

# CHARTING STARS

## *Engaging Conversations*

Written by  
**Jody C. Isernhagen, Ed.D.**  
Principal Investigator, STARS Comprehensive Evaluation  
Associate Professor of Educational Administration  
132 Teachers College Hall  
University of Nebraska-Lincoln  
Lincoln, NE 68588-0360  
402.472.1088  
jisernhagen3@unl.edu

**Shirley J. Mills, Ph.D.**  
Secondary Investigator  
Department of Educational Leadership  
Educational Complex 1.606  
University of Texas-Pan American  
1201 West University Drive  
Edinburg, TX 78539-2999  
smills002@rgv.rr.com

Study Contracted by the Nebraska Department of Education  
Also supported by  
College of Education and Human Sciences,  
University of Nebraska-Lincoln

Submitted to the Nebraska Board of Education  
October 2008



<p><b>Table of Contents</b></p> <p><b>Charting STARS: <i>Engaging Conversations</i></b></p>
---

**Section 1: Introduction ..... 7**

**Section 2: Executive Summary ..... 11**

**Section 3: Research Studies**

**Study I: Nebraska-led Math Portfolio Peer Review Process .....41**

**Theme 1: *Alignment*..... 44**

**Theme 2: *Sufficiency* ..... 47**

**Theme 3: *Clarity* ..... 48**

**Theme 4: *Appropriateness* ..... 50**

**Theme 5: *Scoring Procedures* ..... 53**

**Theme 6: *Summarizing the Review Process* ..... 55**

**Theme 7: *The Pilot Integrated Visit Review Process*.....59**

**Theme 8: *New Learnings*..... 60**

**Study II: STARS Enhancement: The Impact of Revisions to the Quality  
Accountability Act..... 65**

**Study III: 2001-2007 Reading and Math Achievement ..... 73**

**Study IV: 2002-2007 Writing Achievement..... 81**

**Study V: 2001-2007 Achievement for Special Populations ..... 86**

**Study VI: Student Mobility Effects on Achievement across Levels of Poverty ..... 95**

**Study VII: The Effect of Nebraska’s Standards and Accountability System  
(STARS) on School Improvement Practices ..... 111**

**Study VIII: Nebraska Educators Review the Local Math Assessment Process:  
Reliability of Peer Review of Assessment Portfolio 2007-08..... 117**

**Section 4: Appendices:**

**A. IRB Approval Letter..... 135**

**B. Researchers for the Project..... 139**

**C. 2007-08 Study I: Nebraska-Led Math Portfolio Peer Review Survey ..... 143**

**D. 2007-08 Study I: Nebraska-Led Math Portfolio Peer Review**

<b>Interview Protocol.....</b>	<b>147</b>
<b>E. 2007-08 Study II: STARS Enhancement Research Survey .....</b>	<b>155</b>
<b>F. 2007-08 Study VII: The Effect of STARS on School Improvement Practices Survey .....</b>	<b>163</b>
<b>G. 2007-08 Study VIII: Review of the Local Math Assessment Process- District Assessment Portfolio Worksheets.....</b>	<b>167</b>
<b>H. 2007-08 Study VIII: 2007-08 Peer Reviewer Training Evaluation Survey .....</b>	<b>175</b>

## TABLES & FIGURES

### Tables

1	Percent Proficient or Higher (Change) on Criterion-Referenced Tests in Reading .....	76
2	Percent Proficient or Higher (Change) on Norm-Referenced Tests in Reading.....	76
3	Percent Proficient or Higher (Change) on Criterion-Referenced Tests in Math .....	77
4	Percent Proficient or Higher (Change) on Norm-Referenced Tests in Math.....	77
5	Reading and Math District Average Portfolio Ratings (Gain/Loss) 2001-2007.....	78
6	Math District Average Portfolio Ratings (Gain/Loss) 2001-2007.....	79
7	Statewide Writing Assessment 2002-2007: Mean District Percent of Student Scores at the Proficient Level or Higher .....	84
8	District Average Reading Criterion-Referenced Assessments English Language Learners (ELL) 2001-2007.....	88
9	District Average Math Criterion-Referenced Assessments English Language Learners (ELL) 2002-2007.....	89
10	District Average Writing Criterion-Referenced Assessments English Language Learners (ELL) 2002-2007 .....	90
11	District Average Reading Criterion-Referenced Assessments Special Education (SPED) Students 2001-2007 .....	91
12	District Average Math Criterion-Referenced Assessments Special Education (SPED) Students 2002-2007 .....	91
13	District Average Writing Criterion-Referenced Assessments Special Education (SPED) Students 2002-2007.....	92
14	Quality Criterion 1 – The Assessments Match the Standards.....	122
15	Quality Criterion 2 – Students Must Have the Opportunity to Learn.....	123
16	Quality Criterion 3 – Assessments Are Free of Bias and Sensitive Situations.....	124
17	Quality Criterion 4 – Assessments Are at an Appropriate Level.....	125
18	Quality Criterion 5 – There is Consistency in Scoring.....	126
19	Quality Criterion 6 – Mastery Levels Are Appropriate .....	127

### Figures

1	Nebraska-led Portfolio Peer Review Process Survey Category Average Scores 2007-08.....	13/43
2	STARS Enhancement Survey Respondents Levels of Experience.....	67
3	Criterion 1 Reviewer Agreement by Indicator .....	127
4	Criterion 2 Reviewer Agreement by Indicator .....	128
5	Criterion 3 Reviewer Agreement by Indicator .....	128
6	Criterion 4 Reviewer Agreement by Indicator .....	129
7	Criterion 5 Reviewer Agreement by Indicator .....	129
8	Criterion 6 Reviewer Agreement by Indicator .....	130



