Journal of Contemporary Research in Education

Volume 4
Number 1 *Fall 2015*

Article 3

10-1-2015

The Effects of Continuous Improvement Practices Using State Value-Added Data for Reading Educator Program Enhancement

Kristin A. Gansle
Louisiana State University

Gerlinde Grandstaff-Beckers
Southeastern Louisiana University

Angelle Stringer
East Ascension High School (Louisiana)

Nancy Roberts
Louisiana Resource Center for Educators

Jeanne M. Burns
Louisiana Board of Regents

See next page for additional authors

The Effects of Continuous Improvement Practices Using State Value-Added Data for Reading Educator Program Enhancement

Authors

Kristin A. Gansle, Gerlinde Grandstaff-Beckers, Angelle Stringer, Nancy Roberts, Jeanne M. Burns, and George Noell

Journal of Contemporary Research in Education
4(1) 4-23

The Effects of Continuous Improvement Practices Using State Value-Added Data for Reading Educator Program Enhancement

Kristin A. Gansle
Louisiana State University
Gerlinde Grandstaff-Beckers
Southeastern Louisiana University
Angelle Stringer
East Ascension High School (LA)
Nancy Roberts
Louisiana Resource Center for Educators
Jeanne M. Burns
Louisiana Board of Regents
George Noell
Louisiana State University

Abstract

Louisiana's value-added evaluation of teacher preparation programs has provided a salient impetus for program improvement; however, due to the nature of the assessment, teacher preparation programs need to use additional sources of data to identify actionable responses to the value-added results. This paper describes one teacher preparation program's approach to continuous program improvement in reading education and describes some of the limitations and benefits of value-added assessment results for that purpose.

Challenges to business and industry that increased during the 1980s and 1990s, including surges in global competition, changes in markets, and escalation in the necessity to master ever-improving technology have dramatically heightened the need for organizations to collect and interpret data that informs accountability systems and contributes to organizational improvement (Locke & Jain, 1995). These challenges to business have had attendant effects on the educational system that feeds business its intellectual capital. The emphasis on quality and quantity in the development of that capital has augmented the need for accountability and the verification of teaching outcomes. In no content area is this need more evident than in evaluating reading instruction. Large-scale national testing has indicated that primary and secondary students in the United States are ill-prepared for reading decoding and comprehension (National Center for Education Statistics, 2007), critical skills for gaining knowledge from content-specific texts. Despite the fact that these testing data exist, are repeated measures of important outcomes, and could be used to evaluate teacher performance to improve educational systems to a competitive benefit (Reusser, Butler, Symonds, Vetter, & Wall, 2007; Stata, 1989), it has not been unusual for systems to either use them in only a punitive manner or choose not to use them at all. However, systematic use of student data in teacher evaluation is increasingly apparent (Papay, 2010); the advent of databases that link these data to evaluate teacher performance in the classroom are making the use of such evaluations possible (e.g., Anderman, Anderman, Yough, & Gimbert, 2010;

Ballou, Sanders, & Wright, 2004; Gansle, Noell, & Burns, 2012; Hershberg, Simon, & Lea-Kruger, 2004; McCaffrey, Lockwood, Koretz, & Hamilton, 2003).

Value-added analysis or modeling (VAM) was originally developed in industry to support continuous improvement (CI: Bhuiyan & Baghel, 2005; Schroeder & Robinson, 1991) and is now possible within education in those domains for which educational systems collect critical outcome data (e.g., test scores, graduation rates, discipline referrals). VAM allows for the description of achievement outcomes for students at the individual teacher level in a given content area. What sets this approach apart from traditional single-measurement assessments is that teachers can be evaluated based on the extent to which their students' *observed* achievement is different from what would be *predicted* for them given information that is known about students and their classroom contexts. A range of variables that contribute to students' achievement are measured and included in the model. These are generally comprised of demographic data and previous achievement, attendance, teacher, and classroom information. These variables are used to predict what the current year's achievement score in a given content area (i.e., reading, English-language arts, mathematics, science, or social studies) would be under the conditions specified by the demographic and prior achievement variables, and this is compared to the student's measured achievement scores. The differences between predicted and observed scores are then used as an assessment of teachers' instruction (see Noell, Gansle, Patt, & Schafer, 2009, for a detailed description).

Although there is an ongoing debate surrounding the use of value-added data to evaluate individual teachers (Baker et al., 2010; Glazerman, Loeb, Goldhaber, Staiger, Raudenbush, & Whitehurst, 2010; Hanson, 1988; Harris, 2009; Raudenbush, 2004; Tekwe et al., 2004; Viadero, 2008), states and school systems are increasingly using these data-based systems as an input to assessing teachers and making consequential employment decisions (Boyd et al., 2006; Heitin, 2011; Isenberg, Hock, & Mathematica Policy Research I, 2011; Lasley, Siedentop, & Yinger, 2006; Sawchuk, 2011). Indeed, with 33% of fourth-grade students and 24% of eighth grade students scoring below basic in reading (National Center for Education Statistics, 2011), and the increasing focus on accountability for teachers, there is a move toward data-based instructional problem solving in reading: effective, feasible, and time efficient instruction and intervention are crucial to the success of our children (Ross & Begeny, 2014).

School systems, however, are not the only educational institutions that are using value-added data. In evaluating the effectiveness of teachers, one potential logical source of variation among them that might be addressed in intervention is the teacher preparation program (TPP) that recruited, prepared, and recommended them for certification (Cochran-Smith & Zeichner, 2005; Darling-Hammond & Bransford, 2005; Rice, 2003; Wilson, Floden, & Ferrini-Mundy, 2001). TPP evaluation uses data from multiple teachers rather than individual teachers, collected across contexts and over time, which serves to ease *some* of the extant concerns about the use of value-added methods to evaluate individual teachers (see Gansle et al., 2012, for a discussion). This type of evaluation also provides data regarding the most important outcome of training teachers: the effectiveness of TPPs in training their

completers to positively affect the achievement of the students entrusted to them.

