

AASCU Academic Affairs Winter Meeting
February 12, 2011
Submitted by Virginia Horvath

Plenary Session: State Data Systems and Teacher Effectiveness— The Louisiana Case Study

Jeanne M. Burns, Associate Commissioner for Teaching and Leadership Initiatives, Louisiana Board of Regents

Bradley O'Hara, Provost and VP for Academic Affairs, University of Louisiana System

Lisa Abney, Provost and VP for Academic Affairs, Northwestern SU (LA)—by phone

Opening comments from Jolanda Westerhof: the situation in Louisiana is where we are, and we can learn much from this case study

Jeanne Burns

It's good to be at the point of reflecting on what we've accomplished and sharing with you our process and results.

In LA, we are now at a point that we know teacher education is important, especially if states want K-12 students to be career and college ready. Outcome data are linked directly to the success of graduates (one of multiple measures).

- First state in U.S. that links growth of student achievement to new teachers and the teacher preparation programs that taught the new teachers.
- LA can examine growth of student achievement over multiple years to identify areas in need of program development.

Example of Results: 2009-2010 Value-Added Teacher Preparation Assessment Results

Each university receives data measuring the amount of growth in achievement of grades 4-9 students taught by new teachers—in science, language arts, reading, math, and social studies.

- Level 1: greater than students taught by experienced teachers
- Level 2: growth comparable to students taught by experienced teachers
- Level 3: growth comparable to students taught by new teachers
- Level 4: growth below students taught by other new teachers
- Level 5: growth significantly below students taught by other new teachers

Each institution then learns the “effect estimates” for all of their programs. HE institutions will effectively have results of how their new teachers are doing for each of their programs. They also receive longitudinal assessment results.

Process in Louisiana

- 1999-2000: Blue Ribbon Commission for Teacher Quality identified inconsistencies in programs and results. New policies created stronger teacher certification requirements, as well as

stronger alternate and undergraduate pathways. There was an agreement between PK-12 and HE to share data.

- Using multiple measures to determine the effectiveness of teacher preparation. This has included redesign of all teacher preparation programs and approval by BoR and BESE, including alignment of programs to state and national content and teacher standards. CAOs were held to be accountable for the redesign of all programs due to the involvement of colleges of arts/sciences/humanities and education faculty. Evaluation of programs was by national/state experts and consultants. Programs were not approved until they met all standards, and they had to do this by 2003.
- In LA, a great group of public and private education deans worked together on this. All programs were approved by this date, but some had much deeper changes than others to get to that point.
- Another measure is that all institutions are required to be nationally accredited to operate. This is important because the state achievement measure is only grades 4-9, so the accreditation ensures that other programs are also strong and infrastructure is in place.
- Teacher Preparation Accountability System: developing a performance score that has three parts.
- One part is a value-added model (developed by George Noell and Kristin Gansle from LSU and A&M). This predicts the achievement of individual students based on prior achievement, demographics, and attendance. It can assess actual student achievement and link that achievement to individual new teachers and teacher preparation programs. And they act on the results.
- Tests: state achievement tests
- New teachers: 1st and 2nd year
- Experienced teachers:
- Criteria for inclusion of programs in the assessment: redesigned programs only, 25 or more new teachers in grades 4-9, teaching within certification, remained with student full academic year.
- 10 universities and 2 private providers included in 2009-10 results

If you ask campuses to be innovative and creative, they may not be successful at first. But having the data about growth of learning can be very valuable to campuses as they consider what to change.

Institutions are identified by a particular score for each program, and the overall results can be useful in looking at which programs are doing a good job.

Next Steps in Louisiana

- Creation of state research team
- Examination of data to identify needs
- Use of data to identify a strategy to address the needs
- Development of an action research study
- Implementation of the action research study

Brad O'Hara

What do system and campus CAOs think about the use of multiple measures to examine teacher effectiveness?

- There was a lot of anger and upset at news of mandated program redesign, along with an appreciation of the amount of work that needed to be done.
- There was also upset in arts and sciences area, where results were being questioned.
- People were able to step up and need these changes.

What roles did deans and CAOs have in this process?