# **CHARTING STARS**

## ***Engaging Conversations***

### **Section 1: Introduction**



The seventh annual report of the STARS Comprehensive Evaluation Project (CEP) is an independent evaluation of Nebraska's School-based Teacher-led Assessment and Reporting System (STARS). This study was approved by the Institutional Review Board (IRB) at the University of Nebraska-Lincoln (UNL) characterized by the highest level of integrity, with respect and equitable treatment for all persons involved in the study in order to maintain confidentiality and protect the privacy of participants in the study (Appendix A). The CEP was originally contracted between the Nebraska Department of Education (NDE) and the University of Nebraska-Lincoln, College of Education and Human Sciences (CEHS) in 2001. The CEP was supported jointly by the NDE and the College of Education and Human Sciences (CEHS).

Dr. Jody Isernhagen, Associate Professor, served as the Principal Investigator and Dr. Shirley Mills, Assistant Professor, University of Texas-Pan American, served as secondary investigator. Dr. Jerald Riibe, Assistant Superintendent, Ralston Public Schools authored Study VII and John Moon, Assessment Coordinator, Nebraska Department of Education was the author of Study VIII. Graduate students serving as authors in the project were: Jackie Florendo, Casey Tallent and Nino Zhvania. All researchers and members of the research team for the Comprehensive Evaluation Project are listed in Appendix B.

### **OVERVIEW**

Nebraska educators have engaged in stimulating conversations due to the implementation of STARS over the past seven years. Nebraska's approach to standards, assessment, and accountability, better known as STARS: School-based Teacher-led Assessment and Reporting System, is grounded in the belief that decisions about student learning should be standards-based and should be based upon classroom knowledge of the student. Each year

educators have continued to perfect the implementation of the STARS Process to better meet the needs of Nebraska students. This process is designed to help educators build new connections between assessment, curriculum, and instruction. Nebraska educators have become assessment literate and data users. Their focus is now upon perfecting their instructional skills and implementing appropriate interventions to best meet the needs of Nebraska students for whom they serve.

The STARS Process has been constantly in transition since its inception. The 2007-08 school year is no different. In the spring of 2008, the Nebraska legislature passed new legislation requiring a statewide test in Reading, Math, and Science. In 2008-09, the reading statewide test will be piloted and will be fully implemented in the 2009-10 for all Nebraska school districts. In 2009-10, the math statewide test will be piloted and fully implemented in 2010-11. Nebraska is again charting new waters with revisiting its standards and developing new statewide tests in reading, math, and science. School districts are now engaged in conversations about how to monitor student learning using a balanced assessment system.

### **SUMMARY OF THE SEVENTH-YEAR STUDY**

Over the past seven years, the Nebraska School-based Teacher-led Assessment and Reporting System (STARS) has required Nebraska's school districts to develop a local assessment system to measure student performance on local standards that were equal to or exceeded the state standards. Eight studies were conducted during the seventh year of the Comprehensive Evaluation Project.

#### ***STUDY I: Nebraska-led Portfolio Peer Review Process***

In 2006-07, a new Nebraska-led Portfolio Peer Review Process was introduced with great success. This new process provided many opportunities for new learning. The Portfolio Review Process study investigated educator's perceptions of the technical quality of their district assessments according to the requirements of the six quality assessment criteria. In 2006-07, the focus for this study was on reading. In the current year's study (Study I), the focus was on math and the researchers investigated educator perceptions of the technical quality of math classroom assessments according to the requirements of the six quality criteria (Plake & Impara, 2000):

- The assessments reflect the state/local standards.
- Students have the opportunity to learn.
- The assessments are free of bias and insensitive situations.
- The assessments are at the appropriate level.
- The assessments are reliably scored.
- The assessment mastery levels are appropriately set.

Quantitative survey data was collected before the external review for each district.

Qualitative interview data was collected after the actual Portfolio Peer Review in selected districts.

The Portfolio Peer Review consisted of state trained assessment experts that visited each school district within the state. The visiting team read the previous district portfolio and the recommended changes for improvement from past years prior to visiting the district. The visiting teams then went to their assigned districts; reviewed assessment evidence based on

the six quality indicators, and provided formative feedback to the districts. A district summative rating will appear on the State of the Schools Report in 2008.

***STUDY II: STARS Enhancement: The Impact of Revisions to the Quality Accountability Act***

This seventh-year study (Study II) was an initial exploration of the impact of the revisions to the Quality Accountability Act enacted by Nebraska Legislative Bills #653 and #1157. These bills initiated a revision of the state content standards and, following the completion of this study, the development of state tests. The purpose of the study was to explore the impact of the changes to the Quality Accountability Act and to examine the participants perceptions about the changes to STARS based on the enhancement criteria. After LB #653 was passed, a newly formed STARS Enhancement Design Team began work to enhance STARS through a revision of content standards. A survey was conducted to measure the level to which the team participants agreed or disagreed with the enhancement criteria.

***STUDY III: 2001-2007 Reading and Math Achievement***

The third study (Study III) was a longitudinal achievement study conducted in 2007-08 for the STARS Comprehensive Evaluation Project (CEP). District achievement scores for reading and math were compared on criterion-referenced and norm-referenced measures from 2001 through 2007. District portfolio ratings for reading and math from 2001 through 2007 were also compared.

***STUDY IV: 2002-2007 Writing Achievement***

The fourth study (Study IV) was a longitudinal study that focused on writing achievement. District writing scores on the Nebraska Statewide Writing Assessment (NSWA) from 2002 through 2007 were compared.