Louisiana's Assessment of Teacher Preparation Programs

Louisiana began using VAM in pilot form in 2003, followed by de-identified form, and then in program-identified form to evaluate TPPs (please see Noell & Burns, 2006; http://regents.louisiana.gov/academic-affairs/teacher-education-initiatives/value-added-teacher-preparation-program-assessment-model/). All students in grades 4 through 9 who take the standard state assessments in English-Language Arts, reading, mathematics, science, and social studies participate in the program. If a student is not included in the analysis, it is either because they are exempt from the testing program due to severe disability or they have been retained, making their scores not strictly comparable to others and inappropriate to include in teacher assessment. All teacher preparation pathways in the state are assessed in the same way (e.g., private providers, traditional undergraduate certification, master's degree alternate certification). Students' previous achievement scores, student, class, and school characteristics, and student and teacher attendance are used to predict the next year's scores through Hierarchical Linear Modeling (please see Noell et al., 2009 and Gansle et al., 2012, for descriptions). Annual achievement test scores for students on the *Louisiana Educational Assessment Program* (*LEAP;* Louisiana Department of Education, 2008b) and the *integrated* Louisiana Educational Assessment Program (*i*LEAP*;* Louisiana Department of Education, 2008a) are used in the analysis.

New teachers' scores from TPPs with at least 25 teachers teaching in tested grades and subjects are used to create the program means (Gansle et al., 2012). These scores are compared to the means for both new teachers and for experienced certified teachers throughout the state. Program means are then assigned a rating according to pre-defined performance levels specified by the Board of Regents. Table 1 contains descriptions of the performance levels that have been used in the Board of Regents system that evaluates TPPs.

TPP assessment has historically shown that in Louisiana, there is considerable variation across programs: from much lower than average new teachers to much higher than experienced certified teachers. The most important issue in this CI model is that the evaluation process does not end with the assignment of programs to scores and performance levels. On the contrary, it is merely the beginning of the most important part of the process: evaluation and *revision* of the existing program and structures.

According to State policy (Louisiana Administrative Code, Title 28, February 2011), any TPP that is evaluated and receives a Performance Level 4 or 5 in any content area within a teacher preparation program is assigned a designation of *programmatic intervention* in that content area. Within one year of the release of the assessment results, programs assigned to programmatic intervention must (1) review their existing program with an expert in the field that is recognized nationally as well as with a content area specialist that is designated by the Louisiana State Superintendent of Education. Following that review, (2) a corrective action plan must be designed to remediate the perceived deficits in the program, including a time

frame for when results of the corrections made might be anticipated in future value-added assessment assessments. Programs that do not improve are at risk of losing state approval to prepare teachers in that content area.

The Louisiana Resource Center for Educators (LRCE)

LRCE is a private teacher preparation program provider and source for teaching materials and continuing education in Baton Rouge, Louisiana. They offer a practitioner program called Certification Solutions that has been preparing teachers since 2003. It has selective admissions criteria through which individuals with a bachelor's degree from an accredited institution may gain teacher certification. Those selected to attend the Certification Solutions program can achieve teacher certification concurrent with employment as teachers in between 15 and 36 months.

Admission into the LRCE program is predicated on the submission of records indicating passing Praxis I and Praxis II content area scores. In addition, a personal interview, evaluation of a writing sample, and a law enforcement background check are necessary for admission. During the summer prior to beginning a practitioner year, candidates participate in seven weeks of intensive, full-time training sessions on classroom organization and management, instructional delivery, childhood development, adolescent psychology, technology in the classroom, lesson planning, differentiated instruction, school law, reading in the content areas, and special education. In addition, candidates observe and complete clinical teaching hours at area schools with supervision from program staff. At the close of the summer institute, candidates are eligible to teach on a practitioner license in a partner school and receive full-time salary and benefits. Partner schools must be a Louisiana public school or a State-approved private school. During the practitioner year, candidates attend content-specific learning team meetings every two weeks which are taught by master classroom teachers called Learning Team Leaders. They also receive guidance from program mentors known as Practitioner Advisors, who observe the candidate in the classroom setting. Practitioner Advisors are professional educators with classroom and supervisory experience. Candidates are required to pass the Principles of Learning and Teaching or Special Education components of the PRAXIS, and Early Childhood, Elementary, and Special Education candidates must pass the Teaching Reading PRAXIS. Following successful completion of these requirements and the practitioner year, positive evaluations from the school administrator, Practitioner Advisor, Learning Team Leader, and Certification Solutions staff, candidates may obtain their Level 1 Louisiana teacher license.

LRCE is neither a college nor university and is therefore not subject to the requirements of accrediting bodies such as the Southern Association of Colleges and Schools (SACS) and the National Council for Accreditation of Teacher Education (NCATE) for peer institutions. However, they and other private providers must demonstrate to the State that they meet state and national state/national teacher and content standards and other criteria to be approved to operate a teacher preparation program within the state that will result in teacher certification. The Louisiana State Department of Education requires all private providers to submit proposals that address guidelines that are aligned with guidelines for Practitioner Teacher Programs within

universities. All proposals are evaluated by national experts and programs must address weaknesses identified by the national experts before the programs are approved by the Board of Elementary and Secondary to operate within the State. Any alterations to the program must be first approved by the Louisiana Department of Education.

Value-Added Assessment & Program Evaluation

1st stage assessment and programmatic intervention. In fall 2008, Louisiana released the first value-added results for LRCE: their result in reading was -6.2 points (test mean is approximately 300, sd = 50; Noell, Porter, Patt, & Dahir, 2008). This indicated that the mean effect of LRCE's teachers on student achievement as measured by the State's standardized achievement tests (*LEAP* and *i*LEAP) in reading was on average 6.2 points below that of experienced certified teachers (which is set as the reference at 0 points). In other words, students in LRCE-trained teachers' classrooms were losing, on average, 6.2 points on the assessment *per year* versus an experienced certified teacher, which put them at Performance Level 5. The next nearest program effect estimate in reading for another teacher preparation program was -2.4 points at Performance Level 3. The mean effect for new reading teachers was -1.8. In English-language arts, mathematics, science, and social studies, LRCE results did not meet the standard for State-mandated programmatic intervention areas.

