, there is another important question to investigate: Is it indeed an indication of program strength that our recent graduates compare favorably with veteran teachers (i.e., suggesting that our graduates will improve with experience and eventually surpass the average)? Or do the frequent users of effective practices in *any* sample of teachers tend to be the newer ones, with effective practices becoming extinguished as teachers spend more time in a school environment “hostile” to effective practices?

To make sense of the behavioral observations described in this first sub-report of the study, and to help the reader begin to gain purchase on the questions that the study raises about effective teacher practice, Sub-Report 2, which uses an anthropological triangulation of interview, survey and fieldnote data to answer the second research question of what factors facilitate or impede the effective practice of pedagogies taught in the credential program.

References

- Daley, G., & Valdés, R. (2006). *Value added analysis and classroom observation as measures of teacher performance: A preliminary report* (Publication No. 311). Los Angeles: Los Angeles Unified School District; Program Evaluation and Research Branch; Planning, Assessment and Research Division.
- Hiebert, J., Gallimore, R., Garnier, H., Givvin, K. B., Hollingsworth, H., Jacobs, J., et al. (2003). *Teaching mathematics in seven countries: Results from the TIMSS 1999 Video Study* (NCES 2003 012). Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- 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.
- National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics*. Reston, VA: Author.
- National Council of Teachers of Mathematics (NCTM). (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.