***STUDY V: 2001-2007 Achievement for Special Populations***

The fifth study (Study V) was a longitudinal study of district reading, math, and writing using criterion-referenced scores from 2001-2007 for English Language Learners (ELL) and Special Education (SPED) students.

***STUDY VI: Student Mobility Effects on Achievement across Levels of Poverty***

The sixth study (Study VI) was conducted to determine the influence of student mobility on the overall performance of districts. Specifically, this study examines how different levels of student mobility affect the performance of districts with varying levels of poverty.

***STUDY VII: The Effect of Nebraska's Standards and Accountability System (STARS) on School Improvement Practices***

The purpose of the seventh study (Study VII) was to compare the perceptions of Nebraska second-grade teachers (a non-reporting grade) with perceptions of Nebraska fourth-grade teachers (a reporting grade) regarding the effect of Nebraska's standards accountability system (STARS) on school improvement practices.

***STUDY VIII: Nebraska Educators Review the Local Math Assessment Process:  
Reliability of Peer Review of Assessment Portfolios***

The eighth study (Study VIII) was a study undertaken by the NDE to evaluate the Peer Review Process by establishing the reliability of reviewer decisions in collecting information from district assessment portfolios.

**YEAR SEVEN COMPREHENSIVE EVALUATION FORMAT**

This comprehensive report has been designed to serve multiple audiences and provide the most pertinent information available on the implementation of STARS based on the data collected during the 2007-08 school year. All of the studies connected assessment, curriculum, and instruction that supported increased student achievement.

This report is divided into four sections beginning with an introduction of the total report (Section 1); an executive summary of the findings of all studies conducted during the seventh-year study (Section 2); complete research papers of the eight major studies (Studies I-VIII) conducted during the 2007-08 school year (Section 3); and the Appendices (Section 4).

**ACKNOWLEDGMENTS**

We offer special thanks to the many educators in school districts across Nebraska for sharing their knowledge and skills to create a better learning environment for students. Their expertise and enthusiasm for tackling the complexities and connections between assessment, curriculum, and instruction helps to create opportunities and ensure learning for all Nebraska students.

A special thanks to: Doug Christensen, former Commissioner of Nebraska Department of Education; Pat Roschewski, Director of Statewide Assessment; Jan Hoegh, Assistant Director of Statewide Assessment; Bob Beecham, Administrator of Education Support Services; Jackie Naber, Office Administrator of Statewide Assessment, and Carol Bom, Office Assistant. All have offered great assistance for the completion of the Comprehensive Evaluation.

We offer our sincere appreciation to Shirley Mills for the many years that she served as a graduate student and currently as secondary investigator for this project. We couldn't have done it without you. A special thanks to graduate students Jackie Florendo, Casey Tallent, and Nino Zhvania for their long hours and dedication to this project. Without their support, this report surely would not have been completed.

We offer our thanks to Jerry Riibe and John Moon who have enriched the comprehensive evaluation with their studies and findings.

A special thanks to our Administrative Assistant, Susan Wilson, as she provided great attention to detail and completion of the final product. To Cindy DeRyke, Diane Gronewald, Shelia Hayes, and Tammie Herrington, thank you for your support throughout the year. A special thanks to Marjorie Kostelnik, Dean of the College of Education and Human Sciences; L. James Walters, Associate Dean; and Larry Dlugosh, Chair of the Department of Educational Administration, for their continued support for the Comprehensive Evaluation Project.

**REFERENCES**

Plake, B. S., & Impara, J. C. (2000). *Technical criteria for Nebraska school district assessment*. Lincoln, NE: Buros Center for Testing.

# **CHARTING STARS: *Engaging Conversations***

## **Section 2: Executive Summary**



### **INTRODUCTION**

In 2001 the United States reauthorized the Elementary and Secondary Education Act, the law that addresses national PreK-12 education involvement, to include what was referred to as a “common sense” pillar of accountability for results. The Act became known as “No Child Left Behind” (NCLB, 2002) and has driven education policy since that time. The Act required states to develop assessments to provide evidence of accountability. While education is a “states’ rights” issue in the United States, the federal government influences local education policy by tying federal funds to compliance with federal initiatives. These revenues average about 7% to 10% of local school district budgets nationwide.

“Today’s schools are less focused on merely sorting students and more focused on helping all students succeed in meeting standards,” and schools need to move from total reliance on “assessments *of* learning” to the use of “assessments *for* learning” (Stiggins, 2007, p. 22). Madaus (1988) has suggested the use of criterion-referenced tests (CRTs) as an alternative testing system to support “assessment *for* learning” as advocated by Stiggins (2007). While norm-referenced tests (NRTs) ascertain the rank of students, CRTs determine “what test takers can do and what they know” (Bond, 1996, p. 2). Nebraska school districts use a balance of assessments including both NRTs and CRTs, from pencil-and-paper criterion-referenced tests to performance assessments, as suits the individual district (Gallagher, 2007). This system is entitled STARS: School-based Teacher-led Assessment and Reporting System. The requirements of the federal No Child Left Behind Act (NCLB) have been integrated into the accountability requirements of Nebraska’s system.

Since the implementation of Nebraska’s unique assessment system seven years ago, the state’s educators have been compelled to stretch and grow in their understanding of the relationship between assessment, curriculum, and instruction. This growth has led to a common language about assessment related to the six quality criteria developed by the Buros Center during the initial stages of assessment implementation (Plake & Impara, 2000).

