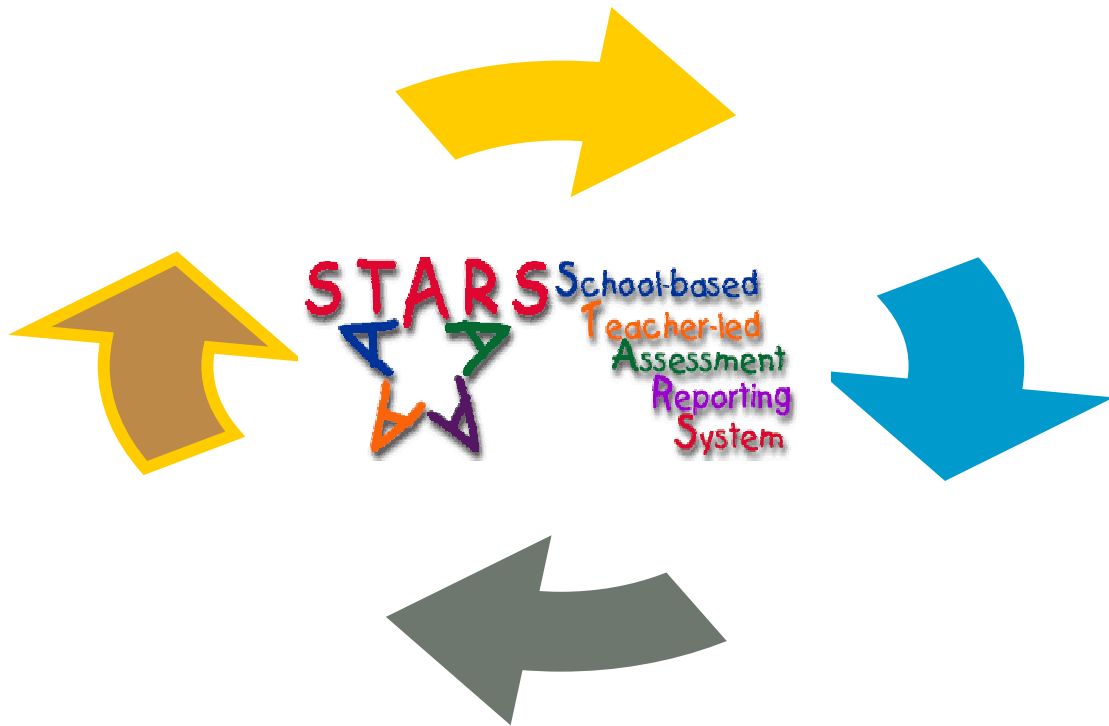# Charting STARS:

New Conversations

Year Three Report
Comprehensive Evaluation of Nebraska's School-based, Teacher–led Assessment and Reporting System (STARS)

Written by Chris Gallagher, Ph.D.
Coordinator and Principal Investigator, STARS Comprehensive Evaluation
Associate Professor of English
202 Andrews Hall
University of Nebraska-Lincoln
Lincoln, NE 68588-0333
402.472.1835
cgallagher2@unl.edu

Submitted to Douglas Christensen, Commissioner of Education

August 2004

Charting STARS: New Conversations

Year Three Report
STARS Comprehensive Evaluation

Executive Summary

This third annual report of the Comprehensive Evaluation Project (CEP) includes several studies:

- **An interview study** conducted by researchers at UNL under the direction of Chris Gallagher. This study is part of a larger, three-year interview study. All eight districts participating in this study were also participants in 2002-03, and two of those were also participants in 2001-2002. This year's interview study focused on how schools are *using data.*

- **Two quantitative studies**: examinations of 2001 and 2003 CRT and NRT data generated by STARS; *Study A* was conducted by Dr. Jody Isernhagen (University of Nebraska-Lincoln) and Dr. Leon Dappen (University of Nebraska-Omaha) and *Study B* was conducted by Dr. Gregory Schraw and Dr. Keith Zvoch of the Center for Evaluation and Assessment at the University of Nevada-Las Vegas.

- **A sufficiency study**, conducted by the Buros Institute for Assessment and Outreach with assistance from the CEP, which examines the potential of NRTs and CRTs to classify students into four performance categories.

- **A survey of elementary educators** regarding local assessment processes.

- **An interview study** of principals, conducted by researcher Elaine Specht, regarding school leadership

Among the most important findings from this years' studies are these:

- ♦ Key indicators of school and student performance – including portfolio ratings and percentages of students proficient on standards – show gains from 2001-2003. (quantitative studies)
- ♦ Assessment quality on both norm-referenced and criterion-referenced assessments in use in Nebraska is not where it needs to be in terms of sufficiency. (sufficiency study)

♦ Trends identified in last year's report with respect to curriculum and instruction, leadership, professional development, and local standards, assessment and accountability (SAA) processes continue this year. (interview studies)

♦ Schools are at various stages in developing more data-informed, research-based SAA processes in which assessment is tied to curriculum and instruction for the purpose of school improvement, but most are making progress. Schools are using data for multiple internal (formative) and external (summative) purposes. (interview studies)

♦ Though some resistance to STARS remains, many educators harbor fairly positive perceptions of how assessment is going in their schools. (interview studies, survey)

♦ Key challenges – including time, workload, and educators' attitudes – continue. New challenges – for instance, dealing with federal accountability and the attractions of "quick fix" assessment options – have emerged. (interview studies, survey)

In last year's report, we asked the question, "Is STARS sustainable?" At that time, we claimed that STARS was at a "crossroads" marked both by signs of progress and continuing challenges. Thus, it was too early to answer our key question.

It is still too early to answer that question definitively. As we suggest above, old challenges remain while new ones emerge. Moreover, resistance and skepticism remain both internally and externally.

However, supporters of STARS will find in this report many reasons for optimism. On the national scene, as we have suggested, STARS continues to garner a great deal of attention. The school reform organization Fairtest, for instance, identified Nebraska as the only state practicing what it calls "authentic accountability."[1] Moreover, the federal government has now formally approved STARS as the state's assessment and accountability system.

Perhaps more importantly, the conversations in Nebraska districts have changed over the three years of our study. We mean this in two ways. First, the *language* used by Nebraska educators has changed. We have witnessed enormous growth in assessment literacy, especially among teachers, many of whom now comfortably "talk assessment." Second, the question that Nebraska educators ask about STARS has moved from "Why do we have to do this?" to "Can it work?" to "How can we make it work for *everyone*?" To be sure, some educators are asking new questions out of resignation; they understand that whether they like it or not, STARS is here to stay. But a growing number of educators are asking new questions because they have come to believe that Nebraska's local discretion model is the right thing to do for Nebraska schools, communities, and children.

---

[1] Monty Neill, Lisa Guisbond, and Bob Schaeffer. *Failing Our Children*. (Cambridge, MA: The National Center for Fair and Open Testing, 2004).

Like the schools, the NDE – and particularly its Statewide Assessment office – has become increasingly sophisticated in its policies, practices, and procedures. It is moving into a position in which it can capitalize on the strengths of the system and respond effectively and efficiently to both stubborn and emergent challenges. As we discuss in the conclusion, we believe the NDE has one more important step to take toward ensuring the sustainability of STARS, but there can be little doubt that it is in a far stronger position at this time than it has been in the past.

Recommendations:
1. Continue to use multiple measures of student performance.
2. Disseminate step-by-step processes and methods for districts to improve their local assessments, especially on sufficiency.
3. Continue to focus professional development offerings for teachers and principals on working with (i.e., interpreting and using) data.
4. Use local talent to model the building of data processes and the integration of data into school improvement.
5. Target assessment literacy opportunities to eighth and eleventh grades.
6. Require all districts to have an assessment coordinator.
7. Continue to educate various constituencies about the nature and purposes of STARS. Focus special attention on local media and educators.
8. Sponsor a "leaders of learning" council or academy for principals and superintendents.
9. Continue to work with the measurement community to develop appropriate metrics for the kind of data generated by STARS.
10. Research the kinds and quality of support schools receive from various outside sources, including ESUs, consortia, higher education, and the NDE.
11. Research the effects on schools of using assessments designed elsewhere (consortia, ESUs, higher education, districts).
12. Partner with schools to experiment with alternative, creative uses of time.
13. Conduct a careful, honest self-study of the NDE. Strive to integrate structurally those parts of its mission that are integrated in its philosophy.

"It is the policy of the State Board of Education that the information available from the Nebraska standards, assessment, and reporting system should be used primarily to promote school improvement for improving student learning." – State Board of Education Policy on Using School Performance Information (Approved April 4, 2003)

"Effective school accountability is about using new data as a roadmap to school improvement and higher student learning." -- Commissioner Doug Christensen

"You gotta have data." – Nebraska Educator[2]

This is the third annual report of the Comprehensive Evaluation Project (CEP), an independent evaluation of Nebraska's School-based, Teacher-led Assessment and Reporting System (STARS). The CEP was contracted between the Nebraska Department of Education (NDE) and the University of Nebraska-Lincoln Teachers College (now the College of Education and Human Sciences) in 2001. The CEP is supported jointly by the NDE, the College of Education and Human Sciences, and UNL's College of Arts and Sciences. Dr. Chris Gallagher (Associate Professor, English) serves as Coordinator and Principal Investigator.

<u>Year Three Report</u>

The Nebraska Department of Education (NDE) dubbed 2003-2004 "The Year of Accountability." The Nebraska State Board of Education adopted a new policy regarding school performance data and accountability (quoted above). The federal government continues to scrutinize state assessment and accountability systems for compliance with the provisions of No Child Left Behind.[3] Local and national audiences are watching closely as the state's unorthodox School-based, Teacher-led Assessment and Reporting System unfolds.[4]

In response to these internal and external pressures on STARS, the Comprehensive Evaluation Project (CEP) was radically expanded in 2003-2004. This expansion included a variety of partnerships with researchers conducting qualitative and quantitative research. We also became part of a national conversation about the

---

[2] As in our previous reports, we will use the term "educator" to refer to both teachers and administrators.
[3] In June 2004, the federal government formally approved STARS as the state's assessment and accountability system.
[4] See, for instance, Monty Neill, Lisa Guisbond, and Bob Schaeffer. *Failing Our Children*. (Cambridge, MA: The National Center for Fair and Open Testing, 2004); Tracy Dell'Angela, "Nebraska Shuns State Tests," *Chicago Tribune*. April 5, 2004.

validation of accountability systems as part of a multi-state consortium sponsored by the Council of Chief State School Officers.

Readers may wish to review the Validation Framework that has been collaboratively designed by Pat Roschewski, Assessment Director for the NDE, and the CEP (Appendix A). It places the various studies included in this report in the context of a long-term validation of STARS.

As with the previous reports, we have endeavored to develop a document that is accessible and helpful to multiple audiences. Wherever possible, we have used nontechnical language. Readers who wish to know more about the technical details of the studies included in this report may contact Chris Gallagher.

The report is divided into several sections. The first is devoted to an **interview study** conducted by researchers at UNL under the direction of Dr. Gallagher. This study is part of a larger, three-year interview study. All eight districts participating in this study were also participants in the 2002-2003, and two of those were also participants in 2001-2002. This year's interview study focused on how schools are **using data**.

We then present sections on several other studies. Each section follows the same pattern: after a brief introduction, we present the executive summary of the study, written by the respective researchers. We then place the study in the context of the Comprehensive Evaluation, discussing *findings*, *implications*, and *recommendations*.

The studies included in this report include the following:

♦ **Two quanitative studies**: examinations of 2001 and 2003 CRT and NRT data, as well as portfolio ratings; *Study A* was conducted by Dr. Jody Isernhagen (University of Nebraska-Lincoln) and Dr. Leon Dappen (University of Nebraska-Omaha) and *Study B* was conducted by Dr. Gregory Schraw and Dr. Keith Zvoch of the Center for Evaluation and Assessment at the University of Nevada-Las Vegas

♦ **A survey of elementary educators** regarding local assessment processes

♦ **A sufficiency study**, conducted by the Buros Institute for Assessment and Outreach with assistance from the CEP, which examines the potential of NRTs and CRTs to classify students into four performance categories

♦ **An interview study** of principals, conducted by researcher Elaine Specht, regarding school leadership

We conclude this report with a progress report and a look ahead.

Overview

In last year's report, we asked the question, "Is STARS sustainable?" At that time, we claimed that STARS was at a "crossroads" marked both by signs of progress and continuing challenges. Thus, it was too early to answer our key question.

It is still too early to answer that question definitively. Old challenges (time, workload) remain while new ones (dealing with federal accountability, the attractions of "quick-fix" assessment options) emerge. Moreover, resistance and skepticism remain

both internally and externally. However, supporters of STARS will find in this report many reasons for optimism.

On the national scene, as we have suggested, STARS continues to garner a great deal of attention. The school reform organization Fairtest, for instance, identified Nebraska as the only state practicing what it calls "authentic accountability."[5] Moreover, the federal government has now given its final approval to STARS.

Perhaps more importantly, the conversations in Nebraska districts have changed over the three years of our study. We mean this in two ways. First, the *language* used by Nebraska educators has changed. We have witnessed enormous growth in assessment literacy, especially among teachers, many of whom now comfortably "talk assessment." Second, the question that many Nebraska educators ask about STARS has moved from "Why do we have to do this?" to "Can it work?" to "How can we make it work for *everyone*?" To be sure, some educators are asking new questions out of resignation; they understand that whether they like it or not, STARS is here to stay. But a growing number of educators are asking new questions because they have come to believe that Nebraska's local discretion model is the right thing to do for Nebraska schools, communities, and children.

Like the schools, the NDE – and particularly its Statewide Assessment office – has become increasingly sophisticated in its policies, practices, and procedures. It is moving into a position in which it can capitalize on the strengths of the system and respond effectively and efficiently to both stubborn and emergent challenges. As we discuss in the conclusion, we believe the NDE has one more important step to take toward ensuring the sustainability of STARS, but there can be little doubt that it is in a far stronger position at this time than it has been in the past.

---

[5] Monty Neill, Lisa Guisbond, and Bob Schaeffer. *Failing Our Children*. (Cambridge, MA: The National Center for Fair and Open Testing, 2004).

- our supporters in UNL's Teachers College, including Dean Marjorie Kostelnik, Michael Toland at the NEAR Center, Associate Dean Jim Walter, and Professor Ali Moeller

- our supporters in UNL's College of Arts and Sciences and English Department, including Dean Richard Hoffmann, Chair of English Linda Pratt, and Linda Rossiter

Interview Study: Using Data

"You gotta have data" – *Nebraska Educator*

Our methodology and protocols for conducting the interview study this year were identical to those employed last year (see Year Two Report, Appendix D for a full description of methodology). However, our focus this year was considerably narrower than in previous years. We visited fewer districts (eight) and asked more in-depth questions about a single topic: using data. This emphasis reflects our sense that the schools in our study – like their cohorts across Nebraska – are at a critical juncture: they are operating in a data-rich environment, faced with the task of learning how to analyze, respond to, and use the data. Our questions seek to shine a light on this crucial part of the standards, assessment, and accountability (SAA) process about which we wrote in considerable detail last year.

Our research questions may be distilled to four:

♦ In these schools, who is responsible for collecting, interpreting, sharing, and using assessment data?
♦ How important are the data? For whom?
♦ What are the schools learning from the data?
♦ What are schools doing with the data? How are they responding?

(See Appendix D for a copy of the interview questions we designed. We piloted the interview questions and revised them according to what we learned from read-aloud protocols and debriefings with educators who participated in the pilot.) We will use these four questions to structure this report. We then discuss continuing challenges and, finally, recommendations. First, we describe this year's sample and study participants.

The Sample

We conducted interviews in eight schools in eight Nebraska districts. In total, we conducted 46 structured, open-ended interviews: 22 with teachers, 6 with assessment coordinators, 8 with principals, 7 with superintendents, and 4 with local school board members.[6] Some questions were repeated across the interviews for the purpose of comparison; others were tailored to the position held by the participant. The findings here are based on analysis of audiotapes as well as detailed researcher notes from each interview.

Although the sample size is small, we aimed for a "maximum variation sample," with the intent of learning about as many different situations within the state as possible.

---

[6] The total is 47 because one superintendent was also a principal.

We chose schools based on the following characteristics: school size, district size, geographic region, educational level, and student demographics.

All eight schools included in this year's study were also included in last year's study. Of those, two were also included in the previous year's interview study.

The following is a brief composite portrait of our participating schools:

- *Number of schools, districts*: 8 schools, 8 districts
- *Size of district*: 3 small, 3 medium, 2 large
- *Type of community*: 3 rural, 4 suburban, 1 urban
- *Geographic region*: 4 eastern, 2 central, 2 western
- *Type of school*: 3 elementary, 3 middle, 2 high school
- *Student diversity*: 5 schools with moderate or high racial diversity
- *Free/reduced lunch*: 4 schools with more than 50% of students receiving free/reduced lunch
- *Mobility*: 3 schools with mobility rates of 25% or more
- *Ratings*: 2003 assessment quality = 7 high, 1 low; 2003 student performance (reading) = mixed
- *Statewide Writing Results*: mixed
- *AYP Classification*: 6 Not Met, 1 Met, 1 N/A

In these schools, who is responsible for collecting, interpreting, sharing, and using assessment data?

The schools we visited are at different stages of implementing what we will call a "data process" – that is, a process for collecting, interpreting, sharing, and using assessment data. One school does very little with assessment data and has no defined process for doing so. In contrast, three schools have very mature data processes; one, for instance, was described this way: "We do an umbrella effect on everything that we do. Individual teaching strategies are always underneath a departmental strategy, which is underneath a building strategy, which is underneath a district strategy." The other schools are shifting toward using data regularly for instruction and school improvement. In general, the schools are using data more regularly and systematically than they have in the past.

We found a variety of "responsibility models" at work in these schools. In particular, different people or organizations were identified as the driving force behind the data process. Educators in one school reported that their ESU coordinates their data process; in four schools, an assessment coordinator or curriculum director is primarily responsible; in two schools, that responsibility falls to teachers; and in one school it falls to the principal. In all but one school, the assessment or curriculum person (for the school or the district) is actively involved in all phases of the data process. The exception is a school in which there is no assessment coordinator either for the building or the district. In this case, the "extra load" (as the principal put it) falls to the building administrators. This is the school in which we detect no defined data process.

Most schools use a team approach to the data process. The team may consist of "key staff," and it may be a formal body, such as a "curriculum council." Or the team may consist of teams of teachers, often organized by department. The following are descriptions of typical data processes:

[W]e go through the classroom teachers and the different content area groups, and obviously they have some very good ideas too on what needs to be done with the data. We have had a couple of data analysis opportunities this year for teams of teachers and/or whole staff. And we're just ready to move that data analysis more from the district level to the building level and have some building level action plans developed under the School Improvement Plan.

Instead of us [administrators] giving them [teachers] the information, I still provide the paper that has the scores on it and put it in a usable format but we let them evaluate the data. So, somewhere in September, the beginning of October, each department makes a report to the whole staff: "This is the data from this year. We're really pleased with this. We really think we need to work on this. This will be part of our building action plan." And then each department writes a piece to the building action plan. We set a goal as a building and then they all write how they're going to address that goal in their departments, based upon their data [and] teacher observation.

In general, administrators – principals and superintendents – are responsible for sharing the data with several audiences, including teachers, local school boards, and the community. They craft their message carefully for each group, typically accentuating the positive, especially for the community and board. Presentations to teachers generally involve significant amounts of data; other groups receive "talking points."

Teachers are becoming more engaged in the data process in most of these schools. They are increasingly responsible not only for using data but also for interpreting them. For example:

[T]he teachers get together. We have curriculum meetings. We review all the data. We actually have…teachers are contracted an extra two days following the end of the school year, which is kind of our data retreat time where we look at everything. And then [we] identify areas that we need to improve or curriculum changes that we need to make.

One superintendent offered an extensive rationale for involving teachers in all phases of the data process. He insisted, first of all, that "the success of our entire program lies on the backs and the shoulders of the classroom teachers. So, they are the ones that interpret the data because they have the immediate feedback." He then explained how and why his district put teachers at the center of their process:

[I]nitially it was really interesting because we allowed the teachers to all go their own way in trying these different things. And you know, some of them became really frustrated because those of us who are concrete sequentialists, we like to be told what to do and how to do it…So, for as many different teachers as we have, we had different ways of doing things. Well, from that, we were able to come back year two and look at all the variety. And I was able to say to the teachers, if I had told you what to do, I would have never thought of doing this or never thought of doing that. And we were able to take all of…the best practice scenarios that were taking place in the various classrooms and put those together. And now we are coming up with a – I mean, we kind of did it backwards – coming up with a district way of doing things.

Interestingly, a superintendent in a different part of the state described a teacher-led process that he similarly described as "backwards":

> We started by the central office driving that [data] down to administrators and then administrators essentially handing it to their building staff. Now, we've evolved to the point – and it's taken us several years to get there – but our school improvement is driven by our classroom teachers. And it comes back up. So we give them the data. If you teach for us, I give you the data, you look at where your kids performed in the third grade and where they met the expectations…the standards…and where they did not. And then the expectation I have of you as a classroom teacher is you're going to tell me as a building administrator how I'm going to tackle the fact that our kids don't have the mechanics of punctuation…or whatever the case may be. So then it comes back up and then it becomes a building goal and then it comes up the other way to the district. So…maybe do it a little bass-ackwards, but I think the people that are closest to the students need to drive that instruction and instruction will improve.

As we have reported in previous reports, this kind of "bottom-up," teacher-led approach to assessment, curriculum, and instruction is more and more common in Nebraska schools.

By contrast, in three schools, teachers are not regularly interpreting or using data. In fact, in these schools, teachers are cut off from the data process almost entirely. As one superintendent admitted, "for the most part, the school's assessment information did not get to the teachers." A teacher in this school confirmed this: "I wouldn't know whether or not the students are low in one of those [math] strands," adding, "which I think I should know because then we could work harder on it." Similarly, teachers in another school reported that they do not get data back from assessments. One teacher added, "that would be a good goal…to have everybody sit down and look."