Shortly after release of these results, LRCE began programmatic intervention in reading instruction. Although evaluation of teacher candidates occurred on a regular basis, no formal formative assessment or evaluation that specifically addressed intensive reading instruction had been used by LRCE prior to the implementation of programmatic intervention. In order to improve their capacity in this regard, LRCE's first step was to design an informational survey of individuals who had just completed the summer institute on effective reading instruction and classroom management (available from the authors). This survey of three pages asked open-ended questions to assess candidates' level of comfort with and use of five specific fundamentals of reading instruction in their summer teaching (e.g., vocabulary, comprehension, phonics, phonemic awareness, fluency), as well as specific teacher behaviors such as grouping students, managing several groups, transitions, and learning centers.

Based on feedback from the survey, staff was increased to two doctoral-level and one masters-level reading educators in addition to those trained in general education. This allowed for doubling the concentrated reading instruction that previously had been offered at the summer institute to 35 hours. Five core areas of reading were designated as the focus of this instruction: phonemic awareness, phonemes, fluency, vocabulary, and comprehension. In addition, this training in effective literacy instruction was provided to the Learning Team Leaders and Practitioner Advisors who would work with the practitioner teachers throughout the academic year. This created a *strategy* for a coherent plan of instruction in reading aligned with the critical areas identified by the National Reading Panel's findings (2000).

However, LRCE would be unable to rely on frequent measurement to inform and improve their program changes. Value-added assessment occurs only once per year, and there is a consequential delay between when the evaluative data become available

and when teachers are prepared. For example, a candidate who completed a TPP in 2012 would be eligible to receive their Level 1 teaching license in 2012-13 and count as a new teacher for a TPP with the spring 2013 achievement testing data. However, several months of data cleaning, database merging, and value-added analysis typically makes value-added results ready for release by the Board of Regents in the summer following (2014). This delay necessitates additional data collection and evaluation for TPPs to engage in effective CI. In order for LRCE to truly improve their program, it would be necessary to collect the data that would be formative in nature on a more frequent basis. Additionally, value-added data are global outcome indicators. They do not provide any indicators that programs might use to make constructive change and do not answer questions regarding what to do in terms of instructional modifications. In order to act, programs have to closely examine their processes as well as their results. Consequently, a direct observation of teaching behavior was designed to score the presence or absence and quality of the several aspects of reading instruction on a 3 point scale: 1 (*emerging*), 2 (*acceptable*), and 3 (*proficient*, please see figure 1 for the instrument used for direct observation). This instrument was used following the first redesign of curriculum and training procedures.

The summary report on the formal evaluations using this instrument indicated that of the 40 teachers observed, an average of 34 practitioners per item were rated either *acceptable* or *proficient*. Evaluators determined that the redesigned program produced teacher performance at or above the level expected of first-year teachers. However, the extent to which these reported behaviors would be detectable by value-added scores had yet to be determined.

2nd stage assessment and programmatic intervention. The following year's results in reading during fall 2009 were similar in magnitude but better in terms of level. LRCE's mean teacher reading effect estimate was -6.3 points (Noell et al., 2009). This indicated that the mean effect of teachers on student achievement as measured by the state's standardized achievement tests (*LEAP* and *i*LEAP) in reading was on average 6.3 points below that of experienced certified teachers. Although the size of the effect estimate for the program in reading was approximately the same as the previous year, the mean new teacher effect was -2.8 in reading in 2009, which led to the difference in level as the LRCE effect was closer to the mean of new teachers. It is important to recognize that these results were obtained for teachers who completed the program before the programmatic changes described above had been implemented.

Although LRCE's level in reading had increased to Performance Level 4, programmatic intervention was still required according to state policy. At this point, LRCE contracted with one national reading expert and one State reading expert per the Louisiana Department of Education's directive. The possible selections for the national expert had been provided as a discrete list by the Louisiana Department of Education. The State expert could be chosen from any available in Louisiana but had to be approved by the state based on an evaluation of the expert's credentials in curriculum, standards and pedagogy in reading, practical experience, service, and scholarly contribution to the field.

National expert. The national reading expert chosen suggested that LRCE

perform an assessment that specifically rated the levels of student engagement and presence of features of effective instruction in phonological awareness, phonics and word study, fluency, vocabulary and oral language, comprehension, and writing. The items were tailored to either early reading skills or later reading skills (checklists are available from the authors). Each teacher candidate was observed and the level of student engagement was rated by Practitioner Advisors and Team Leaders as one of three choices on the data collection sheet: low (less than 80%), medium (80% to 90%), and high (greater than 90%). Five features of effective instruction were marked as present or absent with respect to each of the areas above (phonological awareness through writing): evidence of explicit, systematic instruction; efficient use of time; opportunities to respond; immediate corrective feedback; and differentiated instruction. Further, appropriateness of the teacher's lesson pace, and conduciveness of the environment for learning were evaluated. For the 30 teachers evaluated at the PreK-3 level, the only areas in which fewer than 85% of practitioner teachers were rated as successful were in teaching fluency (77%), teaching vocabulary and oral language (75%), and utilizing differentiated instruction (60%). At the Grades 4-9 level, for 35 teachers, the only area in which fewer than 84% of practitioner teachers were rated as successful was in teaching writing (77%).