Building on the statewide district portfolio project used to measure the success of the assessment process, school districts have showcased their assessments, curriculum, and instruction to teams of experts in Nebraska and across the nation. As assessment discussions and conversations continued across the state in 2007-08, Nebraska educators continued to commit themselves to an improved assessment process. In 2009-10 Nebraska will implement a state test. The use of both classroom-based assessments and statewide tests will provide data for policy makers and classroom teachers.

## **SEVENTH YEAR RESEARCH STUDIES**

There were eight major studies conducted during the seventh year of the STARS Comprehensive Evaluation Project (CEP). These eight studies are summarized in this section of the report and presented as complete reports in Section III.

### **STUDY I: Nebraska-led Math Portfolio Peer Review Process**

**Jody Isernhagen, Ed.D, Associate Professor, University of Nebraska-Lincoln**  
**Jackie Florendo, M.Ed., Graduate Assistant, University of Nebraska-Lincoln**  
**Casey Tallent, M.A., Graduate Assistant, University of Nebraska-Lincoln**

#### **Introduction**

The first study (Study I) was an investigation of educator perceptions of the technical quality of their district math assessments according to the requirements of the six quality assessment criteria (Plake & Impara, 2000):

- The assessments reflect the state/local standards.
- Students have the opportunity to learn.
- The assessments are free of bias and insensitive situations.
- The assessments are at the appropriate level.
- The assessments are reliably scored.
- The assessment mastery levels are appropriately set.

The portfolio review consisted of teams of two state-trained assessment experts that visited each school district within the state. The visiting team went to their assigned districts and participated in conversations about the evidence of assessment quality based on the six assessment quality indicators, and provided formative feedback to the district. Two outside assessment experts located in each of the regional areas assisted the peer review teams when there were questions. The assessment reviewers read the district math portfolio and recommended changes for improvement.

#### **Methodology**

This seventh-year mixed-methods research study focused on the Nebraska-led Assessment Math Portfolio Peer Review Process. The District Internal Portfolio Review Team members were surveyed prior to the Portfolio External Review. For the purpose of this research, Nebraska school districts were divided into two classifications, non-rural and rural, based on population characteristics unique to Nebraska. Non-rural districts were defined as metro-area districts in large and mid-sized cities, large towns, and the urban fringe. All other districts are classified as rural. Of the 254 Nebraska public school districts in 2007-08, 4.3% were non-rural and 95.7% were rural. Each school district was asked to distribute up to ten surveys to internal review team members. Of the 254 Nebraska school districts in 2008, 750 educators from 189 districts (74.4% of districts) returned surveys for the Comprehensive Evaluation Research Study. Fifty-one surveys were disallowed because they were returned after the completion of the District Portfolio

Review including all surveys from one district reducing the number of participating districts to 188 total districts. Of the 188 districts responding to the survey within the prescribed timeframe, 4% were non-rural and 96% were rural.

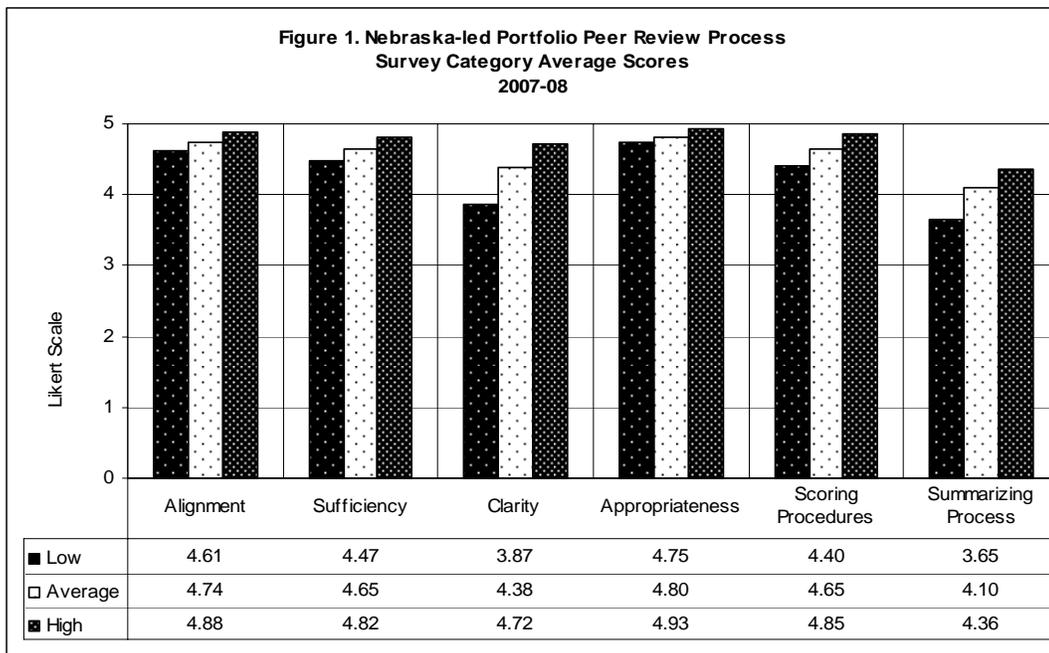
Participants responded to the 49-item survey using a five-point Likert scale for each item, with “1” representing “none of the time,” “2” “very little of the time,” “3” “some of the time,” “4” “most of the time,” “5” “all of the time.” The survey was structured to explore six themes: Alignment, Sufficiency, Clarity, Appropriateness, Scoring Procedures, and Summarizing the Review Process (Appendix C).