In the school least far along in developing a data process, the principal attributed the problem at least in part to teachers' attitudes:

> We have to do it and it's giving us data that we can use to help drive the decisions that we make. But some people are at that angry stage and they're not ready to accept and go forward and use that data in a positive way. Because it really is giving us data. It's giving us so much data that it's overwhelming and I wish they'd give us more training in that area of how to interpret the data and what to do with the data after you have it.

We also observe that in schools that participate in large tests – district CRTs, consortium or ESU tests – teachers know far less about what their data indicate than their counterparts in schools where smaller, building- and classroom-designed assessments are administered throughout the school year.

Finally, local school board members are generally recipients of assessment data, but they are not intimately involved in data processes, including making decisions about how to respond to data (except, perhaps, to approve district-level changes). A couple board members have very negative attitudes about data: "I'm not a promoter of data…My feeling is that we spend a lot of time collecting and disaggregating data…and [educators] could maybe be using that time better teaching children in the classroom." A board member elsewhere told us that "we have the principals do a very short report at a board meeting and…this is terrible to admit, but I really can't remember when the last one

was." She added, "They don't explain what the numbers really mean." For the most part, the board members we talked to were not able to talk with specificity about assessment results in their district, even though a couple sat on curriculum committees.

How important are the assessment data for the schools? For whom?

In only one school did participants offer mixed responses to our question about the significance of assessment data, and this is the school making the least progress on its data process. In the other schools, there is near unanimity that assessment data are important – but for different reasons.

Several educators noted that assessment data are important owing to *external* forces and pressures. These forces include the *state*:

- ♦ "It's important because it's our measuring device; our funding's apparently going to hinge on that. So yes, we have to take it serious."
- ♦ "Because of the necessity to do it, we pay quite a bit of attention to it."

Also, *district- and school-level* decisions hinge on assessment data:

- ♦ "I feel like my job is at stake. I feel like if those scores continue to be bad, they're going to say you're not doing what you need to do."
- ♦ "[W]e have been told that jobs and salaries will be based on that, so…Yeah."
- ♦ "If…the overall average doesn't come back at a certain level, we're told we have to change things. We're – I don't want to say threatened – but we're told that we'll be put on probation."

In addition, the *local community*, including the media, put pressure on the schools to take assessment data seriously:

- ♦ "[The community was] embarrassed by those scores."
- ♦ "It usually goes on the front page of the [newspaper]."
- ♦ "[U]ltimately, I think the reputation of our district and the view the public has of the effectiveness of our educational system will be at stake."[7]

In all these instances, educators take assessment data seriously because others take them seriously. In a couple of cases, teachers expressed a generalized but amorphous fear: "I have no idea what's at stake, but I know…I know it's bad, you know."

However, participants also talked about *internal* reasons for taking assessment data seriously. The most common internal reason was that data drive school improvement. Even a participant in the school that has no data process reported that data

---

[7] One board member insisted that the community does not view assessment results as important: "I don't get people calling me and saying, you know, why don't we do better or why did we do so good? I do get questions like, why didn't my kid play on the basketball team or in the game or something like that more than I would something like this." This view was the exception, however; most educators and board members reported that their communities do care about assessment results, even if they are not well informed about them.

are important because, he predicted, "we're going to be using it for school improvement."
Educators in other schools offered similar notions:

♦ "[J]ust using those results to guide instruction so we can improve student learning, is very important."
♦ "[T]he core teachers – some of them, not all of them even, but some of them – begin to see, 'Oh, this has a direct connection [to] me and how I'm doing.' And it raises the stake, you know. And when you're invested in that, then you're, 'OK, I really better do what you do to put your best foot forward.'"
♦ "It's extremely important. We use a variety of assessments; we don't just use one assessment to make determination. So our teachers have to use at least three sources of data to…draw conclusions and determine what the needs of our students are, so the STARS is one leg of that and a very important leg. It's a criterion-referenced part of it and it's used extensively."
♦ "[N]ot only is it important for individual schools, but we take that all the way down to…individual classrooms. One of the things we've got to get beyond – and this is state I believe is not there yet – we've got to get beyond the notion that the assessment is an indicator of teacher ability. I think it is a diagnostic tool that helps us identify where we need to improve. And improvement shouldn't be negative in connotation."

In each of these cases, assessment data are important to teachers and administrators because they have meaningful internal (formative) uses. Only a few participants did *not* talk about internal uses of data; several participants spoke *only* of internal uses.

Responses were more mixed when we asked how important assessment data, or tests themselves, are for *students*. Especially in schools with large CRTs, teachers tended to suggest that students are apathetic:

♦ "You get somebody that just, you know, fills in the circles and…this is not important to them."
♦ "[T]hey're not given how their results were and so, it's kind of, to them, meaningless. You know, they spend all this time and worry and try to do well, and then they don't ever find out how they did."
♦ "Students are numb with all these CRTs."

On the other hand, some students *do* take assessments and assessment data seriously, at the urging of their teachers:

♦ "I said…through probably two centuries…we've been educating kids in this country, with the thought that if you're not an educated citizen, you can't take care of our freedoms. And so we talk a lot about that and I'm hoping that I can get my kids to take not only my class more seriously, but all their other classes."
♦ "[Students] are reminded constantly the importance of that score and that they are a part of this whole number that's going to come back as far as what people think of our building, our district."

♦ "[Students] become, very knowledgeable…all the lesson plans are tied to standards. And then students are being taught those. You know, here is the standard that we're learning about. And then when they're assessed on it, they want to know how well they did. Even on the writing results…I'll be requested when the results come in. The kids will want me to come in to the classroom and explain how they did."

Students' investment in assessment results is often a positive phenomenon, as when they feel a sense of pride in their work – including, in one school, a sense of tribal pride. As some teachers pointed out to us, however, these kinds of expectations can place an enormous burden on students, especially in the lower grades. In any event, as one might expect, different students react in different ways to assessments and assessment data. Some care deeply; some care not at all. What seems clear, at any rate, is that students are increasingly *aware* of the pressures of assessment on their schools, their teachers, and themselves.

In sum, participants told us that assessment data *are* important for them and for their schools, if less so for their students. One participant articulated the importance of assessment data from multiple viewpoints:

> From district-level, it's extremely important because that's how we report to the state our proficiency of our students. At the building level it's important because once you start to collect years of information, you start to look at various classes. And you start to see maybe this class is really strong in mathematics, and this class isn't as strong. So maybe you can find out what the class who's strong, what they're doing. And then from…the classroom point of it, once you see your results of assessments, you can make instructional decisions. Maybe you need to reteach something. Maybe you change how you taught something. Maybe you need to emphasize something more. Maybe you did a great job and you don't need to change anything.

What are the schools learning from the data?

We asked educators what they were learning from their assessment data about their students, their instruction, and their curriculum (i.e., programs). Participants varied greatly in their knowledge about what the data show. To each question, one principal answered, simply, "Nothing." On the other end of the spectrum, two principals pulled up data on their computer and shared specific results with us during the interview.

Educators from each school were able to indicate general trends in assessment results – that writing scores were up, for instance, or that math scores were down. Beyond this, however, there is great variability in what these educators know about their school's assessment data.

Some participants were able to be specific about strengths and weaknesses in their students' performance. A principal, for instance, explained his results this way: "our pre-algebra skills and our algebra skills both are exceptional. Geometry skills are exceptional…Working with ratios, decimals, fractions, doing just the basic computation like multiplication and division and measurement would be the areas that we've been working on." (He then spoke with the same level of specificity about reading.) Another principal shared with the interviewer both spreadsheet data and "talking points" he had created to share with his teachers and students. He carefully and systematically talked

the interviewer through trend data in reading, mathematics, and writing, pointing out which areas (and at what grade levels) the students excelled and had difficulty. Finally, in one school, teachers and administrator alike pointed to the same results; for instance, in separate interviews, the five participants identified decimals, fractions, and percentages as problem areas for their middle-school students.

Most of these schools are beginning to disaggregate data and to identify achievement gaps among their students. Three schools find significant achievement gaps along poverty lines. For instance, one principal explained that "[i]n the beginning, when we wrote our goal at the beginning of the five years, the achievement gap was between minority and white students...We have done more research on this and it's really between students in poverty and students not in poverty." Two schools have identified racial achievement gaps. Most of these schools are coming to important realizations about various subgroups; for instance: "our kids at all levels are making progress [but] we are not substantially closing the gap [for immigrant children]."

Only seldom were participants able to talk about what they were learning about specific curricular programs they had instituted – and even then, the conclusions seemed vague and anecdotal. A superintendent, for instance, suggested that "we're seeing some documentation of some of what we've decided before; the decisions that have been made in terms of reading instruction – in our case, Four Blocks – [and] consistently sticking to that structure." In most cases, educators told us that the assessment results indicated that their purchased programs in reading or math were "working" – though they did not specify how they came to this conclusion.

Teachers, in particular, talked about the data confirming what they already knew about their students. A typical teacher comment: "I'm not seeing anything that surprises me." These teachers generally view assessment results as "another piece of evidence" – "a lot of what STARS says, I guess, pretty much will back up what you already know. You spent the whole year with this group. You pretty much know where things are standing." This sentiment seems most pronounced in schools where assessment is embedded into the everyday practice of the school.

Finally, many participants indicated that they have been examining trend data for only a short time and so cannot speak conclusively and in detail about school- or district-wide trends. They all planned to continue and enhance this focus, however.

What are the schools doing with the data? How are they responding to what they learn?

As we have indicated, most of these schools are developing more sophisticated data processes. As one superintendent put it, "We all used to give the standardized achievement test and get it back and hope it didn't look too bad and put it away. And we never really thought about it as a diagnostic kind of thing." Today, however, schools are using data for several purposes. In the schools we visited, we saw educators

- developing and refining their data processes
- developing and refining their curriculum
- developing and refining classroom instruction
- developing and refining their assessments
- developing and refining their professional development programs

These trends are very similar to those identified in last year's report; in particular, the schools (save one) are making significant progress on integrating assessment into school improvement.

Data Processes

One school is only beginning to develop its data process, and it faces serious challenges in doing so. As the superintendent of that district suggests, his schools are working toward integrating assessment into curriculum and instruction but "right now our perception is that we can't necessarily do that." His staff, he says, is "being overwhelmed by all the data and not really getting a lot of direction on what we're supposed to do with it." As a result, this district plans to use a new online assessment, with the notion that this assessment will help it manage and interpret its data.[8]

A principal in another school informed us that there was no data process there, either: "I knew you were coming. I asked our staff. We do provide the data to them. Not one soul uses it." But this claim was contradicted by other educators in that building. The assessment coordinator for the school, for instance, said, "we're beginning to graph and interpret a lot more data in order to make good decisions as part of our school improvement action plans." As a result, she added, "our teachers are beginning to say, 'This is what the data shows me. This is the area of weakness. What I want is an instructional strategy so I know what to do about it." The teachers in this building confirmed that they do interpret the data and attempt to adjust curriculum and instruction to what they learn (though not always happily; morale is a serious issue in this school). This school's data process is clearly underdeveloped – but it is just as clearly beginning to take shape, despite the principal's resistance.

All the other schools reported that their data processes are maturing. Educators in three schools stressed the importance of "triangulating data": drawing on results of multiple assessments to make decisions.

This process description is not uncommon:

> [E]verybody's job is to improve student performance. And the only way you can improve student performance is, obviously, you have to be able to measure performance, regularly. And so we look at this as just a way of measuring how we're doing as a school, annually. And we do, you know, measure more than that. But, at least annually, we're going to take some data that we can compare from year to year to see if we're making gains or we're regressing. Then it drives all of our curriculum changes and everything we do in the system. Our personnel, our programs, everything is driven around student performance.

In some cases, educators are taking a philosophical approach to developing data and school improvement processes, as does this superintendent:

> [I]t's always been a situation in school where time was the constant and learning was the variable. That's got to be reversed. And the standards, particularly, suggest a need to do

---

[8] We have visited more than one school that looks to an online assessment or a consortium because it is "overwhelmed." While understandable, it is our perception that this move typically overestimates the assistance the school will receive and often sacrifices local work for the perceived efficiency of a partnership. (See our year two report for a discussion of "unhealthy partnerships.")

that because now, suddenly, learning becomes the constant… But, recognizing the differences in kids, particularly high risk and high needs kids like we have, it has to be time now be the variable in order to get to that higher standard if, again, we don't believe we can leave some segments of the population behind, which we have done in this country—willingly in my judgment…[W]e need to be able to look at things like single track year-round programs. Or, begin to look at academies, or intervention years, or something like that. We need to do better jobs of assessment when newcomers come into the district…We need to decide what the readiness of the things are, including support that the families might need to be able to be better ready to learn.[9]

Others are taking a much more incremental, day-by-day, nuts-and-bolts approach. One principal says his school is focused on "quality instruction day after day, doing everything you can do, working hard, making sure the process is right, playing by the rules, doing all your things, developing kids." At any rate, in these schools, most educators would agree with this principal when he adds that today, "you don't just do things because you've done them in the past…You do them because they show success."

Curriculum
    In the area of curriculum, educators are using data to make the kinds of changes that our previous reports have identified as increasingly common in Nebraska schools. Specifically, schools use data to *focus the curriculum*:

- ♦ "When we get the…information back it has caused some discussion within our English department and our curriculum people in English on the different levels to maybe try and be more consistent, you know, as far as what we teach and how we teach it and to try to get more of a scope and sequence."

- ♦ "When we revised the English curriculum…the elementary started with theirs, then the others [did theirs]. And we took each piece that we felt was important and we aligned it…all the way across – vertically as well as horizontally."

"Focusing" the curriculum can be either positive or negative. It is beneficial, of course, to fill gaps and reduce redundancies. On the other hand, as we noted in last year's report, we see some schools narrow the curriculum to tested material, sacrificing valuable academic work in favor of testing convenience. What administrators call teachers' "pet projects," teachers often describe as important learning opportunities for their kids:

- ♦ "I have to cut out things that I would normally do…We have an advanced English class and… normally we read a Shakespeare play and last year it just didn't fit. I mean, you know, the amount of time it takes to just go in and take the test…"
- ♦ "It just has taken away from my class time. Things that I feel are really valuable…OK, let's take poetry for an example. I mean, there's no standard that says specifically that students will be able to appreciate…poetry. But, that's really I feel like an important thing when you're studying literature and English...Well, that's not something we can test on a 15-question multiple-choice question test."

---

[9] The superintendent does not provide a reference for his ideas about time and learning, but they echo those of a report called *Prisoners of Time* (Report of the National Education Commission on Time and Learning, April 1994).

- "Now what becomes important is what's reported. And…you throw everything into that at the risk of jeopardizing what most educators believe is important: And that is, the best education is well rounded."

Educators are also using data to *reorganize the curriculum*:

- "[W]e did the…low math facts—and money and clock and, you know, the basic kind of things. But now…you've got to take into account geometry and algebra and…logical reasoning and patterns and, I mean, just…a lot of other things."

- "[I]n that advanced math course, the standards that we've set and the expectations that we have for our students are much higher than they have been in the past."

In general, educators report that they expect more of students, especially, it seems, in high school math, where a great deal of curriculum work is being done.

Educators also use data to *integrate the curriculum*:

- "The math department wrote [lesson plan advice] for the other teachers. And they implement them in their classrooms to help reinforce what we teach in our pre-algebra and math 7 classes and math 8 classes."

- "[We used to have a] PE teacher who would say, 'I need you to do a report on an African American athlete' or something like that, and they [would] want 250 words…Now we're really giving them [teachers] some specific guidelines and help[ful] strategies."

Most of the schools in this study integrate writing, reading, and/or math across the curriculum.

Finally, educators are using data to *supplement the curriculum*, offering new programs or enhancing existing programs such as early childhood/preschool, after-school tutoring, and summer school. A superintendent explained that programs like these are crucial because "the thing we've all known forever is that some of them [students] just need more time."

All the schools in this study are constantly revising their curriculum. In two schools, the teaching staff collectively overhauled the curriculum by examining responsibilities for teaching and assessing. The result was a visual map of the curriculum, which in both cases was hanging on a wall in the school. Teachers in these schools were able to point out (literally and figuratively) not only their own responsibilities, but also those of their colleagues teaching in other grades or content areas.

Instruction

Almost all the schools in this study have developed an "information loop" between assessment and instruction. As they did in previous years, participants reported that their instruction is becoming much more sensitive both to group and individual needs. Three schools are targeting their "SES students" – that is, students who live in poverty. Others are targeting English Language Learners (ELL) and ethnic or racial groups such as Hispanics and African-Americans. At the same time, in several of these schools, "[e]verything is individualized; they [students] work at their own pace."

These changes are driven by data: by what the schools are learning about their students. As a superintendent told us, assessment data are "causing us to…rethink and relook at the way we organize instruction…[and] the way we use time." In this school, as in others, there has been a move away from simply teaching out of textbooks; instead, teachers are using textbooks in ways that match their curriculum and the needs of their kids – or are moving away from textbooks altogether. As a result, instruction in these schools is more flexible and responsive to students than it has been in the past.

In addition to studying their students, some of these schools are also examining instructional research. An assessment coordinator told us, "[W]e have a research team as a result of some of this data that we're looked at that's really studying, 'What are the best research-based programs? The best research-based assessments?'" Another assessment coordinator reported that in her district, "[W]e've identified research. Probably our biggest changes are coming k-3. We're really looking at going direct instruction…to try to improve the reading skills, because really, if kids aren't reading at grade level coming out of third grade, the chances of them catching up are pretty slim." We heard more this year than in previous years about "research-based instruction," perhaps because federal programs emphasize this approach.

For the most part, however, participants spoke about instructional change in terms of placing more or less emphasis on certain areas. For instance, several schools are focusing on vocabulary in response to writing scores and measurement in response to math scores.

Participants were most specific with respect to writing. They told us about practice assessments, focused reading, peer editing, writing to prompts, emphasis on the five-paragraph essay, explicit use of Six Traits in the classroom, revising, vocabulary exercises, sentence and paragraph structure exercises – all in response to writing scores and in preparation for the Statewide Writing Assessment (SWA). Surveys conducted by the CEP have revealed that teachers report little change in their teaching of writing as a result of the SWA (other than increased focus on Six Traits; see our year two report, Appendix F); this is not true in these schools, where the assessment seems to be driving writing instruction. It is difficult to say whether we are seeing evidence of a general shift in instructional practice or if this small sample is not representative of schools in the state. In any event, teachers whose responsibilities include the teaching of writing are focused on preparing students for the SWA, sometimes employing creative tactics:

> I wrote a persuasive essay for why basketball should not be happening at the school, you know. And I went through the cost…how much it cost to have the lights, and how much time it took kids away from homework. And I came up with three pretty good reasons and then I said, 'OK, now you have to say why do you think there should be basketball here' because, this town is a basketball fanatical town. I mean…it lives, eats, and breathes basketball.

Finally, as in previous years, we heard complaints from teachers about teaching to the test, "which kind of limits you on creativity," as one teacher put it. This sentiment was most pronounced in schools which administered district, ESU, or consortium CRTs.

Assessments

Most of these schools are constantly making changes to their assessment programs. Three schools are focused on spreading assessments out over the grades in an attempt to take pressure off of teachers and students in the reporting grades (4, 8, and 11). One superintendent explains how they have rearticulated responsibilities:

> [W]e've went back and basically established grade level and course level standards. And so that even though you may assess, the responsibility is going to be at the grade level. For example, fourth-grade standards—instead of possibly having the second-grade class teacher, and the third, and the fourth all being responsible for assessing a single standard, the responsibility may now lie just on the third-grade teacher. That doesn't mean that the second-grade teacher and the fourth-grade teacher still can't assess that standard, as long as it's a peer reviewed and approved …you know, it's been through all the quality criteria. They can still assess, but the responsibility lies only with that third-grade teacher. So what we've done is, we've done away with a lot of our redundancies. Initially, we had some holes and we filled all our holes. And that was a really good thing. Now we've gone back and tried to eliminate a lot of our redundancies.

We have been hearing for some time about schools filling holes and reducing redundancies in *curriculum*; note that this participant is talking about *assessments*.

In addition to changing *when* they assess, schools are also changing the assessment instruments themselves. We have seen that one school plans to move to an online assessment. Another has recently joined a consortium with a regional "STARS test." In a couple schools, by contrast, teachers are moving to smaller assessments: "I try to give lots of, say, smaller quizzes rather than…a few big tests…because it just makes them think about what they've been doing and maybe reinforce a little better." Finally, one school is focusing on sufficiency, as the superintendent explains:

> [O]ne of the things that we're doing now is…you know, it's not required for a couple years, but breadth of the assessment. That is required here currently now…So when a teacher submits an assessment to the peer review committee, it has to contain not necessarily an equal number, but a fairly equal representation of questions from the four different levels of performance, whether it's beginning, progressing, proficient, or advanced.

In two schools we visited, student portfolios travel with students from subject to subject and from grade to grade – "from kindergarten all the way through," as one teacher told us. Among other things, the portfolios record students' progress on standards.

Professional Development Programs

Most of these schools are using data to make decisions about how to focus professional development. For instance, teachers are being trained in areas such as data analysis, working with students in poverty, and writing assessment. (This last continues to be extremely popular: in several of these schools, *all* teachers – k-12 – have received Six Trait training.) An assessment coordinator explains how professional development has changed in her district:

> [P]rofessional development, again, is going to be geared more towards direct instruction,

research-based interventions. I think our days of sending people off to workshops at the most desirable locations to just attend a kind of a buffet [are over;] it's going to be much more focused on what we're trying to do as a district.

Indeed, as we indicated in last year's report, the term "professional development" in Nebraska is no longer synonymous with "attending inservices." Teachers describe on-site collaborative work with their colleagues as the most powerful form of professional development:

> [W]e're having more grade level meetings where we're able to discuss the things we're concerned about. You know, the grade level meetings in years past have been kind of… just hit and miss type thing. But we have scheduled Wednesday afternoons now when we meet with our grade levels. And I feel the fourth-grade level meetings have been very, very helpful because we discuss strategies on what can we do or what are you doing in your classroom that's helping in this area or that area?

Similarly, we heard from educators about the value of working with their colleagues on writing. One participant described a process whereby teachers share their students' writing as a conversation starter both for instructional improvement and curriculum development.