State expert. Following these evaluations, the curriculum of the summer institute was reviewed by the state expert. Additional instructional materials were assembled to better align the LRCE curriculum with empirically-derived best practices in direct instruction in reading as indicated in Carnine, Silbert, Kame'enui, and Tarver (2010). The curriculum detailed in *Direct Instruction Reading* (Carnine et al., 2010) indicates that teachers must be knowledgeable in the five areas of essential skills for reading process and procedures defined by the National Reading Panel (NRP; 2000): phonemic awareness, phonics, vocabulary, fluency, and comprehension. Review of the previous reading curriculum in use at LRCE indicated that despite those five areas having been addressed by the original assessment and revision of curriculum, the scope and sequence of the curriculum was not appropriately aligned with state standards to meet the needs of all students, and furthermore, the curriculum did not demonstrate explicit and systematic instructional approaches to each of those five areas outlined above. Adjustments were instituted to align the scope and sequence to meet standards.

An appraisal was done of the procedures for evaluating, selecting, and modifying programs to meet needs of all students based on researched based best practices of the National Reading Panel (2000) and Carnine et al. (2010). The extent to which the LRCE staff taught techniques to candidates for effectively presenting lessons, pacing tasks, motivating students to do their best work, and diagnosing and correcting errors was evaluated. This evaluation again demonstrated a lack of awareness and implementation of systematic, explicit, instruction. Further, the program was evaluated to determine the extent to which students were instructed to use assessments to create and modify instructional programs, and whether they were taught to use strategies to maximize time spent with students engaged in literacy instruction.

Based on the state expert's review, a meeting was held in which she and the program staff addressed areas of concern within LRCE's reading program by delineating skills and incorporating a

sequence of instruction aligned with empirically-supported outcomes such as pacing tasks and assessment to allow for increased student engagement (Fisher & Frey, 2008; Greenwood, Arreaga-Mayer, & Carta, 1994; Palincsar & Herrenkohl, 2002; Shanahan & Shanahan, 2008). Several sets of master lesson plans were rewritten to delineate the scope and sequence of reading instruction specific to various K-12 settings that were aligned with the evaluation tool and State grade level expectations to identify necessary and specific aspects of appropriate reading instruction. Following lesson plan creation, the state expert assessed the resources and professional literature available to candidates at LRCE. Deficiencies in the resources available were identified. A library of the empirically supported practices and professional literature in phonemic awareness, fluency, phonics, vocabulary, and comprehension to be used in the Summer Institute was created (a list of these is available from the authors). These resources were intended to range from an introductory level to an advanced level and were made available to candidates to be used in learning team seminars throughout the academic year. The assessment system and subsequent curricular changes implemented by the state expert were aligned with the tenets of the National Reading Panel (2000) as well best practice sequenced instruction as documented through the available professional literature.

The state expert provided professional development to Team Leaders as well as on-site evaluators (Practitioner Advisors) in content areas on systematic, explicit instruction in literacy. Participants were provided with an overview of the lesson, assessed for practical knowledge by dividing them into groups and asking them to create lesson plans for literacy or integrating literacy into content area lesson plans. The facilitator then used an "I do…We do…You do" approach whereby the activity was first demonstrated to the participants, after which they were guided through the activity with feedback, and finally, the group individually practiced the skill while the facilitator evaluated the participants' ability to complete the task (Juel & Minden-Cupp, 2000; Stanovich, 1994).

Next, Team Leaders and Practitioner Advisors taught lessons to the group in order to demonstrate explicit instruction in literacy. An exit evaluation was conducted to determine remaining supports needed which subsequently were addressed individually. Remaining supports requested were additional research and references on explicit systematic instruction and integrating content literacy strategy instruction. Throughout the program evaluation and redesign of the program, numerous strategic planning meetings with LRCE management team and the experts were convened to discuss evaluations, findings, content of the curriculum, empirically-based instructional practices and future directions of the reading program.

2nd round results. Following the second round of programmatic intervention, the value-added score released by the Louisiana Board of Regents during fall 2011 was -5.0 points (Gansle, Noell, Knox, & Schafer, 2010). Although this was a Performance Level 4 result, LRCE was informed that results for recent completers were showing gains, and they chose to stay the course with the last round of program changes after consulting with the value-added assessment team. It is important to note that these results would not include the impact of the second round of more extensive program improvement efforts.

Building on the previous data gathered in 2010-2011 academic year, LRCE sought to continue gathering data depicting the practitioners' ability to provide effective literacy/reading instruction. Using previous assessment instruments as well as site field notes, evaluators (Learning Team Leaders and Practitioners Advisors) were able to pinpoint strengths and challenges of candidates' abilities to teach literacy foundations and adjust practices accordingly. Although the state expert designed the assessment and evaluation procedures, she has taught the LRCE staff to implement the assessment and evaluate the results, and make program changes according to those results. LRCE continues to collect the data that Learning Team Leaders and Practitioner Advisors use to make changes to the Summer Institutes and program curricula. Specifically, concerns with pacing, literacy centers, and classroom environments have been addressed in the past, and they continue to compile data on the effectiveness of literacy instruction and adjust instruction accordingly.

Next round value-added results. During fall 2011, the State made a decision to use an adaptation of the VAA-TPP to examine the effectiveness of teacher preparation programs. Louisiana Department of Education and the VAA-TPP worked together to adapt the Value-Added Teacher Preparation Assessment to create a value-added teacher evaluation model to assess practicing teachers in grades 4-9 in tested content areas per the requirements of a recent change in law. Results for LRCE in reading were quite similar across either assessment approach and yielded the same substantive conclusions. The fall 2011 result for LRCE in reading was 0.4 points (standard error of measurement: 1.0 points; Gansle, Burns, & Noell, 2011). For new teachers, the mean effect was -1.2 points (standard error of measurement, 0.2 points; Gansle et al., 2011). These results for LRCE have been considered as a substantial improvement in their reading score and they have since moved out of programmatic intervention.

Comparison of results across reading and mathematics. Although the results described above appear to indicate that the program was improving over time, it is possible that there were other factors that might account for the changes in reading scores of new teachers trained by LRCE. Their mathematics scores had not been sufficiently low to meet standards for programmatic intervention; as a result LRCE had made no program changes to their mathematics instruction classes or activities. Although this was a program evaluation, rather than a controlled study, we opted to compare the mathematics scores and the reading scores for the same years' new teachers. A graph of this comparison is contained in Figure 3. New teachers' reading scores made considerable gains over the 4 years, while the mathematics scores remained largely the same over the same period, suggesting that the changes in reading scores were related to the changes made in the LRCE program.