In the second phase of this study, open-ended interviews were conducted in two districts from each of four geographical areas identified by the NDE for training of portfolio reviewers. Detailed views were collected about the Nebraska-led Portfolio Peer Review Process in the sample districts. An interview protocol for educators (Appendix D) was utilized to gather qualitative data. The six themes based on the survey were evident in the interviews. They were: Alignment, Sufficiency, Clarity, Appropriateness, Scoring Procedures, and Summarizing the Review Process. Two additional themes emerged from the interviews: The Pilot Integrated Visit Review Process and New Learnings.

### Findings

Educators’ perceptions of the Nebraska-led Math Portfolio Review Process were explored using a survey based on six categories. This survey was administered to members of the school based review team prior to the visit to school districts by the portfolio experts. These items were rated by participants on a “1” to “5” Likert scale with “5” being the highest. Noted in Figure 1 are the mean, the lowest, and the highest average scores per category. The highest average score (4.93) was in “Appropriateness” while “Summarizing the Process” was the lowest average score (3.65).

Six themes based on the six quality criteria were identified within the interviews. Two additional themes emerged during the interview. These eight themes are addressed in the summary of Study I.



### *Alignment*

In the area of Alignment, staff shared how they were involved in aligning standards by stating, “We just took the standards and then at each grade level we went through what we taught and . . . it definitely evolved. We started with a big document where we listed all of our objectives by grade level under the standard that we felt that they met.” Alignment responses ranged from 4.61 to 4.88 with an average of 4.74 on the Likert scale. The strongest item rated by all respondents within the Alignment category was “our district involved staff in the alignment of the assessments to the standards” (4.88). Respondents also indicated that “there is a documentation process for alignment of assessments to standards” (4.80) and “districts’ support teachers working collaboratively to ensure assessments measure the standards” (4.79). The lowest rated area for alignment was “our district had assessment items reviewed by external personnel” (4.61).

During the interviews, participants were asked if they had recommendations to make to other school districts regarding Alignment. Recommendations included the need for standards to be aligned with a strong curriculum, involvement of teachers across grade levels as well as the involvement of English Language Learner (ELL) and Special Education (SPED) teachers. Teachers shared that it must be an ongoing process to make sure that the curriculum continues to reflect the standards.

### *Sufficiency*

Sufficiency continues to be one of the most complex criteria for schools and districts to understand. A teacher shared that they initially wrote questions and then identified the levels, “We just wrote questions and we didn’t really think about the levels and then the next year we’re like, oh . . . we need to have levels so we know where the kids are going to fall. Are they truly a beginning student or a progressing, or proficient, or advanced?” However over time teachers recognized the importance of writing questions based on performance level descriptors. Sufficiency responses ranged from 4.47 to 4.82 with an average of 4.65 on the Likert scale. The strongest item rated by all respondents within the Sufficiency category was “our district reviewed assessment items/tasks for sufficiency results” (4.82). Respondents also indicated that “our assessment items/tasks are distributed across all performance levels” (4.75). The lowest rated area for sufficiency was “our assessment items/tasks use a variety of appropriate formats” (4.47).

During the interviews, participants were asked if they had recommendations to make to other school districts regarding the area of Sufficiency. Recommendations included the importance of involving teachers in the process at the school and having strong performance level descriptors for developing questions.

### *Clarity*

Clarity was the second lowest rated item on the survey and was the lowest rated item in 2006-07. A female rural middle school math teacher shared the growth they made as a school in the area of clarity by stating, “The assessment coordinators proctored all assessments. We definitely have standardized the process because it’s the same person giving them (assessments) all the time in each building. The proctors give the same directions every time to every kid for every subject.” Clarity responses ranged from 3.87 to 4.72 with an average of 4.38 on the Likert scale. The strongest item rated by all respondents within the Clarity category was “our assessment directions for teachers are standardized across the district” (4.72). Respondents also indicated that “our assessment directions for students are standardized across the district” (4.69), and “our assessment directions for students are clear” (4.68). Additionally, respondents indicated that “assessment directions for teachers are clear” (4.68). The lowest rated area was “our district provides parents with reports that give an appropriate explanation of assessments results” (3.87).

Additionally, respondents also rated this item low, “our district sends individual reports each school year” (3.93).

During the interviews, participants were asked if they had recommendations to make to other school districts regarding the area of Clarity. Recommendations included the advantages of immediate feedback to students when using online assessments, using assessment feedback for motivating students to master the standards, and the importance of an established standardized procedure for administering assessments.

### *Appropriateness*

Appropriateness continues to be the highest rated area by Nebraska educators. A male rural superintendent shared, “There’s been a conversation and a review between the different grade levels to see that they’re building upon one another. That’s happened both in house and it’s been also somewhat orchestrated by our ESU, too, in terms of different schools coming together at different grade levels.” Responses on Appropriateness ranged from 4.75 to 4.93 with an average of 4.80 on the Likert scale. The strongest item rated by all respondents within the Appropriateness category was “our assessments were screened for fairness, bias, and sensitivity” (4.93). Respondents also indicated that “our assessments were reviewed by internal or external groups” (4.84). The lowest rated areas were “our assessments are appropriate for the assessed grade level” (4.75) and “our assessments indicate our expectations for our students” (4.75).

During the interviews, participants were asked if they had recommendations to make to other school districts regarding Appropriateness. Recommendations included having conversations with grade level teams that include SPED and ELL teachers, understanding local biases and participating in bias training regularly, and involving students and parents in conversations about the assessment process.