An administrator elsewhere described involving teachers in data analysis as a professional development opportunity: "[We] hire subs on a normal day, bring in all the language arts teachers in the district, look at our assessments, look at our curriculum, and make necessary changes." He explained that the goals of this activity are "providing alignment and looking at assessment and making that a part of their normal work." Put simply, he added, "[w]e have to just give them dialogue time. I mean, even though they have common planning times, we need to make sure that all those first-grade teachers sit, from time to time, and really discuss how students are performing in the district."

Professional development is changing also for administrators, particularly principals. Here are two superintendents talking about professional development for principals:

♦ "[W]e've had probably a total of four or five days training…for our administrators so they will really be administrators who are curriculum and instruction driven, and be able to be better supervisors in that area to actually help the teachers more with their instructional practice."

♦ "I wanted them to become more instructional leaders and not so much disciplinarians."

These changes are consistent with shifts in administrator's roles from conventional managers to "leaders of learning" (see our year two report).

*How They're **Not** Using Data*

None of these schools uses data to compare schools within their district or to compare themselves against schools elsewhere in the state – though several participants note that the media (particularly local newspapers) make these kinds of comparisons: "[W]e try to approach it…[as] a good opportunity for us to see what students are doing

and learn how to improve the learning opportunities for all these kids and try not to point fingers and compare between districts or between classrooms, between teachers. But you know that gets done." As this statement implies, the schools are comparing the performance of their students and their school over time:

♦ "[W]e have four elementary schools and we don't compare and say, 'Well, this elementary school is doing better than this.' We say, 'How have you helped the kids grow in your classroom?'"

♦ "[K]ind of like teacher evaluations, you never compare one against the other. You compare one student against where they were at…and where they are now. And that's what we use it for. It's not…it's not at a classroom level, it's not at a building level. It's at a student level. And you look at individual students. Where are they now and where do we have to get them."

In other words, these schools approach student- and building-level evaluation as criterion-referenced, not norm-referenced. The point, as a superintendent put it, is not to rank-order schools or kids, but rather to ensure that "we all improve."

Continuing Challenges
District-level and school-level variables such as district size, region, student demographics (percentage ELL, percentage SPED, poverty, etc.), portfolio ratings, AYP classification, and so on appear not to determine whether or not schools have well developed data processes or how the educators perceive those processes. (This is consistent with other studies featured in this report.) The only such variable that seems to have an influence on how teachers in particular experience and perceive assessment is whether they are administering assessments designed elsewhere. To be specific: in schools where assessments are controlled from elsewhere – whether it be the district, the ESU, or a consortium – teachers find it more difficult to use assessment data for instructional purposes and have more negative perceptions of assessment. We repeatedly heard from teachers in these schools that they had to wait too long for test results to be instructionally useful. By contrast, in schools that rely on classroom- and building-level assessments, the results are generally available to teachers immediately because they are primarily for diagnostic purposes and only secondarily for reporting and accountability. Even in schools with relatively sophisticated data processes, when the assessment was controlled from outside the building, teachers had a difficult time using – and sometimes even gaining access to – the data. The one exception is a school that is part of a large consortium, but uses the consortium test as only one source of data on student learning. Each study participant from this district insisted upon this; the superintendent, for instance, said, "[The consortium test] is an important piece, but understand it's only one piece." "Triangulation" of data is the key for this district and its schools.

As our studies show again and again, *time* continues to be the primary challenge for most educators. In this study, educators continued to insist that assessment takes time away from teaching: "you'll hear teachers say time and time again, 'By the time I get done documenting everything I need to document, and doing this plan, and checking this out, I don't have time to teach. I want to teach.'" This is a very common sentiment among Nebraska educators, as our previous studies have revealed. The administrators we

talked to this year seemed well aware of this concern; indeed, they shared it: "we need more time with our staff. We're locked into a 185-day calendar and that's to teach kids, but we need somehow to find more [time] so we can upgrade their skills. It's hard for teachers to teach and upgrade their skills all at the same time. It's not like a factory, where you shut the factory down for a week and you retool."

Another concern – one that has grown considerably over the course of our three-year study – is *funding*. In particular, several participants told us that "[s]mall schools feel the pressure because of limited resources." But at the same time, a participant from a large district insisted that her district did not "get the aid that they need." Several educators implied that their work with data was made more difficult by a lack of funding from the state and federal levels; a couple indicated resentment that they were being asked to do so much for so little. As an assessment coordinator explained, "I just feel like we spend a lot of resources…trying to figure out the reporting piece and trying to administer it and all of that. And I'm not sure what we get back is really useful as far as improving education for our kids."

We have already hinted at yet another significant challenge for educators: local media's use of assessment data to rank and compare schools. This concern has not abated over the three years of our project; indeed, educators continue to complain that their ability to shape their story is seriously constrained by local media outlets – usually newspapers – that either lack understanding about assessment and the purposes of STARS data, or are outright hostile toward the schools. Federal Adequate Yearly Progress labels have only exacerbated this problem by providing ready means for showing schools in a "negative light."

Educators shared other concerns with us as well – for instance, that *"the target keeps moving"* at the federal and state levels and that *schools with high challenge factors such as poverty, mobility, large populations of ELL and SPED, are disadvantaged in reporting*. Regarding the latter, more than one educator suggested that these factors should be better highlighted on the state report card or that ratings ought to be weighted with these in mind.

But more than any of these challenges, we observe that *educators' perceptions of assessment and their affective response to it* can stand in the way of (or, alternatively, ensure) progress in developing data systems. Above, we quote a principal who described his staff as at "that angry stage [where] they're not ready to accept and go forward and use that data in a positive way." In the interviews in his school, we found ample evidence to support this description:

- ♦ "The process started out to be a reasonable process of local development and now, with the validity and reliability and all the other things that came about…I think the teachers have just felt really inundated and there's a time lag between when we give it, when we get it back, and how we use that for a particular group of kids…The intent was to make it happen for kids, not…all of the energy going into the preparation. That's been the disappointment for me." (superintendent)
- ♦ "[A]ssessment is a bad word to [students]." (teacher #1)
- ♦ "Assessment as far as I am concerned isn't being used for constructive program development. It's being used to punish." (teacher #2)

♦ "[J]ust make them stop. Make them…make them let us have school. We really do know what we're doing. It's just…it shouldn't make you sad. It shouldn't…it shouldn't make you feel like…it shouldn't make you sad." (teacher #2)
♦ "I don't think I should ever be held responsible for somebody who just sits there…and does nothing. I don't want to be responsible for everybody." (teacher #3)
♦ "I'm not sure why Nebraska is doing their own. If it's good enough for everybody else maybe, why [re]create the wheel, you know?" (school board member)

In short, we found remarkable unanimity in this school about the nature and purposes of assessment and accountability. There is little support for, and little understanding of, programmatic evaluation in the school; we were told again and again that STARS is "ridiculous" because it compares different cohorts. This kind of attitude is not confined to this school, of course, but this example demonstrates how educators' understanding of and attitude toward assessment shapes the school's ability to develop viable data and SAA processes.

Conclusion

In most of these schools, there is near consensus that, as one superintendent put it, "You gotta have data." Consider the following descriptions of the role of data in these schools:

♦ "[Y]our data drives your student achievement. It's just like a road map. If you don't know where you've been…how in the world can you improve?" (principal)
♦ "[W]e're using that information to drive our goal with school improvement." (assessment coordinator)
♦ "[Assessment data] pulls me up off of my chair and out helping each of the students individually; it gets me out there which … just overall it helps me. It strengthens what I'm already wanting to do and this is just proving what I think is true: it's gonna help my students." (teacher)
♦ "It is school improvement. I mean, it's not about…assessment scores or anything like that. It's about school improvement, bottom line." (superintendent)

The common factor here is that educators view data not only or even primarily as an outcome, but as a means to achieving an outcome. That outcome is school improvement.

At the same time, as we have noted, resistance toward assessment and data continues. In last year's report, we noted a deepening divide between those who are "on board" with STARS and those who are not. Our sample is too small this year to draw further conclusions about this trend. However, like last year, educators tended to speak in stark terms about their support for or opposition to STARS (even though they were not asked about this).

For instance, several educators addressed the issue of a state test. A couple reported hearing increasing support for a state test from burnt-out teachers and administrators. Others offered their own assessment:

♦ "[C]urrently the state doesn't want us to compare ourselves to anybody else. And actually I think that's a coward's way out, personally. You know? But I do think that we need to, we need to know how we compare to other schools, to other districts within our

state, other states, other nations.   And the way that the state is [doing] this testing procedure right now, there's no way of doing that."

♦ "I am one of the small minority in the state that would like to see a state test. I don't want to see what we're doing now. I do not like what the state is doing because the media are comparing the scores. And what we do here…is totally different than what they do in any other school. And, how can we compare those scores? You're comparing what our kids have scored on what they've done in this school to what [kids elsewhere] have done in their school, and it's not the same. If there was a state test, then I could see comparisons. But, I don't like what's going on right now. I'm totally against it."

However, as the latter passage suggests, this view may be a minority opinion. In any event, we also heard from participants who would resist a state test and who endorse STARS and local assessment:

♦ "I like the idea that the district has been given the opportunity to come up with their own assessments as long as they are within the standards of the state…[A] district and school knows the needs [of kids] and what needs to be assessed."

♦ "[T]hese [district-level] tests were created by teachers and [what] I like about that is, it's really down to earth and again, it's related to what we teach, to our curriculum …because it is aligned. And I like that."

♦ "I would have said, six years ago—I did say six years ago when I moved here [from Texas]—"Why do we have to do all this? Why don't we just give a state test?"…Now, six years later, I've taken a change because I've seen what kind of information you can gain from tests that are written to your curricula, that you have written and then, that the teachers have said is important to them. And it's a lot of work from at least the instructor's perspective. You know, you get handed a test in April, you have to give it— that's not a whole lot of work….But if you can take this information that we're getting based on our test, on our kids, and use it wisely, that is the way to go. And I wouldn't want to be in any other state right now, even though it's a lot of work. I would not want to be in any other state right now."

At the end of the day, all the schools in our study, save perhaps one, are facing up to the reality of STARS. Over the time we have spent studying them, the conversations in most of these schools have changed from "Why do we have to do this?" to "*Can* we do this?" to "How can we do this *better*?" Sometimes, this shift is a product of resignation: "[Y]ou know what? If we're doing it, we may as well get as much information as we can from it." In other cases, as we have seen, educators believe that local assessment is right for their students, communities, and schools. In any event, in the schools that are working most effectively, the process is locally meaningful *first*, and fulfills external mandates as a secondary priority, as this superintendent explains:

It doesn't matter what system they throw at us. The process that we have in place is a good process and we'll be able to make it adapt and work for whatever the case is because it's child centered. And as long as it remains child-centered, we're going to be OK. It's not perfect. It's going to change. We know that. We expect that. We expect to improve every year. I mean…you know, that's the expectation that we have for our kids.

That's the expectation that we have for ourselves. Every teacher here is expected to improve. We establish our own goals. We actually cooperatively establish goals and we work towards reaching them.

Recommendations

This interview study reveals that progress is being made in most of these schools toward integrating data into school improvement. But even in the most advanced schools, a good deal of work remains to be done. Much of the responsibility for this work resides at the local level, but the NDE can support schools in developing their data processes in the following ways:

Continue to focus professional development offerings on working with data.
Educators in Nebraska are increasingly using data for various purposes and there is therefore a need for continued professional development in this area. This includes teachers, who are often responsible for interpreting and using data. The NDE, perhaps in partnership with ESUs, should focus professional development opportunities for both administrators and teachers on this crucial aspect of developing effective SAA systems. Special care should be taken to help educators see data as a means to achieving critical school outcomes, but not as itself the chief outcome. In one of the schools we studied, the language of learning has been almost entirely usurped by the language of data, such that "getting the numbers up" – rather than actual student learning – seems to be the priority. Educators need to know how to work with data and how to fit that work into their larger goals.

Use local talent to model the building of data processes and the integration of data into school improvement. Nebraska schools are at varying points in developing their data processes and their SAA processes. Some have already developed exemplary processes, while others are just beginning this work. We advise using the former to assist the latter, perhaps as part of school improvement workshops and other NDE activities.

Require all districts to have at least one assessment coordinator. In many Nebraska schools, assessment coordinators are not only leaders in working with data, but also the catalysts for developing SAA processes. As technical quality expectations increase and as data continues to roll into and out of schools, it seems imperative to have a key person assigned to lead the SAA charge. Schools in our study that have formalized a position in assessment have improved their data processes. By contrast, the one school in our study that lacks a dedicated assessment person is also the school with the least developed data and SAA processes.

Continue to educate various constituencies about the nature and purposes of STARS. Focus special attention on local media. As we have indicated in previous reports, some progress has been made in educating some members of the media across the state. However, educators continue to report that local newspapers insist on placing schools in a competitive light. We understand that the NDE, and Commissioner Christensen in particular, has "stayed on message" throughout the evolution of STARS. But it is clear that efforts to dispel inappropriate interpretations of assessment results (state and federal) are still needed in many communities. We recommend a series of

"media workshops" or an "education information fair" specifically targeted to local media in all parts of the state. These events should feature state and local educators who can explain in forceful and accessible language what kinds of interpretations of assessment data are appropriate and inappropriate. In our view, local media also would benefit from understanding what is happening in other states and why Nebraska has chosen its unique approach to SAA.

Examine the effects on schools of using assessments designed elsewhere, whether in consortia, ESUs, or districts. We expressed concern last year about "unhealthy partnerships." Commissioner Christensen has begun to address this concern by encouraging schools to involve themselves in consortia only if they ensure the involvement of all staff. We believe a study of these kinds of arrangements is warranted, particularly given the evidence we find in this study that schools that most need to be working on their data processes may be most likely to look for a "quick fix" by consorting, purchasing an online assessment, etc.

Partner with schools to experiment with alternative, creative uses of time. Over the three years of this study, one finding is consistent: time is the most onerous challenge for educators. We agree with the superintendent above who indicated that the time has come in schooling to rethink how time is used. This includes the school day, the school year, and how time is used within those frameworks (for instruction, curriculum development, professional development, etc.). If the NDE is serious about pursuing this recommendation, it should work with schools (and perhaps in particular high schools) to conduct formal experiments, including a systematic research component, with alternative uses of time in school. (Note: we are *not* advocating simply *more* time in school.) Higher education and ESUs might be called upon as partners in such a venture. School reform advocates have been talking for a long time about changing the ways we handle time in schools; Nebraska – with its unique approach to SAA – would be the perfect place to learn more about how these ideas play out in practice. This initiative could dovetail with the NDE's "rethinking the high school experience" project.

Study A Conducted by
Jody Isernhagen, Ph.D., University of Nebraska-Lincoln
Leon Dappen, Ph.D., University of Nebraska-Omaha
and Nebraska Evaluation and Research Center

Study B Conducted by
Gregory Schraw, Ph.D., University of Nevada-Las Vegas
Keith Zvoch, Ph.D., University of Nevada-Las Vegas
Center for Evaluation and Assessment

These studies represent the first quantitative examinations of data generated by STARS, including criterion-referenced results, norm-referenced results, and portfolio ratings. Because we will soon have longitudinal data on mathematics as well as reading, we expect more such studies in the future (see Validation Framework, Appendix A). It is vital that the system be validated by multiple measures, including analyses of student achievement and district assessment quality.

In this section, we present the respective Executive Summaries. The full reports may be found in Appendices E and F. Following the Executive Summaries is a discussion that places the studies in the context of the Comprehensive Evaluation.

Study A: Isernhagen and Dappen

An Examination of 2001 to 2003 Nebraska Criterion-Referenced Tests, Norm-Referenced Tests and District Portfolio Ratings for Reading at Grades 4, 8, and 11

Executive Summary

Nebraska has made a decision to implement STARS (School-based Teacher-led Assessment and Reporting System), a unique system of accountability that focuses upon building assessment literacy among educators and enhancing student performance through the use of a quality assessment system. This system includes both norm referenced (NRT) and criterion referenced (CRT) assessments developed by each district to align with their curriculum.

CRT measures and the determination of proficiency are unique to each district. Each district also uses an NRT of their choosing from those that are available. Assessment Portfolio summary ratings represent scores based on expert review of the process of assessment on six criteria of quality assessment. Based on the unconventional characteristics of these data, traditional statistical approaches for significance of comparisons were not appropriate. Findings are therefore based on descriptive data.

This study reports the first achievement data from this new effort: the 2001 and 2003 reading scores reported for all k-12 school districts for classes 3, 4, and 5. This includes just over 94% of the public school students in Nebraska.

Findings:

- The average percent of students meeting the defined mastery level for their locally developed CRT measure increased at each grade level from 2001 to 2003. At fourth grade, over 5% more students were rated as proficient in 2003 than in 2001. This lends strong support to continuing CRTs in the instructional activities for improvement of achievement.
- NRT scores increased at the fourth and eleventh grades, and decreased slightly at the eighth grade. These scores did not appear to have been negatively affected by the implementation of the CRT program; indeed, there is some indication of slight increases.
- Portfolio ratings from 2001 to 2003 increased strongly at each grade level. This would seem to reflect growing knowledge and skill in assessment practices by district staff.
- While this is initial information, the data are compelling to support the continuation of the Nebraska STARS approach.

These findings provide a base of support for the district reading assessment portfolio process in establishing the credibility of the Nebraska School-based, Teacher-led Assessment and Reporting System (STARS). Future efforts need to follow each academic area (math, science, social studies), as data is available. Research efforts are needed to examine the STARS program as it relates to NCLB. There is a need to engage the measurement community in discussion of how various aspects of STARS (locally developed CRT scores, NRT scores from different companies, district assessment portfolio process ratings) may be examined for significance.

Study B: Schraw and Zvoch
Final Report on the Analysis of STARS Data

Executive Summary

We investigated four research questions related to the relationship between state developed criterion-referenced assessments (CRT: STARS) and norm-referenced assessments (NRT), and the extent to which a variety of student and teacher variables relate to district proficiency and compliance with state and federal standards. STARS and NRT assessments were moderately to highly correlated within and across student cohorts (i.e., within and across academic years and grades). Cohort performance on the NRT assessments was stable over the 2001-2003 timeframe; however, the 2003 cohort showed significant proficiency gains over the 2001 cohort on the STARS assessments among 4th, 8th, and 11th grade students. Student and teacher variables explained a significant, but modest proportion of the district-level variation in proficiency and compliance rates. Three district characteristics, student attendance rates, teacher experience, and teacher education, were most predictive of district performance. Other

district characteristics, including mobility and graduation rates, district size, and student demographics did not improve prediction. We conclude that STARS provides a more sensitive measure of learning progress. We also suggest that the NDE collect data that is more closely tied to classroom practice and procedure in order to better understand why some districts are more able to reach targeted proficiency levels and comply with state and federal accountability standards.

Recommendations

Several recommendations seem reasonable in light of these analyses. We summarize these below.

1. Continue using STARS as a planning and assessment strategy. STARS appears to be a more sensitive measure of yearly progress than NRTs.

2. Promote attendance. In the present analyses, attendance rates were the best predictor of proficiency and compliance by a wide margin, especially among at-risk districts.

3. Recruit and utilize experienced teachers with graduate training. One way to do so is to use experienced teachers as mentors within their school or as in-service facilitators. A second way is to use experienced teachers to serve at-risk students.

4. Consider collecting data on additional variables not included in the present analyses. Variables that assess the quantity and quality of instruction may be particularly important given the important role of attendance rates. Students who attend school more often are more likely to become proficient and meet state and federal standards. Currently, it is unclear what role the quantity (i.e., amount of time) and quality of instruction play.

**

Discussion

The findings from both studies are valuable and point the way toward future research. It is important to note that while the studies used somewhat different databases and methodologies, they come to similar conclusions, suggesting the reliability of the results. Still, as we examine the findings, it is wise to bear in mind several caveats. The first, most important caveat is that data generated from STARS are based on different measures and so are not amenable to conventional metrics. Therefore, we should view interpretations of these data as interesting and meaningful, but not definitive.

CRTs and NRTs vary from district to district, which creates uncertainty when comparing results. The results from CRTs represent the percentage of students who perform at the proficient or advanced levels on Reading CRTs, as reported by districts. The results from the NRTs represent the percentage of students who perform in the top two quartiles on their respective tests, again as reported by districts.

Cut scores are also determined locally. Although districts are all reporting on state standards (with the exception of those reporting on local standards determined by the state to be equal to or more rigorous than state standards), mastery levels are determined by districts. This of course makes it difficult to compare results. On the other hand, it is important to note that districts are required to use an appropriate,

systematic, data-supported procedure for setting mastery levels in order to receive a "met" on Quality Criterion Six. In 2003, 91% of districts met QC 6 for fourth grade, 91% met for eighth grade, and 88% met for eleventh grade. Thus, almost all Nebraska districts have been certified as setting appropriate cut scores.

Additional caveats are warranted as well:

♦ Two of the districts included in both sets of data are very large (relative to the other districts) and a fairly large number of districts are not represented in either study because they are so small.
♦ The two studies have different n's. Dappen and Isernhagen included all class 3, 4, and 5 districts, and excluded very small districts. Schraw and Zvoch included all districts with more than 50 students.

Findings
Study A:

From our point of view, the most important findings from Isernhagen and Dappen's study are these:

♦ Gains in the mean percentage of students proficient on reading CRTs at all grade levels, 2001-2003 (including a 5% change in fourth grade);
♦ Gains in mean percentage of students proficient on reading NRTs at fourth and eleventh grades (slight decrease in eighth);
♦ Gains in average district portfolio rating from 2001-2003;

Study B:

From our point of view, the most important findings from Schraw and Zvoch's study are these:

♦ Gains in the mean percentage of students proficient on reading CRTs at all grade levels, 2001-2003 (statistically significant at all grades);
♦ Stable performance on NRTs, 2001-2003;
♦ 2001 and 2003 student cohorts perform similarly, especially at fourth and eight grades;
♦ Few district characteristics predict district performance:
  • District size, percent SPED, mobility rates are *not* predictive of proficiency or compliance with state and federal goals
  • Student attendance, teacher experience, and teacher education have some (but limited) predictive value

Implications

The increase in portfolio ratings confirms findings from our interview studies that assessment literacy is growing among Nebraska educators. As Isernhagen and Dappen's analysis reveals, many districts made significant improvements in their ratings, jumping two or even three score points.