Discussion and Future Directions

Continuous improvement is increasingly being used in education, and the advent of recent data systems and improvement of analytic capacity of systems have allowed for the incorporation of data-based evaluation of teacher and TPP effectiveness (Anderman et al., 2010; Ballou et al., 2004; Bhuiyan & Baghel, 2005; Gansle et al., 2012; Hershberg et al., 2004; McCaffrey et al., 2003; Schroeder & Robinson, 1991). Value-added assessment can provide TPPs with data designed for

program improvement; what makes this kind of assessment unique is that it assesses what students with similar previous achievement and demographics achieve relative to their predicted achievement (McCaffrey et al., 2003). This is a giant step forward compared to single-data-point measures of educational outcomes where, for example, high socioeconomic status schools are identified as more effective than those serving high poverty student bodies as the result of testing data from a single spring assessment that does not account for the progress those students made. Obviously, end point only assessments are inadequate.

However, one critical limiting issue associated with using value-added scores as outcomes in a CI model is delay. To date, when an alternate certification model program such as LRCE has made major changes to its training, from the moment the new training plan is implemented, it has been a minimum of three years until the first cohort contributes to a value-added result, and this assumes that the initial cohort completed the program in 15 months. Please see figure 2 for a sample timeline. Clearly, not all candidates finish this quickly and commence employment immediately following program completion. This creates a less than ideal situation for monitoring the progress of TPPs toward the improved outputs of quality teachers if changes are made to the program. Essentially, the critical issue is the lack of sensitivity of the measure used to determine teachers' effectiveness (Jenkins, Deno, & Mirkin, 1979). Because it provides only one measurement occasion per year, the standardized testing program used by the state is not designed to assess short-term progress made during the course of or even following intervention (Gansle, Noell, VanDerHeyden, Slider, Hoffpauir, Whitmarsh, & Naquin, 2004). Although it might provide limited data for formative evaluation, additional measures would be extraordinarily useful. Using the new program that Louisiana has in place for individual teacher evaluation, the timeline should improve to a minimum of two years following program changes, but this is still a period of time that makes the formative use of these value-added data problematic.

LRCE chose to use direct assessment of their candidates' teaching skills during training as a more sensitive measure of their progress toward best practice in reading instruction. Although this is clearly not a direct assessment of their *students'* academic achievement, it does provide program administrators with a clearer indication of their *candidates'* skills. The assessment utilized focuses on behaviors that can be demonstrated as related to positive outcomes for students, behaviors that are assessed reliably, and provides results that can be used for program improvement. Continuous improvement may be realized by the use of more sensitive program assessments (Gilham, Lucas, & Sivewright, 1997) that may be combined with summative evaluation opportunities provided by yearly value-added assessment conducted by the state.

Limitations and areas for improvement. Although the intent of the LRCE staff was excellent with respect to assessment of and revision of their curricular practices, the instruments that were used to collect data and the training provided to the staff to use them revealed substantive gaps as they were more closely examined. For example, the Team Leaders and Practitioner Advisors who were charged with collecting data neither participated in formal training nor were held to any specific standard for judging the items. For example, they watched the classroom and determined the

level of engagement without a schedule, procedure, or data collection instrument for determining that percentage of engagement. Observer training, operational definitions of variables, and using specific defined observational techniques certainly would have contributed to an improved observational scheme and perhaps more descriptive and useful data (Cooper, Heron, & Heward, 2007).

Despite these limitations, however, it appears that LRCE made substantive progress toward improving their program in reading through their response to a global outcome measure that suggested concern followed by more micro-analytic and recursive assessments of preparation practices to identify gaps and make changes. Their value-added scores suggested to the state and to their staff that programmatic intervention was needed. The program's evaluation indicated that their curriculum needed revision, which was undertaken and was followed by substantive changes in value-added results. Further, these changes were considerable, especially when compared to the lack of change in mathematics, a content area that was not changed in response to value-added results. LRCE's experience with data-informed continuous improvement may be a harbinger of challenges that will confront teacher preparation programs across the country as these sorts of analyses become more widely available. Value-added results may suggest areas that teacher educators will be concerned about, but they will not provide information about why the results occur in the pattern they do or how to improve them. In order for this type of data to support beneficial program revision it will have to be married with thoughtful and honest assessment of current admissions, preparation, and assessment practices inside programs. This is a process challenge for continuous improvement that will be broadly shared across many preparation programs. The role of the value-added results in this process is to highlight areas of concern, motivate change, focus effort, and provide objective external feedback on the impact of change efforts.

References

Anderman, E. M., Anderman, L. H., Yough, M. S., & Gimbert, B. G. (2010). Value-added models of assessment: Implications for motivation and accountability. *Educational Psychologist, 45,* 123-137. doi: 10.1080/00461521003703045

Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., Ravitch, D., Rothstein, R., Shavelson, R. J., & Shepard, L.A. (2010). Problems with the use of student test scores to evaluate teachers. Washington, DC: The Economic Policy Institute. http://epi.3cdn.net/b9667271ee6c154195_t9m6iij8k.pdf

Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics, 29,* 37-65. doi: 10.3102/10769986029001037

Bhuiyan, N., & Baghel, A. (2005). An overview of continuous improvement: From the past to the present. *Management Decision, 43,* 761-771. doi: 10.1108/00251740510597761

Boyd, D. J., Grossman, P., Lankford, H., Loeb, S., Michelli, N. M., & Wyckoff, J. (2006). Complex by

Design: Investigating Pathways into Teaching in New York City Schools. *Journal of Teacher Education, 57,* 155-166. doi: 10.1177/0022487105285943

Carnine, D. W., Silbert, J., Kame'enui, E. J., & Tarver, S. G. (2010). *Direct Instruction Reading* (5thed.). Boston, MA: Merrill.