### *Scoring Procedures*

In the area of Scoring Procedures, responses ranged from 4.40 to 4.85 with an average of 4.65 on the Likert scale. The strongest item rated by all respondents within the Scoring Procedures category was “our participation rates are documented” (4.85). Respondents also indicated that “our assessments have established scoring guidelines and directions” (4.79) and “our district has local assessment policies in place to assure comparability and consistency across the district” (4.76). An assessment coordinator shared how teachers were able to make wise decisions by using a consistent scoring method, “We use the Angoff method. That gives us our level of difficulty. You can take that information and use it to really look at your cuts and make some very strong decisions about whether they’re appropriate. Then as far as reliability, we use KR 20 and we run those ourselves in the summer. It makes them (teachers) see that maybe their assessments are too easy or if their assessments are too hard.” The lowest rated areas were “our students are given instruction about behavioral objectives during the assessments” (4.40) and “our district provides training for those administering the assessments” (4.48).

During the interviews, participants were asked if they had recommendations to make to other school districts regarding Scoring Procedures. Recommendations included: the need for re-evaluation of scoring procedures every year; using data teams and encouraging double scoring opportunities; participating in training on the consistency methods; and using results for improving student performance.

### ***Summarizing the Review Process***

During the interviews, many leaders and teachers summarized the review process by indicating that they believe the STARS process is strong and that it makes people really think about the importance of assessment and accountability as shared by this high school math teacher, “I think it’s a good process. I can see the value in it. You are a different person at different stages and the process you go through helps you value it even more. It is a lot of work and I think it’s a process that makes people really think about what they’re doing now more than ever before. The accountability piece I like, if you felt there might not have been people before who were aware of what was going on, they pretty much need to be aware now.”

In the area of Summarizing the Review Process for all respondents, responses ranged from 3.65 to 4.36 with an average of 4.10 on the Likert scale. The strongest item rated was “I have the necessary information to prepare the district assessment portfolio” (4.36). Respondents also indicated that “I feel prepared to present my district portfolio to my peer reviewers” (4.32) and “I have had adequate help in preparing the district assessment portfolio” (4.26). The lowest rated areas were “Compensation is provided to prepare the district assessment portfolio when completed outside of the regular school day” (3.65) and “I was provided time within the teaching day to prepare the district assessment portfolio” (3.85).

During the interviews, participants were asked if they had recommendations to make to other school districts regarding Summarizing the Review Process. These recommendations included using teachers as peer reviewers; building an assessment team that keeps teachers in the district informed; providing opportunities for teachers to participate in professional learning communities; and use of data teams as a means to increase communication about the process.

### ***The Pilot Integrated Visit Review Process***

In 2007-08, The Nebraska Department of Education, in an effort to reduce the number of visits a school received in a calendar year, piloted an Integrated Visit Review Process in a limited number of school districts. The integrated visit process conducted multiple reviews in a single visit. For example, some schools connected the portfolio review process with the school improvement process in a single visit while other districts connected Title I with the Portfolio Review Process in a single visit. Initial feedback from pilot schools was positive and districts elaborated on the success of the integrated visit during the Portfolio Math Review Process. A high school math teacher shared that their integrated visit was a positive experience that makes you grow as an individual, “I would highly recommend trying to have more integrated visits like we did. I feel that you could see the bigger picture when you tie it all together. It makes you grow, as an individual and as a district.”

### ***New Learnings***

As leaders and teachers reflected upon the portfolio review process, there were conversations about several new learnings. These conversations focused on the review process being noticeably more manageable the second time around and they were much more confident with the portfolio process itself. Teachers also felt that standards assured them of what was required for a child at a given grade level and that their students were receiving the same quality of education as other students across the state. Learning has improved for both teachers and students. “I feel both processes have made teachers become better teachers. They are much more aware of what students need at all levels,” as shared by a high school math teacher. Other conversations showed that teachers have learned new ways to teach and new methods to use to teach students at various learning levels. “It just made us all become better teachers . . . and it holds you accountable,” as shared by a rural high school math teacher. The process has increased

stakeholder accountability as shared by a rural assessment coordinator, “We kind of want to go in our room and just do our thing . . . But we do need to be accountable to our patrons, to our students, to the parents.”

## **STUDY II: The Impact of Revisions to the Quality Accountability Act**

**Jody Isernhagen, Ed.D., Associate Professor, University of Nebraska-Lincoln**  
**Casey Tallent, M.A., Graduate Assistant, University of Nebraska-Lincoln**

### **Introduction**

The seventh-year study was an initial exploration of the impact of the revisions to the Quality Accountability Act enacted by Nebraska Legislative Bills #653 and #1157. These bills initiated a revision of the state content standards and, following the completion of this study, the development of state tests. The purpose of the study was to examine participants perceptions about the changes to STARS based on the enhancement criteria. After LB #653 was passed, a newly formed STARS Enhancement Design Team began work to enhance STARS through a revision of content standards. Seven enhancement criteria were agreed upon by consensus of the STARS design team. The Enhancement Criteria were:

- The system is in the best interest of ALL students.
- The system promotes best practices for teaching and learning.
- The system meets federal requirements.
- The system meets state requirements.
- The system includes a balance of classroom-based and large-scale assessment.
- The system is manageable.
- The system is fair, equitable and accurate.

The Enhancement Design team also commissioned three other teams, the Standards Advisory Team, the Assessment Advisory Team, and the Reporting Advisory Team. The teams consisted of NDE officials, Policy Partners, and school district personnel. The teams consisted primarily of members who were familiar with the history of STARS.

### **Purpose of the Study**

The purpose of the study was to explore the impact of the changes to the Quality Accountability Act and to examine the participants perceptions about the changes to STARS based on the enhancement criteria.