Second, these studies reveal that almost all indicators of performance (student and district) are on the rise. (The single exception is a slight decrease in student performance

on reading NRTs in eighth grade.) Gains on CRTs are more pronounced than gains on NRTs. This is to be expected, as both studies suggest, because CRTs are designed to be sensitive to local curriculum and instruction. That NRT performance has remained relatively stable suggests that CRT preparation is not interfering with Nebraska students' generally strong performance on national tests.

Finally, Schraw and Zvoch's finding that district-level variables tend not to predict student performance is worthy of note. Analyses of the elementary assessment survey similarly revealed that district-level variables did *not* account for the variation in response. It appears that future studies must focus more specifically on schools, classrooms, and individuals.

Recommendations

These findings support **the continued use of multiple measures of student learning**. (This recommendation is also supported by Impara et al.; see Sufficiency Study.) CRTs and NRTs together provide a more comprehensive portrait than either one could alone. Moreover, Schraw and Zvoch find CRTs to be more sensitive measures of progress than NRTs. This supports their continued use in the context of state and federal improvement goals.

In addition, the NDE should **assist districts in developing CRTs of the highest possible quality**. The sensitivity displayed by these measures and their alignment with local curriculum make them crucial measures of student learning under STARS. This study confirms the significance of CRTs as part of an overall portrait of student performance.

Third, the NDE should **target assessment literacy opportunities to eighth and eleventh grades**. Both studies find that all the indicators are moving in the right direction at fourth grade. At the other two reporting grades, the findings are promising but not as pronounced, especially in the Isernhagen and Dappen study. We suspect part of the explanation for this is owing to disproportionate assessment literacy growth among fourth-grade teachers, who are responsible for assessment and reporting for all reported subjects. Meanwhile, at the higher grades, teachers identify more closely with their content specialization; perhaps the rate of assessment literacy growth is slowed by this identification. Schraw and Zvoch also note the potential impact of inconsistent curricular offerings and assessments at the 11th grade, in particular. At any rate, given the lower performance gains in the higher grades (and especially the slight decrease on NRTs at eighth grade), it makes sense to target assessment literacy opportunities at these higher-level educators.

Finally, **continue to work with the measurement community to develop appropriate metrics for the kind of data generated by STARS**. Readers will note that the two studies here use different methodologies and both suggest pushing the research agenda forward. STARS creates some thorny problems for conventional metrics. Researchers must continue to seek creative ways to examine NRT, CRT, and portfolio data. In addition, it would be useful to track the same group of students over time on the same measures. The advent of the student data-base should help facilitate new, finer-grained analyses.

Elementary Assessment Survey

Study Conducted by
Dottie Heusman and the CEP

Because Dottie Heusman is an employee of the Nebraska Department of Education, we have conducted our own analyses of the survey she conducted with the assistance of the Comprehensive Evaluation Project (CEP).[10] To avoid potential conflicts of interest, CEP staff worked with Ms. Heusman throughout the project to ensure districts' confidentiality; although Ms. Heusman designed the instrument and the sample, at no time was she aware of which districts were included in the study.

The primary purpose of this study was to determine whether student performance in a district influences how elementary teachers perceive five themes related to assessment: Administrative Support, Integration, Collaboration, Accountability, and Portfolio. The secondary purpose was to determine how the elementary teachers perceive these five themes.

We used fourth-grade student performance data for reading from the 2000-01 and 2002-03 schools years to classify Class II and Class III districts into three categories: those with increases in student performance (+10.03% to +79.80%); those with stable student performance (-9.84% to +9.97%); and those with decreasing student performance (-50% to -10.03%). A random sample of 10 districts was chosen from each of the three categories. The sample was reviewed for geographic representation and was found satisfactory.

All second-, third-, fourth-, and fifth-grade teachers in these thirty districts received a copy of the survey from the CEP (see Appendix D for a copy of the instrument). In addition, special education, Title I and building administrators received the survey. Respondents were asked to return completed surveys to the CEP office at UNL.

In all, 128 surveys were returned, for a return rate of 40%. This return rate is rather low, but consistent with those of similar surveys we have conducted. Possible reasons for low return rates include constraints on teacher time, teachers moving into new positions, misidentification in the NDE database, and content considerations (i.e., teachers may respond negatively to the subject matter of the survey). Ninety percent (90%) of the respondents were teachers; 10% were administrators. This group of respondents yields a strong distribution of years of experience (with a mean of 20 years) as well as grade level (grades 2-6). Over 40% of these educators are multi-grade teachers and a small number are special education, Title I, and resource teachers. The respondents are representative of the state geographically.

---

[10] Heusman's full report, including a more detailed description of the protocol, has been submitted to the NDE. Readers interested in more details on our methodology should contact Chris Gallagher.

Survey results were coded into the Statistical Package for Social Sciences (SPSS) and statistics were run by the Nebraska Evaluation and Research (NEAR) Center. A short, technical synopsis provided by Michael Toland of the NEAR Center may be found at the end of this section.

Findings

The major finding of this study is that student performance did *not* affect educators' responses to questions. That is, the student performance category of the district in which the respondent taught did not explain the variability in responses. In fact, no district-level characteristic – such as percent ELL, percent SPED, mobility, or poverty – was found to explain the variability of responses. (See Michael Toland's synopsis below.)

When we consider the themes, the educators surveyed have generally positive perceptions of administrative support, accountability, and portfolio. Their perceptions of collaboration are somewhat more negative, though this is primarily owing to two questions about time. The theme of integration cannot be considered as a whole because the questions within that section of the survey were unreliably correlated (even after reverse coding two questions that were framed negatively).

All items included a four-point Likert scale: Strongly Agree (1), Agree (2), Disagree (3), Strongly Disagree (4).

Administrative Support

This section included five items. The mean for the section was 9.80, suggesting positive perceptions of district-level administrative support. (Five items yield scores between 5 and 20; the lower the score, the more positive the perception.) In particular, 87% of respondents agreed or strongly agreed that their district administration provides support for the language arts standards and assessment process by providing them with professional development opportunities. That same percentage indicated agreement or strong agreement that their district looks at multiple years of data for school improvement.

Integration

The items in this section are not reliably correlated, so we will not consider the grand mean. With the small number of respondents, we can consider individual items only with caution. Still, it is worth noting a couple items. First, 95% agreed or strongly agreed that assessment is an integral part of their classrooms. Second, 52.8% agreed or strongly agreed that a statewide test would be more appropriate for their district than the current assessment system.

Collaboration

This section included four items. The mean for the section was 10.52, suggesting mildly negative perceptions of collaboration. (Four items yield scores between 4 and 16; the lower the score, the more positive the perception.) Again, however, this high mean is explained largely by two questions about whether respondents have enough time for

collaboration. These questions may be more reliable measures of respondents' perceptions of time than of collaboration.

Accountability

This section included six items. The mean for the section was 11.47, indicating very positive views of accountability. (Six items yields scores between 6 and 24); the lower the score, the more positive the perception.) Respondents are split on whether students are accountable, but 94% agree or strongly agree that teachers are accountable for students' learning. Sixty-one percent (61%) agree or strongly agree that district assessments measure what students know and can do and the same percentage indicate that those assessments provide valuable information about students. A slightly lower percentage (56%) indicate that norm-referenced tests measure what students know and can do. Eighty percent (80%) indicate that they have changed their classroom practice as a result of the assessment process, and among those, 61% perceive that the changes benefited students.

Portfolio

This section included four items. The mean for this section was 9.19, indicating mildly positive perceptions of portfolios. (Four items yields scores between 4 and 16; the lower the score, the more positive the perception.) Sixty-eight percent (68%) agree or strongly agree that they understand the six Quality Criteria (QC). Seventy percent (70%) indicate that they believe the QC are good indicators of the quality of their district's assessment process. Fifty-five percent (55%) believe the assessment portfolio rating system is fair, and 69% believe the district assessment portfolio is an accurate representation of their districts' assessment portfolio process.

Implications

The finding that student performance did not affect educators' responses to questions is consistent with our interview study, in which we found that student performance did not seem to affect the extent to which districts had developed their data processes. It appears that district-level characteristics have little influence on individual educators' perceptions of assessment.

When viewed in the aggregate, these survey results indicate generally positive perceptions of assessment among elementary educators. Educators indicate strong administrative support for assessment. They also report looking at multiple years of data for school improvement, which our interview study reveals to be a growing focus in the schools we studied.

Both the interview study and this survey reveal a strong focus on tying assessment to instruction. This trend has been growing stronger over our three-year evaluation. The survey also reveals generally positive perceptions of the ability of local assessments to measure student learning (slightly more positive than of norm-referenced assessments; this is consistent with Professor Bandalos' study, included in our year two report). This suggests another trend that we have seen in our interview study: the alignment of curriculum and assessment.

At the same time, a small majority of educators surveyed would prefer a state test to the current system. Although the relatively small sample size suggests we **should use**

caution when interpreting this result, it is nonetheless cause for prudent concern. The fact that Nebraska educators tend to have more confidence in local assessments than in norm-referenced assessments suggests that this preference is driven less by the perceived relative value of the kinds of assessment than by the time and energy required by local assessments. These educators, like their colleagues in our interview study, indicate a lack of time for collaboration in particular. Comments on surveys consistently reveal this concern:

♦ "This process takes so much time & man power."
♦ "I have no trouble being accountable for what I teach & what the kids have been taught, but this whole process is cumbersome & unmanageable."
♦ "The fact that the process is school-based and teacher-led makes it a more accurate evaluation tool. However, the BIG problem is: we <u>are not</u> given time to accomplish what needs to be done."

Although a number of comments indicate educators' support for STARS, many express educators' sense of being "overwhelmed" by the demands of assessment and not having enough time to teach. This is especially true among fourth-grade teachers, which suggests that many districts continue to place the primary burden for assessment on the shoulders of reporting-grade teachers.

The comments also indicate another reason for this preference, as well: a certain understanding of equity. These comments are typical of this view: "I feel that in order to <u>level the playing field,</u> all students should take the same (statewide) assessment"; "The only way to compare results is if every student in every school is given the same test and it is scored the same." For some educators, then, what is at stake in this preference appears to be a philosophical difference with STARS, which is not intended to generate rankings of the sort imagined here.

Educators' perceptions of district assessment portfolios are positive. Further research is required to determine why the percentage of educators perceiving the portfolio rating system to be fair is as low as it is (55%). Responses to other questions in this section suggest that educators have considerable confidence in the portfolio system.

In sum, this survey confirms that assessment is an important part of the lives of Nebraska educators and that they are using assessment for both instructional and accountability purposes. Further, while many educators would prefer a state test, these results indicate that educators have generally favorable perceptions of assessment under STARS.

Recommendations

Unfortunately, these results do not provide much direction from a policy perspective. That is, they do not help us understand the key differences among schools with increasing, stable, and decreasing student performance. Therefore, our first recommendation is to conduct further research on what distinguishes schools with varying performance. We recommend including assessment quality ratings alongside student performance ratings, since districts with high student performance and low assessment quality are not doing as well as it might appear (and vice versa). Future

research should also focus on school and individual characteristics, rather than district characteristics alone.

Second, continue to educate various constituencies about the nature and purposes of STARS, including educators. **This recommendation echoes a recommendation we propose in our Interview Study: Using Data. There, we focus on the need to educate local media; here, we emphasize the need also to target educators throughout the state. That so many educators express a preference for a state test is troubling. Our interview study would suggest that this preference is waning, but this survey finds a large number of educators harbor this preference. Research should be conducted on *why* educators hold this view and every attempt should be made to address their concerns.**

Third, as we suggest elsewhere in this report, the NDE should provide leadership on alternative uses of time, especially for professional development. **Time continues to be educators' greatest challenge in complying with the demands of STARS; it would behoove the NDE and districts throughout the state if efficient, effective time management models were shared widely.**

Postscript: Synopsis of Results
Michael Toland, Nebraska Evaluation and Research Center

Because of the hierarchical data structure of these data and the primary variables of interest were at a different level than the outcome variables, Hierarchical Linear Modeling (HLM; Raudenbush & Bryk, 2002) techniques were performed on these data. The outcome variables were summative scores on administrative support, integration, collaboration, accountability, and portfolio assessment. Level-2 predictors included district population size, mobility rate, and reading performance groups (negative gains < 10, neutral/moderate gains, positive gains > 10. Also, an additional level-1 predictor, but not of primary interest was years experience as a teacher. However, the general question motivating these analyses was concerned with determining if respondent ratings of their respective districts elementary administrative support, integration, collaboration, accountability, and portfolio ratings varied by characteristics of the district (i.e., district population size, mobility rate, and reading performance groups).

For each outcome variable (five total) a one-way ANOVA with random effects model was performed (a.k.a., unconditional model or baseline model). For our first unconditional model with administrative support as the outcome variable we found that the parameter estimate for the average district mean for administrative support was $\gamma_{00} = 9.94$, p < .001. This indicates that the grand mean for administrative support is statistically significantly different than zero, but this is usually not of importance. We are more interested in finding out if the variability about the grand mean for administrative support varies by district level variables (i.e., does population size of district impact or correlate or explain variability in administrative support scores). The variance estimate for variability about the intercept or grand mean was $\tau_{00} = 2.38$, p > .05. In other words, the level-2 residual or left over variability without any level-2 predictors represents the between group variance in administrative support. The level-1 residual was 8.97, p < .01. Since there are no level-1 predictors, it simply reflects the within group variance in administrative support. Normally you will want τ_{00} to be large and statistically significant to indicate that you have variability to model or variability left to model.

Because this test of $\tau 00$ is not statistically significant different than zero, it is sensible to assume that all districts in our sample have the same mean for administrative support. That is, there is not variation in mean scores about the grand mean for administrative support and no variability about the grand mean is interpreted as meaning that there is nothing left to be explained by district level variables (level-2 variables) like district population size, mobility rate, and reading performance groups.

Consequently, this general synopsis is what is found for all four remaining outcome variables. Additionally, analyses were run to see if years of experience teaching could predict or correlate with any of the outcome variables (administrative support, integration, etc.), and similarly, all results indicated that this variable was not statistically significantly correlated with any of the outcome variables.

Study Conducted by
The Buros Institute for Assessment Consultation and Outreach
A Division of the
Oscar and Luella Buros Center for Testing
University of Nebraska - Lincoln

This sufficiency study aimed to determine the potential for selected norm-referenced and criterion-referenced assessments currently in use in Nebraska to classify student performance into four categories: beginning, progressing, proficient, and advanced. Given the requirement for districts to demonstrate sufficient depth and breadth in their assessments by 2005-06, this study should be of interest to Nebraska educators as well as policymakers.

Impara et al. have submitted their full report to the Nebraska Department of Education. In this section, we present the Executive Summary, followed by a discussion that places this sufficiency study in the context of the Comprehensive Evaluation.

Examining the Potential for Selected NRTs and Locally Developed CRTs to Classify Students into Performance Categories in Reading and Mathematics
EXECUTIVE SUMMARY

James C. Impara, Ph.D
Chad W. Buckendahl, Ph.D.
Abdullah Ferdous, M.A.
Renee Jacobson, Ed.S.

This project had two phases and was undertaken during the period from March 2003 to April 2004. The data for this study were collected in two phases. Phase 1 was the determination of performance levels in reading and mathematics. This took place in July 2003 and a report was delivered in September 2003. This report covers the second phase of the project in which data were collected to determine the extent that 3 NRTs from different publishers and 21 CRTs (11 in reading and 10 in mathematics) could be used to classify students as performing at Advance, Proficient, Progressing, or Beginning levels on selected standards. In addition to this main focus (the determination of the sufficiency of the assessments to make performance level classifications) there was a second study that looked at the consistency of teachers across multiple sites to make the same performance level judgments about a sample of assessments.

A total of 66 teachers met at one of three locations around the state to judge the assessments in reading. An additional 52 teachers met at these same locations to judge the mathematics assessments. After an introduction to the study and undertaking a practice activity, teachers spent approximately a day evaluating the tests at their grade level.

The results of this activity suggest that the NRTs have some utility for classifying students as either Proficient or Below Proficient on the one reading standard that was examined. Some additional, more precise classification is also possible, but such classifications should be made with caution because, in some cases multiple subtests are involved and the proportion of items that focus on the specific standard may be a relatively small proportion of the items on the subtest. This is even truer in mathematics where the utility of classifying student on two standards was examined. On the standard related to Computation and Estimation, the NRTs were comparable to the utility for reading. Similarly, the items were often across more than one subtest, so the proportion of the total items that focus on the standard is not known. However, for the standard related to Data Analysis, Probability and Statistical Concepts, there tended to be fewer items, all in one subtest and fewer opportunities to make performance level decisions.

The CRTs that were examined included some assessments that were locally developed and used only in the local district. Other assessments had been developed in consortia and were used in multiple districts. In general the utility of these assessments to classify students into multiple performance levels is mixed. At grade 4 in both reading and mathematics the assessments of about 67% of the standards provided classification into either Proficient or Below Proficient categories. This percentage declines rapidly to about 50% at grade 8 and about 25% at high school. If more precise classifications are desired (e.g., Advanced, Proficient, Progressing, Beginning) these percentages drop dramatically. The worst case is in attempting to classify students as Advanced. Across all grade levels, assessments and standards in reading (3 x 11 x 6) there were only 13 instances when a rating of Advanced could be made. The situation is only slightly better in mathematics where there are 22 such instances.

The study related to consistency of judgments across settings suggested that some caution should be undertaken in interpreting the results related to the CRTs. Although it was often the case that there was agreement between two of the three sets of teachers, and occasional agreement among all three sets, there were also some assessments on which there was virtually no agreement. In short, there was no consistent pattern of agreement. That is, for some assessments the teachers in the Eastern and Central regions agreed with each other on their item classification judgments, for other assessments, the Western and Central teachers agreed, and for yet other assessments the Eastern and Western teachers agreed. These were not systematic for the same standards.

A number of recommendations were made. Some of these recommendations focus on the process (e.g., making it clearer how the performance level definitions should be used). Other recommendations focus on the results, particularly on the utility of the assessments to be used for making performance level classifications. All of the recommendations are listed below.

Recommendations

Process

- The notion of overlapping participants may be sound, but the time span between the development of the definitions and the operational study needs to be much closer in time so that the participants who develop the definitions do not have time to forget the discussions and rationale for the decisions made at the initial meeting.

- Before undertaking further studies using performance level definitions, the definitions should be reviewed and ambiguities and inconsistencies eliminated.

- A rule of precedence must be developed when more than one standard might be applicable (either directly or indirectly) to an assessment task.

- Clarify the weight of the rubric and the weight of the assessment task when there appears to be a conflict. This may require expanding the performance level definitions to include not only content, but also skills that are needed to demonstrate content knowledge (i.e., how the content is assessed).

- Continually emphasize to the teachers that the decision about the performance level of an assessment task should not be based on their students but instead is based on the performance level definitions provided and on the rules of precedence and weights developed based on the above recommendations.

- In future studies, a better explanation of the outcomes of the process and reassurance that the results will be confidential (to other districts and the NDE) along with assurances that the results will be shared with the district, may help with getting cooperation.

- Districts should be advised that more direction to teachers is needed. For example, rather than just leave open the selection of a reading passage, teachers should be provided a list of passages that are of similar difficulty.

- The NDE needs to work with districts to help them understand the difference between when rubrics are appropriate and when they are not. Moreover, the districts need to know that using the rubric to define the performance levels is not an appropriate procedure, because such a process bypasses the standard setting process and makes the performance level classification arbitrary.

- The NDE should encourage all districts to develop performance level definitions and to apply these definitions independent of the rubrics used to score assessment tasks. If this is not done, then the assessment tasks should be evaluated by the district to make sure that the assessment tasks and the score levels accurately reflect the performance level definitions.

- Practice tests should provide examples of all performance levels for all standards. They should be exemplary tests that represent quality assessments.

Assessment Utility

- Continue using the NRTs for classifying students as either Proficient or Below Proficient, but stop using NRTs for classifying students into four performance levels, especially using the arbitrary cut points of the 75%ile and 25%ile.

- Districts that are using these NRTs to make classification decisions should be doing so with extreme caution. It would be useful to supplement these tests with well-constructed CRTs to be more comfortable in the classification of students into performance levels. The items related to the strand 5 standard represent only a minority of the items on the subtest on which theses items are found, thus using the 50%ile (or any other point) as the dividing point for Proficient or Below Proficient will probably result in many misclassifications.

- Districts should carefully review their assessments to insure that the assessment tasks, and where appropriate the assessment rubrics, provide opportunities for students at all performance levels to demonstrate their knowledge and skill

relative to all standards. Districts should start with their assessments at the high school level.

- The NDE should place little confidence in the districts' classifications of students based on these assessments. The reported data should be collapsed into only two categories Proficient and Below Proficient to obtain a more accurate representation of student performance levels.

**

Findings

      This study demonstrates that both NRTs and CRTs currently in use in Nebraska schools are variable in their utility to classify students into four performance levels. Some NRTs and CRTs are capable of classifying students as Proficient or Below Proficient, and only a few allow for more precise classifications. In general, the assessments include few or no items that allow classifications into the Beginning and Advanced categories. In only a small minority of cases do assessments allow for classification of students into four performance categories. For certain standards, CRTs and aligned NRTs do no provide sufficiency for any classifications at all.

Implications

      These findings are sobering. Some districts rely heavily on NRTs that do not provide sufficiency – sometimes by arbitrarily setting cut points (e.g., below 25th percentile = Beginning; above 75th percentile = Advanced). As for the CRTs, all of the local assessments chosen for inclusion in this study were rated at least Very Good, but they do not provide sufficient opportunities for students to demonstrate their performance level (though we should remember that the district assessment portfolio review process rates portfolios as a whole, not the individual assessments embedded in them).

      When we consider the short period of time Nebraska educators have had to develop their assessment literacy and the quality of their assessments, perhaps it is not surprising that the CRTs do not, for the most part, hold up to technical scrutiny. Similarly, because the NRTs were not designed to show proficiency on Nebraska state standards, their mixed but clearly limited utility is not unexpected. But the fact remains that much work needs to be done in this state if districts are to meet the sufficiency requirements scheduled to take effect in 2005-06.