- They were intimately involved in the pilot that was used to shape the measures.
- When the results of the pilot were rolled out, presidents and CAOs started to question the results of the data, especially reporting on pre-redesign efforts.
- Communication about this very complex model and what the results mean was very important. Each year before the results are released, the system office reviews the results with individual campuses—usually a week in advance. There is a high turnover of provosts in the state, so there is new educating every year for campus understanding to be accomplished. They are asked not to release the results until after the Board of Regents meeting when results are released.
- This has become very high profile each year, much anticipated. Always there is the want to have programs compared across the state, and campuses need to be prepared to answer those questions once the results reach the public.

Lisa Abney

- Assumed her role after the results were released, but had been at Northwestern SU for years and knew of the redesign program.
- It was eye-opening to see the whole picture. It's a complicated data set, one that leads to a lot of questions.
- Northwestern scored well, and we were glad for the national publicity about this. Institution was a normal school with a long tradition of teacher preparation, so it was great to have the program affirmed as well as challenges outlined.

The first university to have a Level-4 score was led by a president who was involved as a former commissioner in the creation of this model.

- Benefits of multiple measures as well as value-added results: media tend to focus on one particular measure, but the institution can look at the full picture. From the system and campus perspective, this effort has enhanced NCATE efforts.
- It's important to recognize that when we use multiple measures, there appears on the surface to be conflicting results. A SPA report was rated one of the best, but this measure had the same program rated low.
- Challenges of a value-added model: trying to mine the data and figure out what is going well and not going well is useful. But the data are tough to mine. We have to think carefully before making an overhaul. Good communication is essential, with deans of colleges of education and provosts all working together.
- They are starting to look at the campuses with good scores to see best practices and have others learn from that.
- There is now heightened awareness and angst about these results. As the data become richer and richer, this will put pressure on the institutions and individuals that are not doing all that well. The other things that are happening are moving into continuous improvement and regular

redesign of programs. Provosts might say we're in redesign burnout because it's all we're doing, but isn't reflective practice what we're supposed to be doing?

In 1999-2000, Louisiana didn't have the capacity to do this, but they knew that had to do it. So they have spent considerable time building this. There are reports based on pilot data and then data that can be examined. We are openly sharing everything that we have done so that other states and universities can benefit from looking at our work, available on the Board of Regents website, along with this PowerPoint.

Q&A

- Relationship to *US News* ratings: if universities are demonstrating that programs leading to students learning, LA is going to go with our own results instead of the new rating system of *US News*. They are looking at outputs in LA.
- Model examining demographics: the models look at free/reduced lunch, prior achievement, teacher absences, learning disabilities, student absences, race, mild mental retardation. All of these are predictors.
- How do you know you're evaluating the effectiveness of the program, instead of the quality of the students who came into the program? We had a Carnegie Foundation grant that identified factors impacting teacher performance once they went into the schools. They looked at ACT scores and found that there wasn't much variation except in mathematics, so all came in with ranges of knowledge that impacted their success. They also looked at pathways that affect success. Some universities had practitioner-teacher programs that were effective, some that were not. Collected data about conditions, teacher qualities, other factors but didn't see huge difference.
- Teachers with higher effect estimates and lower effect estimates: highest estimates were more critical of their institutions than others, suggesting that they have developed skills in critical thinking.
- Propensity test looked at teachers by district to see impact on effect estimates. A low effect estimate is consistent across districts for particular institutions—not within the districts themselves.
- Education deans looked closely at the pilot data and asked questions that helped control for variables and provided bases for refinement of measures. A lot of thought went into the creation of this model before it was released.
- Explanation for the alternate-delivery models, one of which had very good scores and the other had the worst scores? The New Teacher Project—including Teach for America—the new teachers are doing very well. Louisiana has learned that those in that program then leave the state, so there is no long-term benefit for the state. Another private provider has an alternate practitioner teacher program with an alternate pathway. We have to be very careful about looking at the results, even with private and public, when we consider that government is very against universities providing teacher education programs. Just because you're a private provider doesn't mean that you'll achieve results, and these data are being used to try to change legislators' minds about this perception.
- Reaction from experienced teachers where new teachers are rated higher? The teacher unions were positive about what we were doing because achievement of new teachers is an important issue. Louisiana teachers do not know their individual effect estimates, but they've been asked to develop a model for practicing teachers. Every teacher in LA next year will learn his or her

effect estimate. Some of the new teachers may be rated higher than the experienced teachers, who will have to go into a program to enhance their skills.