### **Instruments**

The STARS Enhancement Survey (Appendix E) was designed by the researchers to collect perceptions about the impact of changes to the Quality Accountability Act. Members from the four divisions (i.e., enhancement design team, standards advisory team, assessment advisory team, and reporting advisory team) were asked to participate in an online survey. The STARS Enhancement project members were surveyed following their second team meeting to assess their experiences as team members.

The survey examined the participants perceptions about the changes to STARS based on the Enhancement Criteria. Participants responded to a 33-item survey. Participants used a five-point Likert scale, with “1” representing “strongly disagree,” “2” “disagree,” “3” “neutral,” “4” “agree,” and “5” “strongly agree,” for 22 of the questions. Participants also responded to ten open-response questions and one multiple-choice question. Participants responded to seven

demographic questions, 23 general questions, and team specific questions that were automatically given to participants based on their team membership.

### **Findings**

Among the participants surveyed, most agreed with the enhancement criteria designed by the STARS team. However, the lower average scores on items about improvement in teaching practice (M=3.18) indicated a lower degree of confidence that the enhancements will improve teaching and learning while maintaining fair and equitable practices. Participants did express understanding of the enhancement criteria and indicated a personal responsibility to help improve STARS with the new enhancements. Participants recognized both strengths and limitations of the STARS enhancement process. Many believed that the revised standards would be beneficial in the long term. Some participants also favored the ability to compare student performance across districts. However, some participants saw this as a limitation of the enhancement design. The increased influence of the legislature and the increased time spent outside of the classroom that will be required in the STARS enhancement process were two of the main limitations of the enhancement process expressed by participants.

Participants were also asked to evaluate their experience within their team. The majority felt that the teams worked well together and that the team atmosphere was a forum to openly express opinions about the enhancement of STARS. Team members felt that their expertise was used within the team and that the time spent working in the team was worthwhile. Additionally, many of the Enhancement Teams expressed a need for additional time to accomplish the assigned tasks.

### **STUDY III: 2001 to 2007 Reading and Math Achievement**

**Shirley Mills, Ph.D., Assistant Professor, University of Texas-Pan American**  
**Jody Isernhagen, Ed.D., Associate Professor, University of Nebraska-Lincoln**

#### **Introduction**

Gallagher and Ratzlaff (2007/2008) referred to Nebraska's School based Teacher led Assessment and Reporting System (STARS) as the "road less traveled." Nebraska was the only state that opted to develop their own system of local assessments based on six quality criteria developed by the Buros Center for Testing (Plake & Impara, 2000).

1. Assessments align to state or local standards.
2. Students have an opportunity to learn the content that they will be tested.
3. Assessments will be free of bias or offensive language.
4. The level is developmentally appropriate for all students.
5. Scoring is consistent.
6. The mastery levels are appropriate to subject and grade level.

Nebraska's STARS requires each district to either adopt state standards or develop local standards that are at least equal to or exceed the state standards. Each district then developed a plan for assessing their standards. The plan was based primarily on locally developed criterion-referenced tests (CRT's), which were unique to that district. The STARS assessment results are reported at fourth, eighth, and eleventh grades. Districts also report Average Yearly Progress (AYP) at grades three through eight and one year in high school.

In Nebraska, districts are also required to administer a norm-referenced test (NRT) of their choosing (e.g., Terra Nova, Stanford Achievement Test). NRT's are perceived by many as reliable indicators of student achievement. As Stiggins (2007) indicated, "a major role of

assessment has been to detect and highlight differences in student learning in order to rank students according to their achievement” (p. 22).

### **Purpose of the Study**

The purpose of this study was to examine STARS student achievement data available to date for reading and math for all students. It provided an average of the percentage of students in Nebraska school districts demonstrating proficiency in these areas. The report included locally developed criterion-referenced data, norm-referenced data, and District Assessment Portfolio data.

### **Sample**

Data were included for Class III, IV, and V school districts. The districts in this study represented all public school students in Nebraska. The district data for this study were included on the state website and use of the data was facilitated by the NDE.

### **Findings**

#### ***Criterion-referenced Reading Achievement***

The district average percent of student scores reported by districts as proficient or better in locally defined criterion-referenced reading at the fourth-grade level increased from 74.99% in 2001 to 92.30% in 2007. The district average percent proficient for the eighth-grade level increased from 73.67% in 2001 to 89.79% in 2007. The district average percent proficient at the eleventh-grade level increased from 73.54% in 2001 to 87.48% in 2007. Proficiency on criterion-referenced measures increased at all grade levels each year; the average district gain from 2001 to 2007 was 17.31% at fourth grade, 16.12% at eighth grade, and 13.94% at eleventh grade.

#### ***Norm-referenced Reading Achievement***

The district average percent of students in the top two quartiles on the norm-referenced reading test used by districts at the fourth grade increased from 64.93% in 2001 to 69.25% in 2007. The eighth grade increased slightly from 62.85% in 2001 to 63.61% in 2007. The eleventh grade increased from 59.87% in 2001 to 62.05% in 2007. Proficiency, as determined by the percent of students in districts in the top two quartiles on norm-referenced measures, also increased from 2001 to 2007 with a 4.32% increase at fourth grade, a 0.76% slight increase at eighth grade, and 2.18% increase at eleventh grade.

#### ***Criterion-referenced Math Achievement***

The district average percent of students reported by districts as proficient or better in locally defined criterion-referenced math at the fourth grade-level increased from 78.29% in 2002 to 92.83% in 2007. The district percent proficient at the eighth-grade level increased from 68.58% in 2002 to 86.04% in 2007. The district percent proficient at the eleventh-grade level increased from 66.22% in 2002 to 84.20% in 2007. Proficiency on criterion-referenced measures increased at all grade levels each year; the increase from 2002 to 2007 at fourth grade was 14.54%, at eighth grade was 17.46%, and at the eleventh grade was 17.98%.