Recommendations

      We have anecdotal evidence from our interview study that many districts are already working hard on sufficiency with respect to their local assessments. Moreover, we believe districts *can* meet minimum sufficiency requirements by the reporting deadline of 2005-06. So while we agree with Impara et al. that at present, we should have little confidence in performance classifications, and while we see the attraction of simplifying reporting into two categories (proficient and not proficient), our first recommendation is to **retain the four-level reporting requirement**.

For us, there are three issues here. First, teachers need the kind of diagnostic information they get from a meaningful performance classification. Reducing the number of classification possibilities reduces the amount and quality of information teachers get about their students (assuming the classifications are sound; more on that in a moment), and thus limits their ability to adjust their instruction to what they learn from assessment (a recent focus in many districts, as we report in our interview study). Second, students deserve opportunities to show what they know and can do. Reducing the number of classification possibilities reduces students' ability to show where they are in their learning. Third, educators have been working on assessment with the four performance categories in mind all along; they must be assured that their work thus far is valued and that they are headed in the right direction. The capacity needed to design and implement CRTs with sufficient breadth and depth will not come overnight, but we see nothing in this report to suggest that such capacity is impossible to achieve in time.

If the NDE decides to retain the four performance categories, this study makes clear that it will need also to provide considerable guidance to schools in order to design appropriate assessments. To that end, we add the following recommendations to those offered by Impara et al.:

Continue to allow districts to develop their own performance level definitions and decisions. Imposing either performance level definitions or decisions about student performance would undermine a philosophical cornerstone of STARS: that teachers' professional judgment is paramount in assessing student performance.

Using material from this study, develop and disseminate a step-by-step process local districts can use to conduct their own sufficiency studies. Perhaps the most important outcome of this study is the *process* undertaken by professional educators. We recommend abstracting from the process described in the report *and* the recommendations Impara has made about that process a single, straightforward model for how districts can undertake this work. Disseminate the model with the advice to treat this work as a positive professional development opportunity (as reported by the teachers who participated in the Buros study).

Retain focus on multiple measures. In the absence of assessments (NRTs or CRTs) that allow educators to draw consistently valid inferences about student performance on all of the state standards, it is crucial to use multiple measures.

Study Conducted by Elaine Specht

This interview study focuses on principal leadership for standards, assessment, and accountability (SAA). Specht chose to interview five principals in different parts of the state and in varying school sizes. Each principal's district had received either a Very Good or an Exemplary on its assessment portfolio rating for 2003.

In this section, we present first Specht's Executive Summary. (Her full report may be located in Appendix G.) Following the Executive Summary is a discussion that places Specht's findings into the context of the ongoing Comprehensive Evaluation.

**Leading Principals Leading Schools: Current Roles and Practices of Building
Administrators for Standards, Assessments, and Accountability**
EXECUTIVE SUMMARY

Robert Marzano (2003) asserts that "leadership could be the single most important aspect of effective school reform."[11] In this time of emphasis on Standards, Assessments, and Accountability (SAA), five Nebraska principals were asked to identify their role(s) and practices, both changed and continuing, in implementing these reforms. These building leaders worked in district classified as A through D and in different areas of the state. The principals ranged in administrative experiences from 7 to 25 years. All of the schools had received positive marks regarding assessment quality on the state Report Card, 2003.

The roles and practices that the principals shared in common were their involvement in the SAA process, their attention to the data, and concern regarding student performance results. All, however, did things differently in their buildings. One principal did all the statistical work regarding assessment reliability and mastery cut scores for the entire district. Another principal took over the writing of the building portfolio in an effort to relieve some of the work of the teachers. One principal mainly sent his teachers to work with other teachers on assessment issues. All were very cognizant of public accountability regarding their school's student performances. Several instituted direct professional development in their building to work on an identified area of need and administratively followed up so that the practices are becoming embedded into teacher practice. In several schools the building professional development focus was tied directly into their School Improvement process. Communication by the principal regarding student progress ranged over a wide spectrum from weekly charting of

---

[11] *What Works in Schools: Translating Research into Action*. Alexandria, VA: ASCD. p. 172.

Opportunities to Learn to sharing pertinent articles and research with the staff via technology to discussion of SAA at faculty meetings to casual conversation one-on-one with faculty members. Roles of the principals in SAA were reported as leading from the front, leading from the back, and working collaboratively in learning groups with faculty.

Focusing on SAA, these building leaders were both optimistic about practices that are resulting in improvement of student learning in their buildings and honest in expressing apprehension regarding increasing state and national requirements with fewer budgetary resources in some cases and less time to do this work in all cases. Regarding SAA, there was a common unease underlying reflections, such as "What's next?" with its corollary of "Can we successfully meet more expectations in this building?"

Recommendations for the Nebraska Department of Education:
- continue the frequent and timely communication with districts via teleconferences, emails, mailings, and direct conversation
- assist schools (and building leadership) in the application of systems thinking to establish robust connections to School Improvement
- find ways to work directly with more building principals to energize and sustain their leadership of SAA, embed defensible practices, and thus impact student learning
- continue to forge strong, professional relationships with building principals as a means to enhance their own professional development.

**

Findings

This study confirms and adds to our existing knowledge base about principal leadership under STARS. The CEP has found building-level leadership to be crucial to the successful implementation of a local standards, assessment, and accountability (SAA) process. Likewise, Specht finds that principals are leaders of the SAA implementation process in their buildings.

Specht's study substantiates many of the findings of your year two examination of leadership (see year two report, Chapter 3). In these schools with high assessment portfolio ratings, she finds principals paying increasing attention to the following:

- shared leadership
- using data for school improvement and increased student learning
- teamwork within the building to interpret and use data
- role definition (which Specht calls "setting levels of responsibility")
- improved professional development (including more embedded, ongoing, on-site opportunities)

At the same time, challenges continue to plague principals in this system. At the top of this list, not surprisingly, is time. Work load – a related issue – also surfaces as an ongoing concern. Third, resistance from some educators continues to slow progress toward more mature SAA processes. Finally, there is some concern among principals that there is a "moving target" with respect to SAA; that is, the rules of the game keep changing. We had identified each of these concerns in earlier studies.

On the other hand, Specht takes a closer look than we previously had at the sources of support for principal leadership. All principals identified the NDE as well as their respective ESUs as providing assistance. Interestingly, principals discuss existing schools structures and staff collaboration as key supports for their work.

Implications

In general, Specht finds what we had found in previous interviews and surveys: that the principal's job has changed in Nebraska; the traditional management role has been replaced by one in which building administrators are expected to be instructional leaders in a distributed leadership model. Moreover, as Specht aptly summarizes, "Although they were distressed regarding issues of time and feeling the 'need to be on top of things,' all [five principals] stated that learning in their building is more focused for students."

That these school leaders are leading exemplary assessment processes (as rated by the portfolio reviewers) attests to the effectiveness of shared but focused leadership under STARS. Their leadership approaches match the general trends we identified in last years' report; although this sample is very small and is not generalizable, this match between general trend and quality result is surely a good sign.

It should be heartening to the NDE that it is cited by principals alongside ESUs as a support for their work. Further research would uncover what kind of support is most helpful and how the NDE and the ESUs may provide even better assistance.

Recommendations

Our first recommendation, then, is to **research the kinds and quality of support schools are receiving from various outside sources, including ESUs, consortia, higher education, and the NDE**. As we indicated in last year's report, the time has come for an independent evaluation of the kinds and quality of support offered by the ESUs throughout the state. The same is true of other support agencies.

Second, we renew our recommendation from last year to **sponsor a 'leaders of learning' council or academy.** The Leadership Cohort at the University of Nebraska is doing important work with a small group of enrolled leaders each year. What we have in mind is an ongoing professional exchange in a defined forum for leaders, and especially principals, throughout the state to teach each other. As we indicated last year, topics of discussion and shared inquiry might include "Forming Healthy Alliances with External Agencies," "Motivating Teacher Buy-In," and "Models of Leadership for Learning." Leaders might also share models of data processes and SAA processes in general.

Finally, we echo Specht's recommendations to **provide professional development opportunities reserved for principals to learn about interpreting and using data**. This study indicates that principals are key players in SAA. In our interactions with principals over this three-year study, we have encountered a number of principals whose assessment literacy is somewhat lower than that of their teachers. In some cases, these principals become defensive or resistant to assessment. They need a low-stakes learning environment, sponsored by the NDE and/or ESUs, in which they feel comfortable beginning where they are and improving their assessment literacy.

Conclusion

As STARS has evolved, the Nebraska Department of Education – and particularly its Statewide Assessment office – has become increasingly sophisticated in its thinking, its policies, and its behaviors. It is now in a position to capitalize on the strengths of the system and to respond effectively and efficiently to the stubborn and emergent challenges that a school-based, teacher-led system is bound to engender.

Indeed, in this "Year of Accountability," the NDE has dedicated considerable resources – including time, money, and energy – to shepherding the system into a more stable position. This work has resulted in federal approval of STARS and increased investment in the system within the state. Our observations of NDE meetings and workshops (more than twenty in the past year) lead us to conclude that the Department is doing many things right.

First, Pat Roschewski and her staff have worked tirelessly to communicate expectations, requirements, and anticipated changes to STARS. In doing so, they are modeling good teaching by making learning and improvement goals transparent to local educators. This ongoing effort includes

- ♦ many visits to districts for personalized consultation, including the popular "Chats with Pat"
- ♦ development and sharing of a long-range plan for STARS, which rolls out changes until 2008-09
- ♦ pre-portfolio submission workshops for formative feedback
- ♦ inclusion of Nebraska educators in the portfolio review process
- ♦ guidance in shaping the public message around state and federal accountability results (including a well-conceived press conference)
- ♦ clearer, more specific rubrics and checklists for districts to report to the state
- ♦ development and sharing of a set of documents outlining *and modeling* portfolio requirements (mailing, March 2004).

The documents – and especially the sample portfolios and charts – are clear, succinct, and extremely helpful, as a number of educators have indicated to us. We applaud the Statewide Assessment Office for its efforts to clarify and document its expectations and requirements.

Meanwhile, the meetings and workshops continue to be well attended. Educators are generally pleased with the results of such activities. We have heard several teachers and administrators express their appreciation for the assistance they receive from the NDE. Significantly, the roll-out routinely occasions relief among educators, even those who have been worried about constant changes to STARS. As one educator said at a STARS Advisory meeting, "We can do this. I don't see any problem with this." It is

clear to us that the NDE is increasingly perceived by local educators as a partner, rather than merely an administrative entity. This sentiment is reinforced, in our view, by the inclusion of local educators in several important venues, including

- the Commissioner's STARS Advisory Committee
- panels connected with the "Chats with Pat"
- the State of the Schools press conference
- the District Assessment Portfolio review process.

The NDE continues to include and to listen to the expertise of local educators, increasingly relying on native talent rather than importing it from elsewhere. This is a testament not only to the wisdom of the NDE but also to the level of expertise of Nebraska educators.

The NDE and the State Board of Education have made other positive policy and

procedural changes as well. These include

- a streamlining of the reporting procedure for science and social studies standards
- a slow phasing in of technical requirements (e.g., sufficiency) and a reasonable schedule for portfolio reporting
- an option for reporting science and social studies standards at the fifth grade
- an online portfolio submission process.

Each of these changes promises to simplify and streamline the STARS process – a recommendation we have made each year of our study.

In addition, the NDE has kept its focus on professional development for educators, which we believe is the linchpin of the entire STARS system. Efforts in this direction include

- continued assessment literacy focus (NDE workshops, Rick Stiggins visits, partnering with ESUs)
- further alignment work with higher education, including the development of a higher education framework for all 17 institutions for pre-service skills in assessment.

It is crucial that NDE continue its work with higher education to develop innovative programs that integrate assessment into both preservice and ongoing professional development for educators.

Finally, the NDE has demonstrated its ability to look forward with the following initiatives:

- a simple, flexible integration of 3-8 testing into the existing system *for AYP only*, thus maintaining the primacy of state accountability
- validation of the accountability system

♦ an "assessment of assessments" led by assessment expert Sue Brookhart, to design a rubric for evaluating the quality of local assessments

♦ a national conference, scheduled for September 2005, to highlight "lessons learned" under STARS

♦ clarification of the roles of the various partners: practitioners, technical experts, and policymakers

This last item – an unofficial theme in many meetings run by Pat Roschewski this year – has been handled particularly well in STARS Advisory meetings and in decisions about the roll-out of various technical requirements. In our estimation, it is crucial for the sustainability of STARS that all these crucial players remain at the table, working in concert for student learning.

Looking Ahead: A Remaining Challenge

The Validation Framework included in Appendix A demonstrates both the amount of work required to fully validate the state's accountability plan and how committed the NDE is to achieving that goal. The state's involvement – indeed, its leadership – in a national Council of Chief State School Officers consortium devoted to validating accountability systems indicates the high level of seriousness with which the NDE approaches accountability. This effort demonstrates once again that the NDE is engaged in the same kind of work it expects of schools: the gathering, analysis, and use of data to improve its operations for the sake of student and teacher learning.

In sum, we continue to believe that NDE is providing excellent leadership for STARS. We know of no other state in which the department of education collaborates so regularly and democratically with the schools, especially on assessment and accountability, which tend to be inordinately top-down. Many of the educators we visited this year indicated that they view the Department as supportive and responsive to their concerns and questions. In particular, Pat Roschewski, Sue Anderson, Dottie Heusman, and Commissioner Christensen are regularly praised for their accessibility to educators around the state.

However, based on our observations and our interactions with NDE staff, we believe the NDE has an important step to make yet toward ensuring the sustainability of STARS. Specifically, we believe the organization tends to function more as a congeries of sometimes only vaguely related divisions than as a coherent unit. At present, the NDE is division-oriented and some of its employees seem more devoted to patrolling the borders of their division than to contributing to the good of the entire organization and those whom it serves: educators and students. This *modus operandi* not only does not reward creativity; it is sometimes hostile toward it, wasting precious time and talent.

These observations may seem beyond the purview of a STARS evaluation. But we believe the present organizational structure and work culture of the NDE – and particularly the lack of articulation among the various divisions – are hindering the development of STARS, and thus the ability of the NDE to support improvement in the schools.

These conclusions are based on admittedly limited observations and interactions with NDE staff. However, based on our three years of observing and interacting with

NDE staff, we feel confident that at the least, an organizational self-study, perhaps assisted by an outside facilitator, is warranted.

We believe a self-study is appropriate for the NDE because its staff is enormously talented but often underutilized and sometimes disaffected. One purpose of such a study would be to strengthen the staff's understanding of and commitment to the NDE as a "learning organization": one that learns and one that sponsors the learning of its members (NDE employees) and those it serves (educators and students). How well does the NDE function as a learning organization? How could it improve as a learning organization? Collectively examining questions such as these may help forge a shared vision, prompting individuals and divisions to commit to sustaining a creative, participatory work culture. It may also point the way toward necessary organizational shifts, potentially including the integration or reorganization of existing divisions. At any rate, we believe what is needed is an ongoing process of self-examination and creative thinking about the nature and function of the organization. This is precisely what the NDE is asking of schools; it should ask the same of itself.

Our final recommendation, then, is this:

The NDE should conduct a careful, honest self-study of its organization and its workplace culture. It should strive to integrate structurally those parts of its mission that are integrated in its philosophy.

Goal One				
Educators can appropriately and accurately assess and report student performance on content standards using local assessment systems.				
Indicators for Study	**Research Questions**	**Method**	**Who?**	**When?**
District Assessment Portfolio	Is the DAP process reliable and accurate?	Inter-rater reliability of DAP scoring	Buros	Annually
		Outside Evaluator of Portfolio Process	Ellen-Forte-Fast Dave Frisbie	May 28-November 4
		Inside Evaluation of Process	Eight Nebraska participants	July-September
	What is the correlation (if any) between DAP ratings, CRT and NRT data?	Correlational studies	Jody Isernhagen Leon Dappen	June 04/Ongoing
	What are the characteristics of districts where assessment ratings are high? …where ratings are low?	Surveys	Dottie Heusman UNL Comprehensive Evaluation Project (CEP) Jody Isernhagen & Leon Dappen	August 04/Ongoing
	How do educators experience the DAP? e.g., do they believe it appropriately and accurately represents their local assessment system?	Visitations	(CEP) Pat Roschewski CEP	Ongoing
Local district assessments	Are the local assessments used of sufficient quality to accurately	Examination of assessments	Susan Brookhart (CEP)	July – December 04

		measure student performance?			
			Peer review sessions	Teacher teams	Beginning Fall 04-05
Teacher decisions about Classifying Student Performance	To what extent can teachers make appropriate decisions to classify student performance on locally developed assessments?	Performance Classification Study (Sufficiency)		Peer reviews/ESU Buros & CEP	July 03 – July 04

Goal Two				
Student Performance data generated from assessment and graduation rates are being used to inform local school improvement.				
Indicators for Study	Research Questions	Method	Who?	When?
School Improvement Plan	Are districts basing their school improvement goals on collected data from assessment?	Review of school improvement plans External visitations Case studies	NDE Review teams CEP	Ongoing August 2004
Professional Development	What kinds of professional development are available to educators?	Survey	NDE/ESU	April 2004
	Who is providing professional development?	Case studies	CEP	August 2004
	How effective are Nebraska's professional development opportunities?	Case Studies	CEP	August 2004

Instructional Change	How is instruction changing in Nebraska schools as a result of assessment data?	NCA visits/review of data	NDE	Ongoing
		Case studies	CEP	August 2004
Leadership	Who are leaders in Nebraska's accountability system?	Case studies	CEP	August 2004
	What are the characteristics of effective leaders?	Survey	Jody Isernhagen & Leon Dappen (CEP)	August 2004
	How are local leaders modeling effective leadership assessment practices?	Survey	Elaine Specht, Jody Isernhagen & Leon Dappen	August 2004
	How are the leaders in effective schools aligned in their thinking about assessment practices?	Case Studies	Jody Isernhagen & Leon Dappen (CEP)	August 04

Goal Three				
Student performance in reading, mathematics, and writing, as well as graduation rate will improve as a result of data-informed school improvement efforts.				
Indicators for Study	**Research Questions**	**Method**	**Who?**	**When?**

Student Performance on Reading Standards	Have STARS 2001 reading data improved? 2003 reading data?	Correlational Studies	UNLV (CEP)	August 2004
	Have NRT and CRT reading data improved from 2001, 2002, 2003, 2004?	Survey	Dottie Heusman (CEP)	
Student Performance on Mathematics Standards	What are the characteristics of the districts where student improvement is being made or not?	Correlational Studies	UNLV (CEP)	August 2004
	Have STARS 2002 mathematics data improved from 2002-2004 mathematics data?	Correlational Studies	UNLV (CEP)	August 2004
	Have NRT and CRT mathematics data improved from 2001, 2002, 2003, 2004?	Correlational Studies	UNLV (CEP) Jody Isernhagen Leon Dappen	October 04 October 04
Student Performance on Statewide Writing	Have statewide writing results improved? Grade 4 – 2002 / 2004 Grade 8 – 2003 / 2004	Correlational Studies	UNLV (CEP)	April 2004
	Are Statewide Writing results and reading results comparative?	Correlational Studies	UNLV (CEP)	August 2004
	How are statewide writing data being used in classrooms in districts and for programmatic change?	Case studies		

Goal Four				
The accountability system (both state and AYP) classifies school systems correctly (avoids misclassifications.)				
Indicators for Study	**Research Questions**	**Method**	**Who?**	**When?**
Schools Not Meeting AYP	What are the similarities and differences in the schools being classified as needing improvement and those who met all AYP goals?	Correlational Studies	Research in Action UNLV (CEP)	April, 2004
Schools Meeting AYP	What (if any) relationship exists between those schools identified as not meeting state goals and those who did not meet federal goals? How would confidence bands affect the AYP classification?	Correlational Studies	Research in Action UNLV (CEP)	May, 2004

Goal Five				
The consequences of the accountability system are positive ones that contribute to the achievement of the other four accountability goals.				
Indicators for Study	**Research Questions**	**Method**	**Who?**	**When?**
Schools meeting accountability goals	What are the intended and unintended consequences of accountability in schools meeting accountability goals?	Case studies Visitations Survey	CEP NDE Jody Isernhagen & Leon Dappen; Dottie Heusman (CEP)	August 2004 Ongoing August 2004 August 2004
Schools not meeting accountability goals	What are the unintended consequences of accountability in schools meeting accountability goals?	Case Studies Visitations Survey	CEP NDE Jody Isernhagen & Leon Dappen; Dottie Heusman (CEP)	August 2004 Ongoing August 2004

Validation of Nebraska's Standards, Assessment, and Accountability System

"An Accountability System can be said to have validity when the evidence is judged to be strong enough to support inferences that:

1) the components of the system are aligned to the purposes and are working in harmony to help the system accomplish those purposes; and,

2) the system is accomplishing what was intended (and did not accomplish what was not intended.)"

The accountability system will be validated as "working" if evidence is found that the following purposes are goals of the accountability system are being met.

Goal One   Educators can appropriately and accurately assess and report student performance on content standards using local assessment systems.

Goal Two   Student performance and assessment data are being used for school improvement efforts.

Goal Three Student performance in reading, writing, and mathematics have improved because of school improvement efforts based upon assessment and performance data.

Goal Four   The Nebraska accountability system (both state and AYP) classifies schools or school systems correctly (avoids misclassifications.)

Goal Five   The consequences of the accountability system are positive ones and contribute to the achievement of the districts in the state accountability goals

Appendix B
Institutional Review Board
Approval

December 22, 2003

Dr. Chris Gallagher
English
202 Andrews 0333

IRB # 2001-12-108 EX

TITLE OF PROTOCOL: **Study of School-based Teacher-led Assessment and Reporting System**

Dear Dr. Gallagher:

This letter is to officially notify you of the approval of your project's Continuing Review by the Institutional Review Board for the Protection of Human Subjects. It is the committee's opinion that you have provided adequate safeguards for the rights and welfare of the subjects in this study. Your proposal seems to be in compliance with DHHS Regulations for the Protection of Human Subjects (45 CFR 46).

We wish to remind you that the principal investigator or project director is responsible for keeping this Board informed of any changes involved with the procedures or methodology in this study. You should report any unanticipated problems involving risks to the subjects or others to the Board.