Cochran-Smith, M. & Zeichner, K. M. (2005). *Studying teacher education.* New York: Routledge.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nded.). Upper Saddle River, NJ: Pearson.

Darling-Hammond, L., & Bransford, J. (2005). *Preparing teachers for a changing world*. San Francisco: Jossey-Bass.

Fisher, D. & Frey, N. (2008). *Better learning through structured teaching: A framework for the gradual release of responsibility.* Alexandria, VA: Association for Supervision and Curriculum Development.

Gansle, K. A., Burns, J. M., & Noell, G. (2011). *Value-added assessment of teacher preparation programs in Louisiana: 2007-08 to 2009-10, Overview of 2010-11 results.* Retrieved from http://www.regents.doa.louisiana.gov/assets/docs/TeacherPreparation/2010-11ValueAddedAssessmentOverviewofResultsNov212011.pdf

Gansle, K. A., Noell, G. H., & Burns, J. M. (2012). Do student achievement outcomes differ across teacher preparation programs? An analysis of teacher education in Louisiana. *Journal of Teacher Education, 63,* 304-317. doi:10.1177/0022487112439894

Gansle, K. A., Noell, G. H., Knox, R. M., & Schafer, M. J. (August, 2010). Value-added Assessment of Teacher Preparation in Louisiana: 2005-2006 to 2008-2009. Baton Rouge, LA: Louisiana State University, Department of Psychology.

Gansle, K. A., Noell, G. H., VanDerHeyden, A. M., Slider, N. J., Hoffpauir, L. D., Whitmarsh, E. L., & Naquin, G. M. (2004). An examination of the criterion validity and sensitivity to brief intervention of alternate curriculum-based measures of writing skill. *Psychology in the Schools, 41,* 291-300. doi:10.1002/pits.10166

Gilham, S. A., Lucas, W. L., & Sivewright, D. (1997). The impact of drug education and prevention programs: Disparity between impressionistic and empirical assessments. *Evaluation Review, 21,* 589-613. doi: 10.1177/0193841X9702100504

Glazerman, S., Loeb, S., Goldhaber, D., Raudenbush, S., Staiger, D., & Whitehurst G. (2010). *Evaluating teachers: The important role of value-added.* Washington, DC: The Brown Center on Education Policy at Brookings.

Greenwood, C. R., Arreaga-Mayer, C., & Carta, J. J. (1994). Identification and

translation of effective teacher-developed instructional procedures for general practice. *Remedial and Special Education, 15*, 140-151. doi:10.1177/074193259401500302

Hanson, G. R. (1988). Critical issues in the assessment of value added in education. *New Directions for Institutional Research, 59,* 53-67. doi: 10.1002/ir.37019885907

Harris, D. (2009). The policy uses and policy validity of value-added and other quality measures. In D. Gitomer (Ed.). *Measurement issues and assessment for teaching quality* (pp. 99-130). Thousand Oaks, CA: Sage.

Heitin, L. (2011). Evaluation system weighing down Tennessee teachers. *Education Week, 31*(8), 1, 14-15.

Hershberg, T., Simon, V., & Lea-Kruger, B. (2004). The revelations of value-added: An assessment model that measures student growth in ways that NCLB fails to do. *School Administrator, 61*(11), 10–13.

Isenberg, E., Hock, H., & Mathematica Policy Research, Inc. (2011). *Design of Value-Added Models for IMPACT and TEAM in DC Public Schools, 2010-2011 School Year. Final Report*. Princeton, NJ: Mathematica Policy Research, Inc.

Jenkins, J.R., Deno, S.L., & Mirkin, P.K. (1979). Measuring pupil progress toward the least restrictive environment. *Learning Disability Quarterly, 2*, 81–92.

Juel, C., & Minden-Cupp, C. (2000). Learning to read words: Linguistic units and instructional strategies. *Reading Research Quarterly, 35*, 458-492. doi:10.1598/RRQ.35.4.2

Lasley, T., Siedentop, D., & Yinger, R. (2006). A systemic approach to enhancing teacher quality: The Ohio model. *Journal of Teacher Education, 57*, 13-21.

Locke, E. A., & Jain, V. K. (1995). Organizational learning and continuous improvement. *International Journal of Organizational Analysis, 3*, 45-68. doi: 10.1108/eb028823

Louisiana Administrative Code, Title 28, Education, Part XLV., Bulletin 996, *Standards for approval of teacher and/or educational leader preparation programs*, February 2011.

Louisiana Department of Education (2008a). i*LEAP 2008 technical summary*. Retrieved from http://www.louisianaschools.net/lde/uploads/13125.pdf

Louisiana Department of Education (2008b). *LEAP GEE 2008 technical summary*. Retrieved from http://www.louisianaschools.net/lde/uploads/13306.pdf

McCaffrey, D. F., Lockwood, J. R., Kortez, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability.* Santa Monica, CA: RAND Corporation.

National Center for Education Statistics. (2007). Mapping 2005 state proficiency standards onto the NAEP

scales (NCES 2007–482). Washington, DC: Government Printing Office.

National Center for Education Statistics (2011). The Nation's report card: Reading 2011 (NCES 2012–457). Washington, DC: Institute of Education Sciences, U.S. Department of Education. Retrieved June, 2012 from http://nces.ed.gov/nationsreportcard/pdf/main2011/2012457.pdf.

National Reading Panel (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: National Institute of Child Health and Human Development.