It is the responsibility of the principal investigator to provide the Board with a review and update of the research project each year the project is in effect. This approval is valid until **January 15, 2005.**

If you have any questions, please contact Shirley Horstman, IRB Administrator, at 472-9417 or email shorstman1@unl.edu.

Sincerely,

Marcela Raffaelli, Chair
for the IRB

Shirley Horstman
IRB Administrator

Appendix C
The Researchers

Principal Investigator

Dr. Chris W. Gallagher, Associate Professor English, University of Nebraska-Lincoln. Gallagher's research into teaching and assessment has led to the publication of a book, *Radical Departures: Composition and Progressive Pedagogy,* as well as several articles and reviews in journals such as *Phi Delta Kappan, College English*, and *Composition Studies*. He has also worked for the testing industry as a test scorer and training leader. He has won a national outstanding dissertation award, a university-wide junior faculty award, and a college teaching award.

Secondary Investigators

Dr. Chad W. Buckendahl, Director, Buros Institute for Assessment Consultation and Outreach (BIACO), University of Nebraska-Lincoln, is a 2000 graduate of the Quantitative Methods program at the University of Nebraska-Lincoln and holds a master's degree in legal studies from the College of Law at the University of Nebraska-Lincoln. Dr. Buckendahl is involved with test development oversight for the Oklahoma Commission for Teacher Preparation and in the development of the district assessment and accountability system for Nebraska. Prior to joining BIACO, Dr. Buckendahl worked for The Gallup Organization in various capacities.

Tamy Burnett, M.A. Burnett received her master's degree in English Literature at the University of Nevada-Las Vegas and is currently enrolled in the doctoral program in Rhetoric and Composition at the University of Nebraska-Lincoln.

Sandy Byrd. Byrd has completed the course and examination requirements to receive her master's degree in Medieval and Renaissance Literature from the University of Nebraska-Lincoln. She is completing her thesis and is currently enrolled in the doctoral program in Medieval and Renaissance Literature at the University of Nebraska-Lincoln.

Dr. Leon Dappen, Assistant Professor in Educational Administration, University of Nebraska-Omaha. Dr. Dappen received his doctoral degree from the University of Nebraska-Lincoln and is a licensed psychologist. He works as an evaluator for the Nebraska TeamMates Student Mentoring Program, is an external evaluator for an Omaha Public Schools Magnet Assistance Grant and a Banneker CEMS Math/Science grant, and is involved in North Central Association School Improvement activities for several schools. Dr. Dappen has held positions as a teacher, counselor, psychologist, special education administrator, and assistant superintendent in preK-12 schools.

Dottie Heusman, M.Ed., serves as the National Assessment of Education Progress (NAEP) Coordinator for the Nebraska Department of Education. Heusman holds a master's in Educational Administration K-12 and is currently part of a graduate Leadership for Learning Cohort. She has served as a primary grades teacher, Reading Recovery Teacher, and K-12 Curriculum Coordinator.

Dr. James C. Impara, Buros Institute for Assessment Consultation and Outreach. Dr. Impara is a leader in the field of psychometrics, specializing in test recognition and accreditation programs and providing technical oversight for test development. He served as the Director of BIACO from 1995 to 2004 and is involved in the evaluation of school district assessment systems for the State of Nebraska, including the evaluation of local testing programs in all of Nebraska's school districts.

Dr. Jody Isernhagen, Associate Professor of Educational Administration, University of Nebraska-Lincoln. Isernhagen is an evaluator for the Nebraska TeamMates Student Mentoring Progam, an external evaluator for grants with Nebraska Education Service Unit 6 and Crete Public Schools, the primary instructor for the School Improvement Specialist Program, and the State Accreditation and North Central Accreditation External Leader for four school districts. She has been a teacher, assistant principal, principal, Supervisor of Elementary Education, and superintendent in pre-K through 12 schools.

Renee Jacobson, MBA, Ed.S. Jacobson has served as the Superintendent of Plattsmouth Community Schools for the past five years and is currently completing her doctoral degree in Educational Administration, Curriculum, and Instruction through the UNL/UNO Joint Doctoral Program. She has taught science and conducted economic viability and technical capability studies for private industry.

Dr. Gregory Schraw. Professor Educational Psychology, University of Nevada, Las Vegas. Schraw is Co-Director of the Nevada Center for Evaluation and Assessment. He has published extensively in the areas of metacognition and beliefs and cognition, among others. He is author of three books and recipient of two distinguished teaching awards.

Elaine R. Specht. M.Ed. Specht has successfully completed the coursework and comprehensive examination in partial fulfillment of the requirements to obtain her Ed.D. in Curriculum and Instruction from the University of Nebraska-Lincoln. She serves as the Director of Staff Development at Education Service Unit 10 in Kearney, Nebraska.

Susan Wilson. Wilson is on the staff of the University of Nebraska-Lincoln and is currently enrolled in a bachelor's program in Business Administration.

Dr. Keith Zvoch. Assistant Professor of Educational Psychology, University of Nevada, Las Vegas. Zvoch is on staff at the Nevada Center of Evaluation and Assessment. He has authored several articles and research reports.

Appendix D
Research Instruments

A. Interview Study: Using Data
B. Elementary Survey
C. Interview Study: Principal Leadership

A. Interview Study: Using Data

Questions

Language Arts and Math Teachers

I.   How are STARS assessment data reported back to the school?
     1. When are the results made known to you?

     2. By whom?
     3. To whom?
     4. In what context/forum?
     5. In what format?
     6. What, if anything, is at stake in the scores? (For you, your students, the school, the district?)

II.   What have you learned from the STARS assessment data?
     7. What, if anything, have you learned about your curriculum?
     8. What, if anything, have you learned about the classroom instruction you are providing, individually and collectively (in your building)?
     9. What have you learned about your students?
        a. Have the data revealed any achievement gaps among groups of students? If so, in which areas?
        b. Have the data provided information on individual students' learning?

III.   What has been done in response to the STARS assessment data?
     10. Who makes decisions about how to respond to STARS assessment data? Please describe the decision-making process.
     11. What, if any, curricular or programmatic strategies have been developed in response to the STARS assessment data? Provide examples. (New courses, remediation programs, school improvement plan, etc.)
     12. What, if any, instructional or classroom strategies have been developed in response to the STARS assessment data? Provide examples. (Activities, groupings, shifts in emphasis, etc.)
     13. Have you been involved in any of the activities we have discussed? Please describe.
     14. How is your teaching affected, if at all, by these responses to STARS assessment data?
     15. Are the data you get from the STARS assessments useful to you as a teacher? If yes, how so? If no, why not?
     16. If your data revealed achievement gaps, what, if anything, is being done, either programmatically or in individual classrooms, to serve the needs of those students?

IV.   What is working and what is not?
     17. Are you finding that certain curricular or programmatic strategies are working? Explain.

18. Are you finding that certain curricular or programmatic strategies are NOT working? Explain.
19. Are you finding that certain instructional or classroom strategies are working? Explain.
20. Are you finding that certain instructional or classroom strategies are NOT working? Explain.

Writing Teachers

I. How are Statewide Writing assessment data reported back to you?
1. When are the results made known to you?
2. From whom?
3. To whom?
4. In what context?
5. In what format?
6. What, if anything, is at stake in the writing scores? (For you, your students, the school, the district?)

II. What are you learning from the Statewide Writing assessment data?
7. What are you learning about student writing in your school?
8. What are you learning about student writing in your classroom?
9. What are you learning about how you teach writing, either individually or as a school?
10. Have you been able to identify achievement gaps in particular groups of students?
    a. How were these groups identified? By whom?

III. What has been done in response to the Statewide Writing assessment data?
11. Who makes decisions about how to respond to Statewide Writing assessment data? Please describe the decision-making process.
12. What, if any programmatic or curricular strategies have been developed and/or implemented in response to writing scores? (New courses, remediation programs, school improvement plan, etc.)
13. What, if any, instructional or classroom strategies have been developed and/or implemented in response to writing scores? (Activities, groupings, shifts in emphasis, etc.)
    a. Has your approach to 6-Trait writing evolved in response to writing scores? When? How? Why?
    b. Do you emphasize all six traits equally or emphasize one or some over others? Why?
14. Have you been involved in any of the activities we have discussed? Please describe.
15. At what level have any curricular or instructional changes been made? (Classroom? School? District? ESU? Consortium?)

IV. What has or has not worked?
15. Are you finding that certain curricular or programmatic changes are working? Explain.
16. Are you finding that certain curricular or programmatic changes are NOT working? Explain.
17. Are you finding that certain instructional or classroom strategies are working? Explain
18. Are you finding that certain instructional or classroom strategies are NOT working? Explain.

.

Local School Board Members

I. ***How are STARS assessment data reported back to the district?***
1. When are the results made known to you?
2. By whom?
3. To whom?
4. In what context/forum?
5. In what format?
6. How important are the STARS assessment data for the district? For the community?

II. What are you learning from the data?
7. Are you seeing district-wide trends in STARS assessment results in mathematics?
8. In language arts results?
9. In writing results?
10. Are you identifying areas of strength across the district? Provide an example.

11. Are you identifying areas of challenge or weakness across the district? Provide an

    example.

12. Are you identifying achievement gaps across the district? Please explain, including

    identification process.

III. What has been done in response to the STARS assessment data?

13. Who makes decisions about how to respond to STARS assessment data? Please
    describe the decision-making process.
14. What actions, if any, have been taken at the district level in response to what you have
    learned from STARS assessment data? (New programs, new instructional emphases,
    professional development, shifts in funding priorities, school improvement processes,
    etc.)
15. How, if at all, are STARS assessment data used for the purposes of comparison in this
    district? (Previous district results, between schools, between this district and others, vs.
    state average, etc.)

IV. What is working and what is not?
16. Do you have evidence that specific district improvement strategies are working? Please
    explain.
17. Do you have evidence that specific strategies are NOT working? Please explain.
18. What, if anything, does your district need to do differently or better in response to
    STARS assessment data?

District Superintendents

I.   How are STARS assessment data reported back to the district?

1. When are the results made known to you?
2. By whom?
3. To whom?
4. In what context/forum?
5. In what format?
6. How important are the STARS assessment data for the district? For individual schools? For teachers? For students?

II.   What are you learning from the STARS assessment data?

7. Are you seeing district-wide trends in STARS assessment results in mathematics?
8. In language arts results?
9. In writing results?
10. Are you identifying areas of strength across the district? Provide an example.
11. Are you identifying areas of challenge or weakness across the district? Provide an example.
12. Are you identifying achievement gaps across the district? Please explain, including identification process.

III.   What has been done in response to the STARS assessment data?

13. Who makes decisions about how to respond to STARS assessment data? Please describe the decision-making process.
14. What actions, if any, have been taken at the district level in response to what you have learned from STARS assessment data? (New programs, new instructional emphases, professional development, shifts in funding priorities, school improvement plans, etc.)
15. How, if at all, are STARS assessment data used for the purposes of comparison in this district? (Previous district results, between schools, between this district and others, vs. state average, etc.)

IV.   What is working and what is not?

16. Do you have evidence that specific district improvement strategies are working? Please explain.
17. Do you have evidence that specific strategies are NOT working? Please explain.
18. What, if anything, does your district need to do differently or better in response to STARS assessment data?

Principals

I.   How are STARS assessment data reported back to the school?

1. When are the results made known to you?
2. By whom?
3. To whom?
4. In what context/forum?
5. In what format?
6. How important are the STARS assessment data for the school? For teachers? For students?

II.   What are you learning from the STARS assessment data?

7. Are you seeing school-wide trends in STARS assessment results in mathematics?
8. In language arts results?
9. In writing results?
10. Are you identifying areas of strength in the school? Provide an example.
11. Are you identifying areas of challenge or weakness in the school? Provide an example.
12. Are you identifying achievement gaps in the school? Please explain, including identification process.

III.   What has been done in response to the STARS assessment data?

13. Who makes decisions about how to respond to STARS assessment data at the school level? Please describe the decision-making process.
14. What actions, if any, have been taken at the school level in response to what you have learned from STARS assessment data? (New programs, new instructional emphases, professional development, shifts in funding priorities, school improvement processes, etc.)
15. How, if at all, are STARS assessment data used for the purposes of comparison in this school? (Previous school results, between schools, between grades, between teachers, against state average, etc.)

IV.   What is working and what is not?

16. Do you have evidence that specific district improvement strategies are working? Please explain.
17. Do you have evidence that specific strategies are NOT working? Please explain.
18. What, if anything, does your school need to do differently or better in response to STARS assessment data?

B. Elementary Assessment Survey

A. Background Questions: *Please write your answer in the space provided.*	
Years of teacher experience in your current district:	Total years of teaching experience:
Grade level taught:	

Please answer the following questions using the scale on the right. **B. Administrative Support**	Strongly Agree	Agree	Disagree	Strongly Disagree
1. The administration in my district provides support for the language arts standards and assessment process by providing me with professional development opportunities.				
2. The administration in my district communicates with me in regards to standards and assessments.				
3. Our district looks at multiple years of data for the purpose of school improvement.				
4. The district administration has given me moral support throughout the assessment process.				
5. The district administration listens to me when I have suggestions or concerns about our assessment process.				
C. Integration				
1. Assessment is an integral part of my classroom instruction.				
2. Multiple assessments are given to get more valid student results.				
3. There are too many assessments in my district.				
4. I am very familiar with the assessments given in my district at all grade levels.				

	Strongly Agree	Agree	Disagree	Strongly Disagree
5. A statewide test would be more appropriate for my district than our current assessment system.				
D. Collaboration				
1. Teachers are given enough time to work collaboratively to develop assessments.				
2. Teachers are given enough time to work collaboratively to review assessment results.				
3. I was actively involved in the assessment process from the beginning.				
4. All teachers in my district are involved in the assessment process.				
E. Accountability				
1. In my district, students are accountable for their own learning.				
2. In my district, teachers are accountable for their student's learning.				
3. Our district assessments measure what our students really know and can do.				
4. The norm-referenced assessment given to our elementary students measures what our students really know and can do.				
5. The Language Arts assessments developed by our district provide me valuable information about my students.				
6. I have changed my classroom instruction as a result of our assessment process.				
7. If agree or strongly agree with #6 answer the following: The changes I have made have benefited the students.				
F. Portfolio				
1. I understand the Six Quality Criteria for developing good assessments.				
2. The six quality criteria are good indicators of the quality of my district's assessment process.				
3. The district assessment portfolio rating system is fair.				
4. The district assessment portfolio is an accurate representation of my district's assessment process.				

In the last 3 years I have attended classes/workshops pertaining to language arts at the following locations: (*Please indicate the number of times you have attended.*)

    ____   ESU

    ____   at my school district

    ____   Dept. of Education

    ____   University or college level

Comments: Feel free to use the back of this page to elaborate on any of these topics.

Thank you for taking the time to complete this survey. Your help with this study is important to the Nebraska Department of Education in helping them provide support to districts throughout the state.

C. Interview Study: Principals

Interview Protocol

Interviewee:_____ Date:_____

Organization:_____

Title:_____ Location:_____

Introduction

I want to thank you for taking the time to talk with me. My questions will cover the following topics: the process for implementing SAA in your building; leadership roles and practices associated with SAA; challenges for leadership associated with the implementation of SAA; and supports for leadership associated with the implementation of SAA.

You have had the opportunity to think about and review the questions. I want to understand your perspective so please feel free to discuss your views candidly. I will ask you additional questions for clarification. Are you ready to begin?

INTERVIEW QUESTIONS: PRINCIPALS

1. Who provides building-level leadership for standards, assessment, and accountability (SAA) in this school?

2. As principal, what is your role in the school's SAA process?

3. Please describe your school's progress in implementing SAA.

4. How, if at all, has the implementation of SAA changed your role as principal?

5. How, if at all, has the implementation of SAA changed your practices as a principal?

6. What, if anything, has hindered the implementation of SAA in your building? If appropriate, please tell me how you have responded to obstacles.

7. What, if anything, has supported the implementation of SAA in your building? Please be specific.

8. In your view, what skills or abilities are most important for principals regarding SAA?

An Examination of 2001 to 2003 Nebraska Criterion Referenced Tests, Norm Referenced Tests, and District Portfolio Ratings for Reading at Grades 4, 8, and 11

Dr. Jody Isernhagen, University of Nebraska-Lincoln
Dr. Leon Dappen, University of Nebraska-Omaha

Introduction

With the emphasis to demonstrate increased achievement for all students and all schools to be accountable to their constituents, many states have developed or adopted state mandated tests to assess student academic performance. Nebraska stands alone, rated by the 1999 Education Week Report Card as a "C" largely because it does not measure school performance by a state-wide mandatory test. According to the report, the state is "lagging behind" in accountability (Editor, 1999). Yet, this seems incongruent as the children in the Nebraska rank among the top 10 nationally in most generally accepted measures of academic success (Editor, 1999). Instead Nebraska has made a conscious decision to lead the way in a new system of accountability that focuses upon building assessment literacy among educators and enhancing student performance through the use of a quality assessment system.

Nebraska's Quality Assessment System
The Nebraska School-based Teacher-led Assessment and Reporting System (STARS) is unique in several ways. As described by Gallagher (2004) this system:
- is a system of local assessments, not a state test;
- promotes a balanced approach to assessment, using multiple measures;
- involves evaluation of achievement and of assessment quality;
- uses classroom-based assessments for reporting; and
- includes no high-stakes testing.

In the STARS process districts first adopt local or state standards. Districts then submit an assessment plan that includes locally developed criterion referenced tests (CRTs) to assess the district's standards at the identified grade levels. They may also choose to use some items of their locally chosen norm referenced test (NRT) to assess some of their standards. The districts then submit a District Assessment Portfolio to the Nebraska Department of Education. Portfolios are reviewed and rated by independent

experts using six quality criteria outlined by the Buros Center for Testing (Plake & Impara, 2000):

1. Assessments reflect state or local standards.
2. Students have an opportunity to learn the content.
3. Assessments are free from bias or offensive language or situations.
4. The level is appropriate for students.
5. There is consistency in scoring.
6. The mastery levels are appropriate.

The focus of the STARS assessment system is to keep teaching and learning at the center of the educational process, promoting high-impact, not high-stakes assessment (Gallagher 2004).

STARS began in 2000 and the first scores were released publicly in the State of the Schools Report by the Nebraska Department of Education on their website and in newspapers across the state in the fall of 2004 in the area of reading. The Nebraska Department of Education collects district and school data including: the percent of students meeting local defined proficiency for criterion-referenced assessments; the average percent of students in the top two quartiles on the district chosen norm-referenced test; and the District Assessment Portfolio rating, a summary based on scores from the six quality assessment criteria.

Purpose of the Study

This study examines the first achievement data available for the Nebraska STARS program, comparing the reading scores in grades four, eight, and eleven from 2001 to scores for 2003. Both locally developed CRT scores and NRT scores were examined. District Assessment Portfolio ratings based on these reading scores were also examined. The research questions for this study were:

1. What are the differences in the CRT reading scores in grades four, eight, and eleven from 2001 to 2003?
2. What are the differences in the NRT reading scores in grades four, eight, and eleven from 2001 to 2003?
3. What are the differences in the reading district assessment portfolio ratings in grades four, eight, and eleven from 2001 to 2003?
4. Based on this data, what are the implications for the Nebraska STARS program?

Research Design and Methodology

School Districts Studied

Data were included for Class 3, 4, and 5 school districts. Class 3 school districts (209 school districts) are represented by any school district with territory having a population of more than 1000 but less than 150,000 inhabitants (Nebraska Education Directory). Class 4 school districts (Lincoln only) have a territory with a population of 100,000 or more inhabitants with a city of the primary class within the territory (Nebraska Education Directory). Class 5 school districts (Omaha only) with a territory having a population of 200,000 or more inhabitants with a city of the metropolitan class within the territory (Nebraska Education Directory). The districts in this study represent just over 94 % of the public school students in Nebraska. The district data for this study

were included on the state website and cooperation for use of the data was facilitated by the Nebraska Department of Education.

Score Definitions

For CRTs, the score used was the percentage of students meeting the district defined mastery level for their locally developed measure. For NRTs, the score used is the percent of students scoring in the top two quartiles on the nationally standardized test used by that district (e.g., California Achievement, Iowa Test of Basic Skills, etc.). While the NRT used will vary, the data reported (percent of students in the top two quartiles) is constant for all districts. The score used for the reading District Assessment Portfolio is the summary score assigned by the external expert visiting team based on the previously mentioned six quality assessment criteria. Summary portfolio rating categories are "Unacceptable," "Acceptable," "Good," "Very Good," and "Exemplary" (Plake & Impara, 2000).

District Portfolio Process

Districts complete the portfolio each spring, including requested information supporting each of the six rating criteria. The portfolio is then submitted to the Nebraska Department of Education in the summer of each year. A contract evaluation agency, Buros Institute for Assessment Consultation and Outreach, arranges for a panel of external reviewers, the District Assessment Evaluation Team. This team is comprised of professionals with an earned doctorate in educational measurement and come from Nebraska and several other states. A rubric has been developed for each of the six quality criteria including the qualifications of those involved, the quality of the process used, and the results reported. The external review team is trained in the rubric review process with significant attention given to inter-rater reliability. Each of the six criteria is rated "Not Met," "Met – Needs Improvement," or "Met." These ratings are then summarized as prescribed by the developers of the system for the reading District Assessment Portfolio score.

Data Analysis

The unit of analysis for this study is the composite of class 3, 4, and 5 school districts for the state of Nebraska in reading at grades four, eight and eleven. Based on the fact that CRT scores are unique to each district, NRT scores vary with the standardized test chosen by that district, and portfolio ratings are based largely on the CRT measures, the data is described as unconventional. Traditional statistical approaches for significance, therefore, were not appropriate. Descriptive data was reported and discussed.

Results

   As shown in Table 1, the percentage of students attaining proficiency on the reading CRT increased at every grade, fourth (5.19%), eighth (.82%), and eleventh (.98%). As reflected in Table 2, NRT scores increased at two grades, fourth (2.63%) and eleventh (1.19%), and decreased at eighth grade (-.56%).

   On the reading district assessment portfolio, the median score, assigning a score of "0" for Unacceptable, "1" for Acceptable, "2" for Good, "3" for Very Good, and "4" for Exemplary, increased for each grade. Fourth grade increased from 3.0 to 3.5. Eighth grade increased from 3.0 to 4.0, eleventh grade increased from 3.0 to 4.0. As reflected in Table 3, the number of districts at ratings "0" through "3" was significantly fewer from 2001 to 2003, reducing from 436 to 266. The number with a rating of "4" (Exemplary) greatly increased from 2001 to 2003, from 107 to 277. This indicates a strong general increase in summary portfolio ratings.

   In examining the changes in portfolio ratings at each grade level (Table 4), at fourth grade, 35% of the districts went down in their rating, 12.2% stayed the same, and 52% increased. At eighth grade, 12.8% of the districts went down, 32.6% stayed the same, and 54.7% increased. At eleventh grade, 13.2% of the districts went down in their rating, 32.4% stayed the same, 54.4% increased. Of those that went down in ratings, the strong majority went down one point (a significant number from the highest rating of "4" to "3"). Of those that went up, many increased two or three points, the single biggest number of increases was the group that increased from "3" to "4." At each grade, over half of the districts increased their reading district assessment portfolio rating from 2001 to 2003.

   In 2001, 6.7% of the fourth grade portfolios were rated Unacceptable, 8.8% of the eighth grade portfolios, and 9.9% of the eleventh grade portfolios. By 2003, those percentages were all reduced by half or more, 3.3%, 3.9%, and 3.7% respectively. In 2003, over 96% of the districts were rated Acceptable or better.

Discussion

Differences in CRT scores

   There are gains in the average percent of students mastering the reading CRT scores from 2001 to 2003 at all grades four, eight and eleven. The change score in the 2001 to 2003 reading CRT at the fourth grade represents an increase of over 5% of the students in these Nebraska schools demonstrating proficiency in the reading standards at that grade. This single finding was the most significant change identified from the study. Fourth grade is where students are formally assessed on the new criterion referenced measures for the first time. Students demonstrated this gain most strongly on measures developed to reflect the standards at their level rather than on the norm referenced tests.

   While demonstrating improvement, students at the eighth and eleventh grades did not show the same level of improvement. It will be important to watch the longitudinal progress of these fourth grade students as they move forward and continue to be assessed using criterion-referenced measures to assess mastery of the reading standards at the

eighth and eleventh grade levels. We will also need to examine other academic areas (math, science and social studies), as they are included in the system.

Differences in NRT Scores

While the NRT used will vary from district to district, the NRT score being examined is the same. There are gains in the percent of students in the top two quartiles of the NRT score from 2001 to 2003 for the fourth and eleventh grades. The eighth grade percent of students in the top two quartiles of the NRT score from 2001 to 2003 was slightly lower. There is, however, an overall slight increase in the total percentage of students in the top two quartiles. While the increase is not striking, it is definitely something to watch, as large group NRT scores are generally quite resistant to change. Once again, we must continue to examine the longitudinal effect of these scores and other academic areas, as they are included.

A study conducted by the Buros Center for Testing at the University of Nebraska in 1998 revealed that the five primary standardized achievement tests match only 35%-40% of the Nebraska State Standards (Roschewski, Gallagher, & Isernhagen, 2001). It would, therefore, not have been surprising to see NRT scores decline somewhat through the local standards and CRT process. It is important to note that the attention paid to CRT assessment activities did not appear to have a negative impact on NRT scores in this study.

Differences in Portfolio Ratings

There were strong gains at all grade levels in the reading District Assessment Portfolio ratings from 2001 to 2003. The portfolio is developed each year by local district staff responding to the previous year's external rater comments and any training and improvement activities that have been carried out over the year. This increase would indicate that district staff has increased their knowledge in assessment and its use in instruction and importance in increasing student achievement. This would be expected as the district assessment portfolio is developed based on the six quality criteria that are used by schools when developing their criterion-referenced measures and also used when measuring the quality of the district assessment portfolios. The curriculum standard, instruction, and assessment activities tie together to inform each other and build a continuously self-improving educational program. This support for the viability of the portfolio process is the "linchpin" for the STARS system.

Implications for Nebraska STARS

While these are first-year findings and should be interpreted with caution, they do provide an initial base of support for the district assessment portfolio process in establishing the creditability of the school-based, teacher-led assessment and reporting system (STARS). This is particularly important in Nebraska's accountability system as student performance is not based on a common statewide measure but on locally developed criterion-referenced assessments.

The very nature of STARS and the assessments used makes comparisons very difficult. There is a need to work with the educational measurement community to

examine new ways to determine the success of the model. While approaches such as STARS are strongly supported by the measurement community from a standpoint of student learning, little has been done to support the acceptance of what is referred to as unconventional data that the approach is based on as evidence of school success.

Much more work is needed in continuing to examine longitudinal results and the impact of refinements and corrections as Nebraskans problem solve concerns and the program evolves. One of the biggest challenges will be the ongoing effort to work with the United States Department of Education to ensure the STARS program will allow Nebraska to meet the No Child Left Behind (NCLB) requirements. It should be noted that this goal has become a multifaceted (educational, political, statistical) issue and will require the focused attention from individuals from each perspective who are committed to preserving the STARS philosophical approach.

This study supports the continuation of the STARS system. As noted by Roschewski, Gallagher and Isernhagen (2001), the STARS plan brings together the best of both worlds; student learning is foremost, but public accountability is provided as well. While avoiding the "unintended negative outcomes" of high-stakes assessments, STARS places the responsibility of teaching and learning where it belongs. The teachers and administrators who create high-quality learning environments are directly involved in affirming the quality and in measuring the learning of students in their classrooms.

References

Editor (1999, March 1). Quality Counts-State of the States Report. *Education Week.* Retrieved May 26, 2004 from: <http.//www.edweek.org/sreports/qc99>

Gallagher, C. W. (2004, January). Turning accountability tables: Ten progressive lessons from one "backward" state. *Phi Delta Kappen.* 251-261.

Nebraska Education Directory (2003-2004). *Nebraska Department of Education Directory.* Lincoln, NE. 106th Edition, 170.

Plake, B. S. & Impara, J. C. (2000). Technical quality rubric for Nebraska's district performance ratings. Lincoln, NE: Buros Center for Testing.

Roschewski, P., Gallagher, C. & Isernhagen, J. (2001, April). Nebraskans reach for the STARS. *Phi Delta Kappan.* 611-615.

Table 1
Change in Percentage of Students Demonstrating Proficiency on Criterion Referenced Tests

CRT	Grade	n	2001 % of Students	n	2003 % of Students	Difference
	4	182	74.33 %	182	79.52 %	+ 5.19 %
	8	183	73.70 %	183	74.52 %	+ .82 %
	11	183	73.70 %	183	74.68 %	+ .98 %

Table 2
Change in Percentage of Students in Top Two Quartiles on Norm Referenced Tests

NRT	Grade	n	2001 % of Students	n	2003 % of Students	Difference
	4	180	64.20 %	183	66.92 %	+ 2.63 %
	8	181	62.78 %	183	62.22 %	- .56 %
	11	180	60.21 %	180	61.40 %	+ 1.19 %

Table 3
Number of Districts Changing in Reading District Assessment Portfolio Ratings

Grade	n	Unacceptable 2001	2003	Acceptable 2001	2003	Good 2001	2003	Very Good 2001	2003	Exemplary 2001	2003
4	180	12	6	35	5	9	1	86	78	38	90
8	181	16	7	38	4	5	1	88	75	34	94
11	182	18	7	36	4	8	1	85	77	35	93
Total	543	46	20	109	13	22	3	259	230	107	277

Table 4
Number of Ratings Units Districts Changed Up or Down on Reading District Assessment Portfolios

Grade	n	-4	-3	-2	-1	Same	+1	+2	+3	+4
4	180	1	1	2	59	22	51	16	24	4
8	181	0	1	2	20	59	51	15	27	6

11	182	2	2	7	59	53	12	12	29	5

Appendix F
Quantitative Study B

Gregory Schraw, Ph.D.
Keith Zvoch, Ph.D.

Final Report on the Analysis of STARS Data

Background and Purpose

The purpose of this report is to document STARS and NRT assessment results using data from the 2000-2001 and 2002-2003 academic years. STARS (School-based teacher-led assessment and reporting system) places special responsibility on each school district to create local criterion-referenced tests (CRTs) in reading and writing. The goal of STARS is to tailor assessments to local curricular goals in order to better align instruction and assessment. STARS requires districts to set learning goals and develop assessments to ensure that learning goals are met. The intent of STARS is to increase the reliability and validity of accountability decisions (Marion, White, Carlson, Erpenbach, Rabinowitz, & Sheinker, 2002). STARS also is intended to complement norm-referenced tests (NRTs) such as the Iowa Test of Basic Skills (ITBS). The main goal of the analyses reported here is to evaluate and assess relationships between the locally-constructed STARS assessments and the national NRT assessments and to determine whether district characteristics are predictive of student performance on these measures.

It is important to note four caveats in relation to interpreting these results. First, STARS assessments may differ from district to district depending on the learning goals established by each district. Comparisons among districts should be interpreted with caution and causal comparisons among districts are unwarranted. Second, results of these analyses may be affected by district size. Larger districts yield data that is more stable than that associated with smaller districts. Third, some districts did not report data, which may affect the generalizability of the results. Fourth, results presented here are based on the "school district" as the unit of analysis. A district average may not be representative of all schools within that district.

Research Questions

Four research questions were identified in consultation with Dr. Pat Roschewski. These questions align with goals three and four of the master plan to validate Nebraska's

standards, assessment, and accountability system. These questions focus on data collected in 2001 and 2003. Comparisons with 2003-2004 data will be considered in a separate proposal.

Question 1 addresses the relationship between STARS assessments and NRTs collected concurrently on students in the 2001 and 2003 school years. The main purpose of this analysis was to summarize performance on each type of measure and examine the relationship between CRTs and NRTs within and between the two-targeted years.

Question 2 addresses change in two different cohort groups from 2001 and 2003. By cohort, we mean a group of students who progress through the educational system together. Students from the 2000-2001 academic year represent a different cohort than students from the 2002-2003 academic year. Differences between these cohorts could be due to instruction or individual differences of students. Because cohorts are naturally occurring, rather than randomly assigned, differences across cohorts are suggestive rather causal in nature. Nevertheless, cohort analyses enable us to draw descriptive conclusions about progress within districts and across the state. The main purpose of these analyses was to compare change among the two cohort groups using STARS and NRT data.

Question 3 addresses the relationship between state writing data and STARS reading data for grade 8 students in the 2002-2003 school year. One purpose of this analysis was to determine whether writing and reading scores are correlated across districts. A second purpose was to investigate which variables are most predictive of writing and reading success at the district level. We used two main types of variables, including student demographic information such as percent ELL and percent free and reduced lunch (FRL), as well as teacher demographics such as annual salary and percent of teachers with graduate degrees.

Question 4 addresses the issue of district compliance with state and federal accountability standards. The purpose of this analysis was to identify factors that discriminate between districts that are in compliance with state and federal (No Child Left Behind) standards and those that are not. A variety of student and teacher variables were used to determine which of these variables, or what combination of these variables, best explains inclusion in the compliance versus no-compliance groups. This information may help the NDE and individual districts focus on variables of special importance with respect to reaching state and federal accountability criteria.

Methods and Procedures

The state of Nebraska has 517 school districts. Accountability data on each district is posted on the NDE website. However, to protect student and teacher anonymity, the NDE suppresses data from some of the smaller school districts. Our analyses focused on districts with a minimum of 50 students to allow for dissemination of accountability data. Typically, our analyses included between 200 and 300 districts.

A master database was compiled using over 20 separate data files available on the NDE website. Individual data files usually had some missing data unrelated to district size. For example, approximately 10% of districts did not report average teacher salary. The cause of missing data is unknown to the authors of this report. In addition, it is unclear whether missing data affects in any significant way our interpretations or the

generalizability of the results (Little & Rubin, 1987; Schafer, 1997). Replication across different cohorts would help address this issue.

Variables used in the analyses were screened prior to analyses. Most variables were distributed normally. When a variable was skewed and/or kurtotic, the variable was trimmed by excluding extreme scores unless otherwise noted. For example, when comparing STARS reading and writing data, a disproportional number of districts reported that 0% of students reached proficiency. These districts tended to be small in size, usually with less than 50 students. We excluded schools with less than 50 students in this case.

Several aspects of the master database and analytic procedures should be noted.

1. STARS reading data were analyzed in terms of the aggregate total score reported by NDE. The total score was used for two reasons. The greater reliability of the total score (relative to individual standard scores) was the first reason. The second was to reduce the number of analyses reported to a manageable size.

2. Separate analyses were conducted for 4th, 8th, and 11th grades. This level of analysis allowed us to examine the percentage of students reaching proficiency at each of the three tested levels (and/or compliance with state and federal standards).

3. There were no STARS writing data for the 2000-2001 academic year, so a cohort writing comparison could not be conducted.

4. The English Language Learner (ELL) variable tended to be highly skewed, with most districts reporting very few ELL students. We included this variable in our analyses *without deleting 0% districts* (i.e., districts without ELL students). This may lead to a conservative estimate of the effect of ELL students due to restriction of range among scores.

Results

A number of different analyses were conducted. The general purpose and nature of the statistical procedure for each research question is described below. An attempt was made to reduce large amounts of data into concise summary tables. Complete data files and statistical analyses are available upon request.

Question 1. We examined the mean percentage of students meeting proficiency for 4th, 8th, and 11th grade for STARS, math NRT, and reading NRT scores for the 2001 and 2003 academic years. We also computed the correlation across the 2001-2003 years for each variable, as well as computing the correlation between math and reading within each cohort.

Table 1 shows the mean percentage of students who reached proficiency on six different performance measures. In most cases, 50% to 65% of students reached proficiency. Proficiency levels were higher for grade 11 than grades 4 and 8. One possible reason is that assessments are more specific at grade 11 and may be better aligned to standards. It should also be noted that whereas the 2003 student cohort was more proficient than the 2001 cohort on the STARS assessment, the reverse was generally true of cohort performance for the NRT measures.

Table 1: Means and Standard Deviations for 2001 and 2003 Percent Proficient for STARS and NRT Data

	4thGrade		8th Grade		11th Grade	
Type of Performance Measure	Mean	SD	Mean	SD	Mean	SD
Percent STARS Proficient 2001	.59 N = 306	.32	.48 N = 382	.37	.73 N = 250	.17
Percent STARS Proficient 2003	.63 N = 281	.33	.56 N = 326	.34	.74 N = 242	.13
Percent NRT Reading 2001	.64 N = 209	.14	.62 N = 201	.13	.58 N = 174	.14
Percent NRT Reading 2003	.52 N = 263	.29	.47 N = 305	.30	.59 N = 203	.16
Percent NRT Math 2001	.66 N = 204	.16	.66 N =197	.16	.66 N =173	.14
Percent NRT Math 2003	.53 N = 260	.31	.51 N = 303	.33	.65 N = 205	.16

Note: N equals number of school districts.

Table 2 shows the correlations between 2001 and 2003 district performance measures. Rows 1-3 of Table 2 indicate whether the two cohorts have similar performance over time. In general, the 2001 and 2003 cohorts perform similarly in 4th and 8th grades, but do not perform similarly among 11th grade students (i.e., with the exception 11th graders, districts that have high (or low) levels of proficiency in 2001 have similar rank proficiency levels in 2003). The lower 11th grade correlations may reflect differences (or a change) in either instruction and/or assessment at the high school level.

The data in rows 1-3 also indicate that the 2001-2003 cohort correlations tend to be much larger for STARS assessments compared to reading and math NRT scores. This may be due to better alignment between instruction and STARS assessments at the district level.

Finally, rows 4-5 of Table 2 indicate that math and reading NRT performance is highly correlated within a cohort group. Those who do well in math tend to do well in reading. This relationship is strong in the 2001 cohort and very strong in the 2003 cohort. The change between 2001 and 2003 may be due to better alignment of instruction and assessment as districts develop and implement STARS.

Table 2: Correlations Between 2001 and 2003 STARS and NRT Data

	4thGrade	8th Grade	11th Grade
STARS 2001-2003 Proficiency	.75	.88	.23
NRT 2001-2003 Reading Proficiency	.55	.41	.40
NRT 2001-2003 Math Proficiency	.47	.45	.30
NRT 2001 Math and 2001 Reading	.61	.63	.62
NRT 2003 Math and 2003 Reading	.91	.95	.62

Note: All correlations are significant at the p < .01 level.

Question 2. We examined the mean change in the percentage of students meeting proficiency for 4th, 8th, and 11th grade for STARS, reading NRT, and math NRT scores for the 2001 and 2003 academic years. The purpose of these analyses was to examine whether there is a systematic change in cohort performance over time.

The No Child Left Behind (NCLB) federal legislation mandates that schools demonstrate adequate yearly progress (NCLB: No Child Left Behind, 2002). For NCLB to be met, schools must demonstrate a higher rate of proficiency across successive student cohorts. Thus, NCLB expects the 2003 cohort to have a higher percentage of students reaching proficiency than a 2001 cohort. We computed change scores (i.e., a value of .05 represent a 5% increase in the percentage of students reaching proficiency, whereas a score of -.05 indicates a 5% decrease) and conducted one-tailed significance tests on the change scores to determine whether proficiency rates increased from 2001 to 2003.

Table 3 reveals two different results. The first is that the level of proficiency increased significantly between cohorts when using STARS assessments. The second result indicates that reading and math NRT proficiency levels did not change over the same time period. There are two plausible explanations for this pattern. The first is that the increase in STARS data is due to chance factors. A second explanation is that STARS assessments are better aligned to local district instruction and therefore yield more sensitive measures of progress. While it is impossible to disentangle the effects of chance and better alignment between instruction and assessment, it seems reasonable to assume that local assessments characteristic of STARS would be better able to detect change.

It is important to note that yearly progress shown in Table 3 is small. NCLB mandates progress of 3% to 5% a year on average. In contrast, Linn (2000, 2003) reports a general trend of 1% progress per year. The STARS progress shown in Table 3 is consistent with the trends reported by Linn (2003).

It should be noted that the difference scores shown in Table 3 are based on districts reporting data for both 2001 and 2003; thus, the number of districts is smaller than in Table 1. It is possible that complete data from all districts might have changed the results reported in Table 3.

Table 3: Means, Standard Deviations, and t Tests for Change in Percent Proficient Scores

	4th Grade		8th Grade		11th Grade	
	Mean	SD	Mean	SD	Mean	SD
STARS Percent Change 2001-2003	.050	.15 N = 218 t = 4.18 p < .01	.018	.16 N = 288 t = 1.87 p < .05	.019	.18 N = 228 t = 1.65 p < .05
NRT Reading Percent Change 2001-2003	.003	.14 N = 173 t = .36 p > .70	-.001	.16 N = 180 t = -.10 p > .90	.014	.17 N = 154 t = 1.06 p > .25
NRT Math Percent Change 2001-2003	.008	.18	.007	.16	-.002	.17

	N = 167	N = 174	N = 154
	t = .62	t = .59	t = -.18
	p > .50	p > .50	p > .80

Note: N equals number of school districts. Significance of t values are based on one-tail tests.

Question 3. We examined whether student and teacher variables were significant predictors of the percent of students reaching proficiency on STARS reading and writing performance measures using data from 8th grade students. Tables 4 and 5 list a number of predictor variables which are described in detail on the NDE website. We conducted two regression analyses; one using 2003 STARS reading data, and the second analysis using 2003 STARS writing data. The correlation between percentage of students reaching reading and writing proficiency was .326. Although this correlation is statistically significant, it is somewhat low in our opinion. It suggests that there is a small degree of overlap in reading and writing proficiency at the district level.

The data in Table 4 are based on 219 districts. All variables were entered into the regression equation at the same time to predict the percentage of students reaching proficiency on the STARS reading assessment. Tables 4 and 5 show the mean percentage, standard deviation, beta weight (i.e., the relative contribution of the variable in the prediction equation), t value, p value (i.e., level of statistical significance), and the partial correlation (i.e., the relationship between each variable and percentage proficient when other variables are held constant).

The total correlation of all variables with percent reading proficient was .38. R^2, a measure of the amount of variation shared among the predictor variables and percent proficient, was .14. Thus, the variables collectively explained 14% of the variation in percent reading proficient. This value is considered small (Cohen, 1988) and suggests that the student and teacher variables do not predict proficiency with a high degree of accuracy. Nevertheless, some variables are better predictors than others. Attendance rate and the percentage of teachers with a Master's Degree were statistically significant predictors. Districts in which attendance rates are higher and employ more teachers with advanced degrees tend to have higher proficiency rates. Percent ELL also explained a significant proportion of variance.

Table 4: Results of Regression analysis on STARS Reading Data

Variable	Mean	SD	Standardized Beta	t-test value	*p* value	Partial Correlation
ELL %	.018	.118	-.168	-2.17	.03	-.15
SPED %	.155	.045	-.041	-0.58	.56	-.04
FRL %	.348	.147	-.097	-1.15	.25	-.07
AR	.957	.018	.180	2.00	.04	.13
Graduation %	.897	.250	.012	0.17	.86	.01
Mobility %	.093	.072	.108	1.13	.26	.07
Accomd %	.050	.073	-.107	-1.49	.13	-.10
ATS	35576	3243	-.066	-0.76	.45	-.05
AYTE	16.48	2.77	.034	0.46	.65	.03
PTM	.287	.133	.187	2.28	.02	.15

Note: ELL % = English language learner %; SPED % = special education %; FRL = free and reduced lunch %; AR = attendance rate; Accomd % = accommodation %; ATS = average teacher salary; AYTE = average

years teaching experience; PTM = percentage of teachers with masters degree. Results based on 219 districts.

The data in Table 5 are based on 234 districts. All variables were entered into the regression equation at the same time to predict the percentage of students reaching proficiency on the STARS writing assessment. The total correlation of all variables with percent writing proficient was .30. R^2 was .09. This value is considered small (Cohen, 1988) and suggests that the student and teacher variables do not predict proficiency with a high degree of accuracy. In addition, slightly less variation is explained in writing proficiency compared to reading proficiency. Two variables, Percent ELL and graduation rates, were statistically significant predictors. Note that the beta weight of Percent ELL is negative. This suggests that as the percentage of ELL learners increases, writing proficiency decreases. This outcome is consistent with the relationship between Percent ELL and reading proficiency displayed in Table 4.