Noell, G. H., & Burns, J. L. (2006). Value-added assessment of teacher preparation: An illustration of emerging technology. *Journal of Teacher Education, 57*, 37–50. doi: 10.1177/0022487105284466

Noell, G. H., Gansle, K. A., Patt, R. M., & Schafer, M. J. (2009). *Value-added assessment of teacher preparation in Louisiana: 2005-2006 to 2007-2008.* Baton Rouge, LA: Louisiana State University Department of Psychology. Retrieved from Louisiana Board of Regents website: http://www.laregentsarchive.com/Academic/TE/2009/2008-09VATechnical (8.24.09).pdf

Noell, G. H., Porter, B. A., Patt, R. M., & Dahir, A. (2008). *Value-added assessment of teacher preparation in Louisiana: 2004-2005 to 2006-2007.* Baton Rouge, LA: Louisiana State University Department of Psychology. Retrieved from Louisiana Board of Regents website: http://www.laregentsarchive.com/Academic/TE/2009/2008-09VATechnical(8.24.09).pdf

Palincsar, A. S. & Herrenkohl, L. R. (2002). Designing collaborative learning contexts. *Theory into Practice, 41*, 26-32.

Papay, J. P. (2010). Different tests, different answers: The stability of teacher value-added estimates across outcome measures. *American Educational Research Journal, 48*, 163-193. doi: 10.3102/0002831210362589

Raudenbush, S. W. (2004). What are value-added models estimating and what does this imply for statistical practice? *Journal of Educational and Behavioral Statistics, 29*, 121-129. doi:10.3102/10769986029001121

Reusser, J., Butler, L., Symonds, M., Vetter, R., & Wall, T. J. (2007). An assessment system for teacher education program quality improvement. *International Journal of Educational Management, 21*, 105-113. doi: 10.1108/09513540710729908

Rice, J. K. (2003). *Teacher quality: Understanding the effectiveness of teacher attributes.* Washington, DC: Economic Policy Institute.

Ross, S. G., & Begeny, J. C. (published online September, 2014). An examination of treatment intensity

with an oral reading fluency intervention: Do intervention duration and student–teacher instructional ratios impact intervention effectiveness? *Journal Of Behavioral Education*. doi:10.1007/s10864-014-9202-z

Sawchuk, S. (2011). Wanted: ways to assess the majority of teachers. *Education Week*, *30*(19), 1, 15.

Schroeder, D. M., & Robinson, A. G. (1991). America's most successful export to Japan: Continuous improvement programs. *Sloan Management Review, 32(3),* 67-81.

Shanahan, T. & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review,* 78, 40-59.

Stanovich, K. E. (1994). Constructivism in reading education. *Journal of Special Education, 28,* 259-274.

Stata, R. (1989). Organizational learning: The key to management innovation. *Sloan Management Review, 30(3),* 63-74.

Tekwe, C. D., Carter, R. L., Ma, C., Algina, J., Lucas, M. E., Roth, J., Ariet, M., Fisher, T., & Resnick, M. B. (2004). An empirical comparison of statistical models for value added assessment of school performance. *Journal of Educational and Behavioral Statistics, 29,* 11-37. doi:10.3102/10769986029001011

Viadero, D. (2008). Scrutiny heightens for "value added" research methods. *Education Week*, *27*(36), 1, 12-13.

Wilson, S., Floden, R. Y., & Ferrini-Mundy, J. (2001). *Teacher preparation research: Content knowledge, gaps and recommendations.* Washington, D.C: U.S. Department of Education.

Kristin A. Gansle is a professor of education and serves on the Special Education Program faculty of the College of Human Sciences and Education at Louisiana State University. Her research interests include evaluation of teacher preparation programs, improvement of teacher preparation programs, and assessment of academic skills for children with high-incidence disabilities. **Dr. Gansle** is the corresponding author for this article and can be contacted at ckgansle@lsu.edu.

Gerlinde Grandstaff-Beckers is an assistant professor in the College of Education, Department of Teaching & Learning, at Southeastern Louisiana University. She has eleven years of teaching experience in the public schools. Her areas of expertise are content literacy and at-risk and struggling readers. She has several publications and has conducted numerous presentations at the national, regional, and state level.

Angelle Stringer is an instructional coach in Ascension Parish Schools. Her interests are encouraging and mentoring teachers to provide high quality and challenging instruction for all students. Dr. Stringer has been a classroom teacher, college instructor, alternative certification program director, and teacher mentor. She is a perpetual student and plans to continue further studies in curriculum and instruction education and English literature.

Nancy Roberts is the founder and Executive Director of the Louisiana Resource Center for Educators. With the support of the community and the state, Roberts has taken a small organization with a budget of $30,000 a year to a statewide service for teachers with an annual budget of $2.2 million. Ms. Roberts holds Bachelor of Science and Master's degrees in Education from LSU.

Jeanne M. Burns has previously taught and served in district administrative roles in Florida and Louisiana and has been a faculty member at Stetson University and Southeastern Louisiana University. She is currently the Associate Commissioner for Teacher and Leadership Initiatives for the Louisiana Board of Regents. She has directed and co-directed millions in grant programs to improve teacher preparation and educational leadership preparation in Louisiana, all of which are initiatives that have been jointly supported by the Governor, Board of Regents, Board of Elementary and Secondary Education, and Louisiana Department of Education.

George H. Noell is a professor of psychology at Louisiana State University. His research has focused on improving the quality and implementation of treatment plans for children in need of psychological services. Dr. Noell has been engaged in educational policy work that has sought to develop data analytic systems that could support time sensitive and long term policy making. Dr. Noell's scholarship has been acknowledged by election to scholarly societies, awards, research journal editorial board appointments, and an appointment as editor-in-chief.

Appendix

Table 1

Performance Levels for Teacher Preparation Programs

Level 1	Programs whose effect estimate is above the mean effect for experienced teachers by its standard error of measurement or more. These are programs for which there is evidence that new teachers are more effective than experienced teachers, but this is not necessarily a statistically significant difference.
Level 2	Programs whose effect estimate is above the mean effect for new teachers by its standard error of measurement or more. These are programs whose effect is more similar to experienced teachers than new teachers.
Level 3	Programs whose effect estimate is within a standard error of measurement of the mean effect for new teachers. These are programs whose effect is typical of new teachers.
Level 4	Programs whose effect estimate is below the mean effect for new teachers by its standard error of measurement or more. These are programs for which there is evidence that new teachers are less effective than average new teachers, but the difference is not statistically significant.
Level 5	Programs whose effect estimate is statistically significantly below the mean for new teachers.

Formal Assessment System for Reading Instruction

Teacher: _____ Subject/Grade: _____
School: _____ Date/Time: _____
Evaluator's Name: _____ Title of Lesson: _____

1 Emerging = inadequate performance or progress, needs guidance
2 Acceptable = adequate/acceptable progress or performance with potential for improvement
3 Proficient = progress/performance exceeds normal expectations of a beginning teacher
N/O = not observed yet/not known
N/A = not applicable

Reading Instruction	**SCORE**	**COMMENTS**
Knowledge of Components		
Vocabulary Development		
Comprehension		
Other (phonics, phonemic awareness, fluency, writing)		

Engagement of Students

Provides hands-on activities		
Variety, meaningful		

Effective Grouping

Management / Control		
Monitors Engagement		

Planning

Preparation		
Implementation		

Evaluating

Documents Mastery		
Adapts Instruction		

Classroom Technique

Enthusiasm / Motivation		
Smooth Transitions		

Figure 1. Instrument used in direct observation of teacher behavior

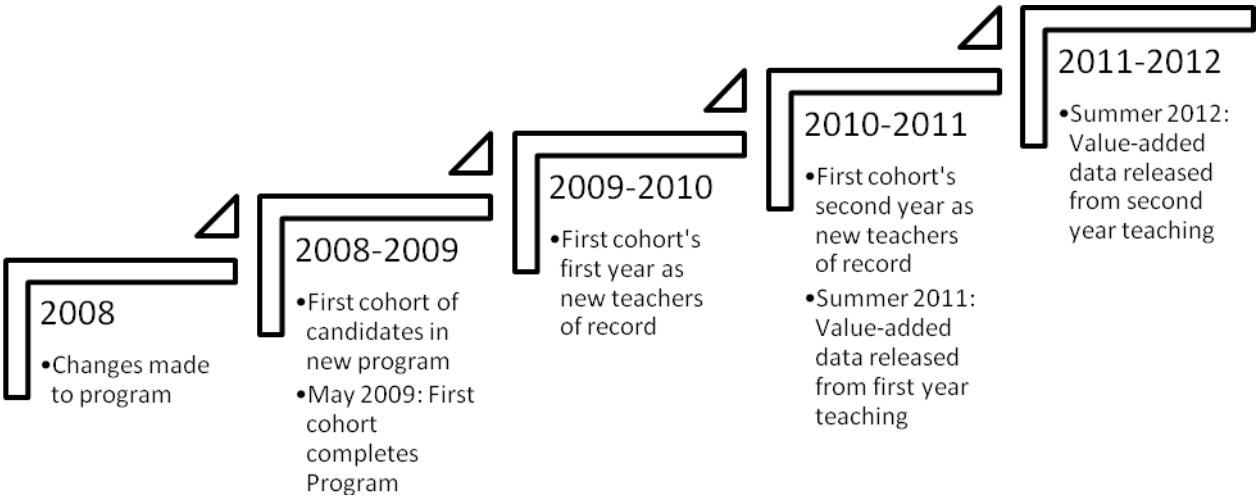
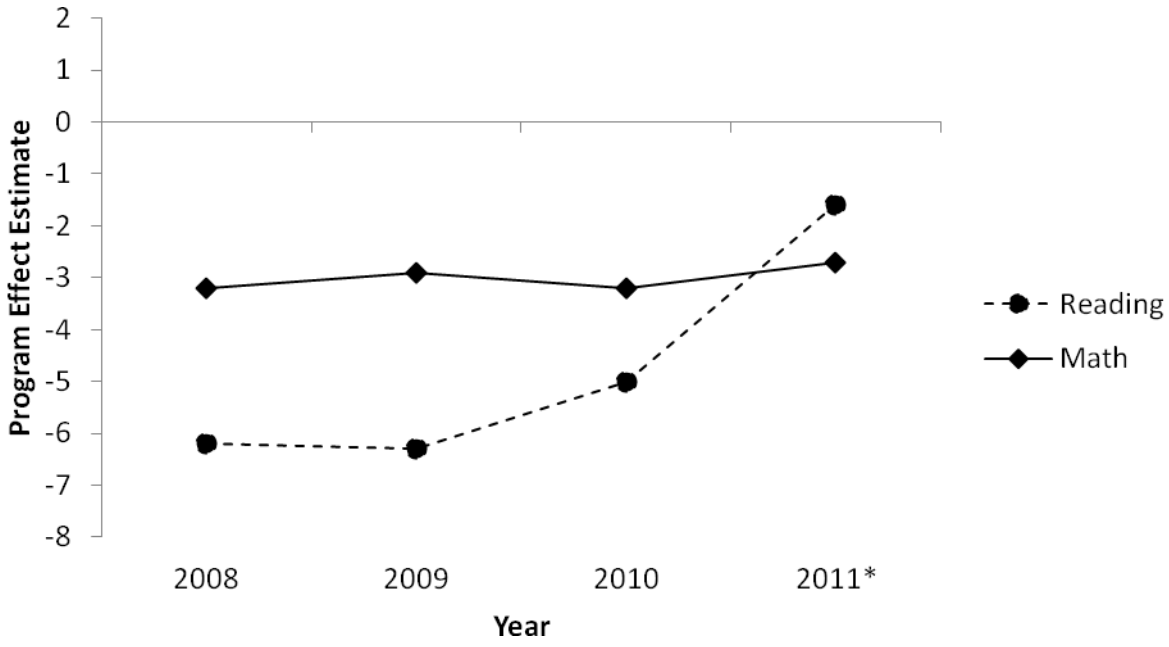


*Figure 2.*Sample timeline for release of value-added report from time of program change



Note. *2011 estimate based on original hierarchical linear model.

Figure 3. LRCE program effect estimates over time for reading and mathematics