Table 5: Results of Regression Analysis on STARS Writing Data

Variable	Mean	SD	Standardized Beta	t-test value	*p* value	Partial Correlation
ELL %	.020	.052	-.173	-2.21	.02	-.14
SPED %	.154	.044	-.088	-1.27	.20	-.08
FRL	.346	.146	-.063	-0.76	.44	-.05
AR	.956	.018	.118	1.32	.18	.09
Graduation %	.901	.243	.154	2.35	.02	.16
Mobility %	.097	.073	.049	0.51	.60	.04
Accomd %	.051	.072	.035	0.49	.64	.03
ATS	35779	3330	.017	0.19	.86	.01
AYTE	16.44	2.71	.029	0.39	.70	.03
PTM	.967	.136	.017	0.19	.84	.01

Note: ELL % = English language learner %; SPED % = special education %; FRL = free and reduced lunch; AR = attendance rate; Accomd % = accommodation %; ATS = average teacher salary; AYTE = average years teaching experience; PTM = percentage of teachers with masters degree. Results based on 234 districts.

The results shown in Tables 4 and 5 indicate that variables have different effects on reading and writing proficiency. However, it is important to reiterate that the predictor variables explained little of the variation in proficiency levels. The relatively low amount of variation explained by district demographic characteristics suggests that data that is more closely tied to classroom practice and procedure may be necessary to better understand why some districts are outperforming others.

Question 4. We examined whether student and teacher variables were predictive of compliance or non-compliance with state and federal NCLB standards. A discriminant function analysis was conducted for students in the 4th, 8th, and 11th grades using the student and teacher variables described in Question 3. Two other variables were added, including number of students in each district and number of teachers in each district. We conducted separate analyses using state and federal compliance as outcomes. In the

93

current analysis, we report the results of compliance with *state* standards rather than *federal standards* for two reasons. First, approximately half the districts in the database do not report on federal compliance. Thus, while 234 districts reported compliance/no compliance at the state level, 109 did so at the federal level. A second reason is that state and federal analyses mirror one another, although variables tend to be better discriminators using state compliance due to a much larger sample size (i.e., statistical power). The results using the federal data are available upon request.

Table 6 presents mean scores for the met versus not met groups for the three targeted grade levels. Table 7 shows corresponding F- and statistical significance values for the differences between the met versus not met groups at 4th, 8th, and 11th grades. Three variables consistently differ between the met versus not met groups. The most robust difference is on the attendance rate variable. Districts with higher attendance rates are more likely to meet state compliance. This pattern also occurred when comparing federal compliance. The most plausible explanation is that higher attendance leads to more instruction and learning; thus, making it more likely that districts will perform well. A second difference occurs with the average years of teaching experience variable. Districts with more experienced teachers are more likely to comply with state standards. However, it is unclear whether teaching experience has a direct causal effect on compliance rates, or whether teachers with more experience gravitate over time to districts with higher student performance and compliance rates (e.g., Carey, 2004; Darling-Hammond, 2000). A third difference occurred with the free and reduced lunch (FRL) variable. Districts with lower FRL percentages were more likely to comply with state standards.

Table 6: Means for Met Versus Not Met Status on State Compliance

	4th Grade		8th Grade		11th Grade	
	Met	Not Met	Met	Not Met	Met	Not Met
ELL %	.01	.04	.02	.04	.02	.03
SPED %	.15	.16	.15	.16	.15	.16
FRL	.33	.36	.34	.42	.37	.42
AR	.96	.93	.96	.94	.96	.94
Graduation %	.87	.80	.89	.85	.96	.94
Mobility %	.10	.13	.09	.12	.09	.10
# Students	1172	558	1171	457	1132	444
ATS	35753	35394	35618	34790	35723	34456
AYTE	16.4	15.1	16.6	14.3	16.6	15.3
PTM	.30	.25	.29	.23	.30	.25
# Teachers	84.8	45.6	80.7	37.2	82.3	35.9

Note: ELL % = English language learner %; SPED % = special education %; FRL = free and reduced lunch; AR = attendance rate; ATS = average teacher salary; AYTE = average years teaching experience; PTM = percentage of teachers with masters degree. Results based on 234 4th grade districts, 256 8th grade districts, and 253 11th grade districts.

Table 7: F Values and Significance for Met Versus Not Met Status on State Compliance

	4th Grade		8th Grade		11th Grade	
	F	Sig.	F	Sig.	F	Sig.
ELL %	2.68	.10	5.40	.02	1.23	.27
SPED %	.41	.52	1.11	.29	2.41	.12

94

FRL	.56	.44	6.48	.01	6.61	.00
AR	20.61	.00	28.83	.00	20.25	.00
Graduation %	.84	.36	.62	.43	3.56	.06
Mobility %	2.72	.10	2.34	.13	.63	.43
# Students	.34	.56	.63	.43	.77	.38
ATS	.16	.69	1.21	.27	3.17	.08
AYTE	3.13	.08	14.51	.00	5.19	.02
PTM	1.82	.18	4.73	.03	2.62	.11
# Teachers	.31	.58	.58	.45	.73	.40

Note: ELL % = English language learner %; SPED % = special education %; FRL = free and reduced lunch; AR = attendance rate; ATS = average teacher salary; AYTE = average years teaching experience; PTM = percentage of teachers with masters degree. Results based on 234 4th grade districts, 256 8th grade districts, and 253 11th grade districts.

The correlation between the set of discriminating variables and compliance status was .36 for the 4th grade data, .44 for 8th grade data, and .37 for the 11th grade data. All three of these correlations were statistically significant at $p < .01$. These findings indicate that discriminating variables behave quite consistently across the three grades levels and explain approximately 10% of the variation in compliance status. Although the 10% variation explained criterion is statistically significant, it is considered low because 90% of the variation in compliance status is unexplained. Thus, districts either reach compliance or fail to reach compliance based on a variety of variables not included in the present analyses.

Conclusions

The present analyses lead to several general conclusions regarding relationships between STARS and NRT assessments and district compliance with state and federal standards. These are summarized below.

1. Scores from STARS and NRT assessments are correlated within and between the 2001 and 2003 cohorts. High correlations are evident in Table 2 for 4th and 8th grade, whereas the correlation drops at the 11th grade. The lower correlation may be due to differences in classes taken in high school or more variability in assessments.

2. Adequate yearly progress is difficult to achieve (Linn, 2003). The Nebraska NRT data comparing 2001 and 2003 cohorts do not reveal any changes in performance between 2001 and 2003. In contrast, STARS assessments reveal modest but statistically significant progress across the cohorts. One explanation is that STARS assessments are better aligned to district and state standards, and therefore provide a more reliable and valid assessment of educational progress compared to NRTs.

3. Few of the variables used in these analyses were good predictors of proficiency or state and federal compliance standards. Only three variables, attendance rates, years of teacher experience, and percentage of teachers with Masters' Degrees, provided significant prediction across analyses. The present findings suggest that increasing attendance rates may have an impact on percentage of students reaching proficiency and compliance with state and federal achievement standards. The NDE should consider ways to increase

attendance, and perhaps further investigate which districts are at risk for low attendance rates. Increased numbers of teachers with long-term experience and advance training also was associated with proficiency and compliance rates. Experienced and more highly educated teachers may be better prepared to plan, implement, and assess learning due to practical experience and additional graduate training. In addition, it should also be noted that Percent ELL appeared to have an important effect on 8th grade district writing and reading proficiency.

4. Most of the variables used in the analyses of questions 3 and 4 were poor predictors of proficiency and compliance. *This should not be interpreted negatively*. For example, Tables 4-7 reveal that number of students and teachers in a district (i.e., district size) is unrelated to proficiency and compliance. Similarly, Percent SPED and Mobility Rates did not predict proficiency or compliance. These findings provide valuable information about variables that are not related to compliance.

Recommendations

Several recommendations seem reasonable in light of these analyses. We summarize these below.

1. Continue using STARS as a planning and assessment strategy. STARS appears to be a more sensitive measure of yearly progress than NRTs.

2. Promote attendance. In the present analyses, attendance rates were the best predictor of proficiency and compliance by a wide margin, especially among at-risk districts.

3. Recruit and utilize experienced teachers with graduate training. One way to do so is to use experienced teachers as mentors within their school or as in-service facilitators. A second way is to use experienced teachers to serve at-risk students.

4. Consider collecting data on additional variables not included in the present analyses. Variables that assess the quantity and quality of instruction may be particularly important given the important role of attendance rates. Students who attend school more often are more likely to become proficient and meet state and federal standards. Currently, it is unclear what role the quantity (i.e., amount of time) and quality of instruction play.

References

Carey, K. (2004). The real value of teachers: Using new information about teacher effectiveness to close the achievement gap. *Thinking K-16, 8 (1)*, 3-32.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Educational Policy Analysis Archives, 8, no. 1*, retrieved

June 7, 2004 from http://epaa.asu.edu/epaa/v8n1/.

Linn, R. L. (2000). Assessments and accountability. *Educational Researcher, 29*, 4-16.

Linn, R. L. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher, 32*, 3-13.

Little, R.J.A. & Rubin, D.B. (1987). Statistical analysis with missing data. New York: Wiley.

Marion, S., White, C., Carlson, D., Erpenbach, W. J., Rabinowitz, S. & Sheinker, J. (2002). *Making valid and reliable decisions in determining adequate yearly progress*. Washington, DC: Council of Chief State School Officers.

No Child Left Behind Act of 2001, Pub. L. No. 107-110 (2002).

Schafer, J.L. (1997). Analysis of incomplete multivariate data. London: Chapman & Hall.

Written By
Elaine Specht

In the spring of 2004 this interview study attempted to take a snapshot of where principals are with the implementation of Standards, Assessments, and Accountability (SAA) in their elementary and secondary buildings across Nebraska. Five building administrators in districts of varying sizes, Class A through Class D, in five geographical areas were asked common questions concerning their role(s) and practices of leadership at this point in time. All five schools had received Very Good or Exemplary on their assessment process on the 2003 District Report Card.

The interview findings point to both celebrations and continuing challenges. The principals ranged in administrative experience from 25 to seven years.

- **The building principal's role in implementation of SAA**

To an opening question of who has led the implementation of SAA in the building, one principal spoke for all: "I guess that would be me." Four out of five interviewed principals led the work of curriculum development and alignment in their building in years preceding the matching of student learning objectives to the standards. Thus, they found SAA to be a next step.

The organizational pattern of who did what regarding STARS within the district was a common part of the response to the question of role. In one district the principals divided up the work. The interviewee was in charge of compiling the assessment data, putting it into a usable format for teachers to use and review via a secure district server, and handling Portfolio Criteria 5 and 6 for all levels of reporting. A principal in a small district although not in charge of the SAA was very aware of the STARS results and where the building faculty was in doing the testing online. Three of the five principals used technology as a major tool for organization and communication in the STARS process.

The building portfolio was a central part of the principal's responsibilities in two of the districts. In a Class C district, the administrator reviewed teachers' lesson plans and charted the "opportunities to learn" (Criterion 2) from every classroom in the building on a weekly basis. This principal also compiled the building portfolio by herself in order to save the teachers from the added responsibility although volunteer teachers did assist with proofreading the final portfolio document. In two other districts the principals were less concerned with the portfolio documentation. In a larger district the Curriculum Director compiled the district portfolio and in another district, teachers in the particular

curricular area with assistance from the school counselor were given time out of the classroom during the year and in the early summer to complete the portfolio. In the school with the on-line assessment system, the building principal was aware of the achievement results from his students but mainly depended on the teachers to make any needed changes in instruction or curriculum.

Other roles that these five principals listed relating to SAA included:
- staying informed about what is going on in their building
- attending trainings with staff and then giving teachers time to work
- finding grant funds to pay teachers for off-contract responsibilities relating to STARS
- looking at NRT scores and disaggregating them

From compiling and developing the building portfolio totally to the less-involved end of the spectrum, principals played a variety of roles in SAA. All principals reported some accountability for the process. A principal on the less active end summarized: "My role is to stay informed, to know what's going on so I can work with the teachers in our building. I work directly with (the other principal in the district) on things. We bounce ideas off each other on how we're going to do things. We provide to our teachers whatever they tell us they need."

- **The state of the school's progress at this point in implementing SAA**

The building administrators who had processes for curriculum building and learning teams in place reported that Standards, Assessments, and Accountability requirements are one more step in improving learning for all students. Although all these schools received high ratings in SAA via the building report card, there was a range of emotion regarding this work.

> "And really our teachers became the leaders on these (district) committees because they'd been there. We had tried to do it in our building…They knew about it. They understood."

> "(B)ut I think the teachers overall…I think they feel like, wow, we do really have a good program in place. They know it was pretty painful."

A high school principal reported that the assessments of the standards are put into the curriculum at the point of what they normally do in the classroom. In another district the principal stated that almost all the fourth and eighth criterion-referenced assessments (CRAs) for science were rewritten due to the results of Criterion 5 and the portfolio requirement documentation of breadth of sufficiency on the assessments. "I think the teachers had a lot better handle the second time we went through and revised them. It was much easier, and they had a better idea of what they needed to do."

- **The changes in role(s) and practices of building administrators in light of five years of SAA**

A principal spoke of change in leading from the front to "floating around." He interpreted his part as "(M)y role is to keep it (the SAA process) moving."

No one seemed to be content with the current status. A principal in a small district labeled SAA as "a work in progress." Although all five principals self-reported keeping a keen eye on the progress of SAA in their buildings, the practice of shared leadership seems to be catching on.

"(It) could be real easy to say no, this is what we are going to do, and you are going to do it, and that's it. Rather than the knowledge of knowing how to pull and get the information to those people and find ways to show them why it is important and why we need to do this together and how it's good for kids."

Another administrator stated that he was much more involved in curriculum (since SAA) and referenced Michael Fullan and the need to "socially process" the information together (with his faculty) for understanding.[12] Several of the administrators referred to educational books that were guiding their practice. David Perkins' The Smart School[13] provided a template for meetings in one of the districts and helped the faculty distribute leadership so that a different teacher leads each learning team meeting during the year. Before the session ends, plans are made for the next session. The book was read by the principal who then shared salient points with the faculty at a three-hour meeting early in the year to set the learning agenda.

All the principals mentioned their own learning curves in the SAA process. Although they were distressed regarding issues of time and feeling the "need to be on top of things," all stated that learning in their building is more focused for students.

One principal was reflective: "The first year or two we kind of wondered and wandered…We learned each year more things that we needed to do." Another stated that "standards have changed my job…Implementation of standards, assessments, and accountability has changed my role as building principal tremendously. I would like to say it's all positive. I can't quite go that far. There have been some positives that have come about. I feel very, very good about the program in place…so good about it that, that's why I haven't left here despite the challenges."

This same principal was at first identifying and tracking every standard in the lesson plans being taught in every grade, every classroom. "That was making me crazy…Now I have them (teachers) help me…And I hold them accountable to it. If I don't get it (lesson plans aligned to standards), I send a little email note…It's sometimes hard because they think I'm a pain in the butt…But if they don't know that I'm here for the kids, they don't know me very well."

Another practice of principals was the attention to SAA information and student achievement data and communicating it to the faculty. "We go over the data as far as the results...with the teachers…mostly individually…we've got one math teacher and we talk

---

[12] Michael Fullan states that if information is to become knowledge, a social process is required. Information stays information until people work through it together in solving problems and achieving goals. Dennis Sparks, "Interview with Michael Fullan*: Change Agent," Journal of Staff Development*, vol 24, no.1.

[13] David Perkins, *Smart Schools: From Training Memories to Educating Minds*. (New York: The Free Press, 1992).

about what's there. This same principal said: "Now we look at those test results first." At faculty meetings principals are using the opportunity to "explain the accountability issues to them…I've tried to explain nationally what's happening, explain NCLB, talk about STARS. I try to give them pieces of information as we go along so it isn't just like we are dumping all of this on (them)."

Professional development is another of the duties of building principals. "I've really tried to beef up our in-services to meet whatever our standards need or our curriculum needs at the time." SAA is changing how principals are looking at professional development. A principal in a Class A district stated that a well-known educator on reading was brought in who was "very good, (but there was) no follow-up." A need to connect to what the faculty has learned in prior district or building professional development and then bridge the gap to the next level was an area identified by this principal as very important. Principals in three districts stated that attending workshops and trainings with teachers helped to bridge that gap.

Another is setting levels of responsibility. "Every year I have all of my teachers report at the end of the year on student performance levels on all of the standards, K through 6…Then we have at least one meeting a year, maybe two, where we talk about what were the weak areas, what were our strengths. (W)e determine a plan to meet our areas of weakness." Principals are in a change process for involving staff in SAA. "As we moved through the process, we involved the staff more. (I)t seemed easier to work with the staff and say, OK, here's what we have to do and just go from there."

A principal in a district in southeast Nebraska declared that SAA had restructured the whole School Improvement process and that next year there would be early dismissal on a weekly basis to work on improving learning for students which will embed SAA into the district's school improvement process.

The challenges have not lessened or disappeared. Principals are actively dealing with them. "(The) fourth grade teachers…were wanting to jump ship a couple of years ago because they felt overwhelmed. Their third grade partners, as we call them, said, 'Hey, we can cover that for you,' and they have very nicely assumed many of them (standards)."

Personally, some of the building principals admitted to moments of doubt: "I've been reading a book called _____ and on days when I'm really struggling with figuring out what in the world am I doing here, and why am I still in administration, I read that book, and it helps pull me back." The work load has increased "10 times as much as it was before. And a lot of it is because of the portfolio." When another principal walked into a fourth grade classroom this year with a box of math tests, "the kids looked at them, and they go, 'Ooooh!' How much it's impacting them. They were like 'Another test?' My gosh, these kids just dreading, you know, fourth grade."

Not all the school principals reported K-12 alignment of student learning objectives or being part of a strong administrative team. "As tactfully as I can…if I had my druthers, I would have a K-12 very articulated program. I can only take control of what I'm supposed to take control of…sometimes I feel I'm on a little island down here. And when you're not on the same page administration-wise in the building, some people are saying on one end of the building that you don't have to do this, it sometimes creates a rift. Then they wonder if ___(is) putting me through something I don't need to go

through…I asked the ___ principal yesterday how you doing on the portfolio? And he said, 'I hope my teachers are working on it.' This is May."

- **Support for implementation of SAA in the building**

The principals in all the interviews verified that assistance had come from the Nebraska Department of Education as well as from their area Educational Service Unit "Without __ and __ (names of ESU professional developers) and the people at the State Department…I do call area principals, too, but everybody's doing something a little bit different." Another principal felt that NDE was becoming more helpful. "The biggest thing that ever happened was that the department got going in my opinion."

Another support was identified as the structures already in place in the building that help teachers to work together and improve practices. "We just folded standards into that." Another principal built the support: "I don't want to call it forced teamwork…Everybody knows they're going to hear it from me every single time. We're in this together, not just those fourth grade teachers out there paddling by themselves. So…more of a team atmosphere where we're all in this, we are all going to be working toward this (identified weak area)…not just the fourth grade teachers." Another principal put it differently: "(We) try to do as much as we can as a group so everyone has input, and everyone understands why we are doing it and the importance of it…that we are not doing to them but with each other or for each other."

Other areas of support that principals included were having professional development in their building and the more rigorous requirements from higher education.

"(I) bring people in to in-service them and give them fresh ideas on it (area of low performance)."

"(W)hen Lincoln (UNL) changed their requirements to four year of math really helped."

- **Obstacles to implementation of SAA in the building**

Time was targeted as a major need by all interviewees. "(I) still have to run the building." Another principal said that trying to run the building and trying to be the instructional leader and tie in the activities takes time away from collecting data or working with the faculty. One principal went further and identified the traditional school year as a major time constraint.

There are still faculty who are fighting Standards, Assessments, and Accountability processes and products. "(R)eluctant people don't bother me as much as people who try to find ways why it's not working…(T)hose are more difficult to handle…(P)eople that have a tendency to not want to take responsibility." Although several principals cited support from the superintendent, the knowledge level of the superintendent regarding the time commitment and work of STARS was limited.

The unease regarding what will schools and their leadership be required to do next continues. "The other big challenge for me would be the things that are always coming down the road. I mean, just like, you know, we are really doing as much as we

could and really starting to move into standards and assessments and have that knowledge at least fairly well where we thought we knew what was going on, and NCLB came and those those types of things…almost stop and say, are we gonna be able to continue doing it that way? What do we do until we find out where we go?"

- **Referencing SAA, identification of the most important skills or abilities for building administrators**

According to the interviewed principals, building administrators need to be strong in several areas:"(You have to be) willing to be the instructional leader in the building, and there's no way around that." Organizational skills and setting priorities were skills that were mentioned by several principals.

Like a melody, all the principals at one point or another in the interview connected their SAA roles and practices to "What's good for kids." Compassion and trust were also addressed. "(A) large heart—not a hardened heart, a large heart. You have to love kids." The skill of the administrator in establishing a level of expectations for children is important: "(M)aybe it's the trust they have in you as an administrator and leader of children. That they trust that you are not going to lead them on a path that is not going to be best for students."

To this question, a principal of seven years responded by saying that the skills or abilities for SAA that a building administrator needs are no different than the ones that make a good principal: knowledgeable, able to communicate, create a team, delegate and give power and resources for what needs to be done.

Another principal put it differently, "As far as standards (and) assessments, I don't know 'cause we are still learning in that area, too. It's not something where you know even though I've been an administrator for…18 years now. Whenever I get to the point where I think I know it all, I better hit my head against the wall and start over again because there is always something new coming up…I love it. I mean, it's 18 years, and I love it."

Recommendations for the Nebraska Department of Education:

- continue the frequent and timely communication with districts via teleconferences, emails, mailings, and direct conversation
- assist schools (and building leadership) in the application of systems thinking to establish robust connections to School Improvement
- find ways to work directly with more building principals to energize and sustain their leadership of SAA, embed defensible practices, and thus impact student learning
- continue to forge strong, professional relationships with building principals as a means to enhance their own professional development