#### ***Norm-referenced Math Achievement***

The district average percent of students in the top two quartiles on the norm-referenced math test used by districts at the fourth grade increased from 68.12% in 2002 to 70.48% in 2007. The eighth grade increased slightly from 67.34% in 2002 to 68.60% in 2007. The eleventh grade decreased slightly from 67.49% in 2002 to 66.49% in 2007. From 2002-2007, proficiency on norm-referenced measures increased and decreased slightly from year to year at all grade levels.

However, the overall gain from 2002 to 2007 was a slight gain at fourth (2.36) and eighth (1.26) grades, with a slight decrease at grade eleven (-1.00).

#### ***District Assessment Reading Portfolio Ratings***

The total district average of Reading Assessment Portfolio ratings across grades four, eight, and eleven, on the “1” to “5” Likert scale, increased at grade four from 3.57 in 2001 to 4.83 in 2007. Portfolio ratings at grade eight increased from 3.48 in 2001 to 4.83 in 2007. Portfolio ratings at grade eleven increased from 3.46 in 2001 to 4.76 in 2007. The total district average of Reading Assessment Portfolio rating across grades four, eight, and eleven increased from 3.50 in 2001 to 4.35 in 2003, to 4.55 in 2005, declined to 4.48 in 2006, and increased to 4.81 in 2007. This was a total increase of 1.31 from 2001 to 2007.

#### ***District Assessment Math Portfolio Ratings***

The total district average for Math Assessment Portfolio ratings across grades four, eight, and eleven, on the “1” to “5” Likert scale increased at grade four from 3.98 in 2002 to 4.75 in 2007. Portfolio ratings at grade eight increased from 3.96 in 2002 to 4.85 in 2007. Grade eleven portfolio ratings increased from 3.96 in 2002 to 4.85 in 2007. The total district average of Math Assessment Portfolio rating across grades four, eight, and eleven increased from 3.97 in 2002, increased to 4.74 in 2004, declined to 4.61 in 2005, remained at 4.61 in 2006, and increased to 4.82 in 2007. This was a total increase of 0.85 from 2002 to 2007.

#### **Summary**

District criterion-referenced measures at grades four and eight continue to show growth over time in the areas of reading and math from 2001 to 2007. District norm-referenced measures have generally increased in reading and math at fourth and eighth grades, with a small decline at eleventh grade from 2001 to 2007. The district assessment portfolio ratings have increased for both reading and math at all grades from 2001 to 2007.

School improvement with student academic achievement as the goal was not intended to be a short-term process. Nebraska is in its seventh year of full implementation of the STARS program and variability still exists in achievement in some areas; however, Nebraska educators have made strides towards improving all students’ math and reading scores on the district average achievement scores. Generally, criterion-referenced and norm-referenced assessment scores have improved. The portfolio scores for reading and math have consistently improved over time.

#### **STUDY IV: 2002-2007 Statewide Writing Achievement**

**Shirley Mills, Ph.D., Assistant Professor, University of Texas-Pan American**  
**Jody Isernhagen, Ed.D, Associate Professor, University of Nebraska-Lincoln**

#### **Introduction**

Conversations with educators across the state of Nebraska have focused on writing skills since early 2000. Nebraska Legislative Bill 812, which amended State Statute 79-760 (Educational Quality and Accountability Act, 1999), required district involvement in a statewide assessment of writing for all students in grades four, eight, and eleven as a part of STARS. This bold step required significant professional development across the state. In the Nebraska School-based Teacher-led Assessment and Reporting System (STARS), districts first adopted state writing standards or local writing standards that were equal to or more rigorous than the state standards.

### **Purpose of the Study**

The purpose of this study was to examine the district achievement data available for the Statewide Writing Assessment for the Nebraska STARS program.

### **Sample**

Data was included for Class III, IV, and V school districts. The districts in this study represented all public school students in Nebraska. The district data for this study were included on the state website and use of the data was facilitated by the Nebraska Department of Education (NDE).

### **Statewide Writing Assessment Prompt Development**

The process of the development of writing prompts for use in the Statewide Writing Assessment relied on the involvement of Nebraska classroom teachers. Participating teachers were recommended by their district superintendent or assessment contact person and selected by the NDE each year to take part in a writing development task force. During the workshop, participants read and discussed examples of current research related to best practices in the teaching and assessment of student writing. A number of examples of writing prompts including those that had been used in previous Nebraska statewide writing assessments were also reviewed.

### **Statewide Writing Assessment Prompts Field Testing Process**

From information gathered at the Writing Prompt Development workshop, school districts representing various sizes and geographic locations were selected to field test the writing prompts with students in grades four, eight, and eleven before the end of the current school year. Multiple prompts were field tested across multiple school districts.

### **Sample**

The unit of analysis for this study was the district average percent of students rated as proficient in Class III, IV, and V school districts for the State of Nebraska in writing at grades four, eight, and eleven. While this statewide assessment took on some formal technical assessment characteristics that would more characterize norm-referenced tests (statewide common administration and scoring, common cut-score) than many criterion-referenced assessments, it was clearly not a comparison with a separate norm group. Descriptive data was, therefore, reported and discussed. However, because the assessment was a common measure across districts and was an equal interval scale, inferential statistics were also used to examine statistical significance between pre/post scores from inception to last scoring.

### **Results**

The district average percent of student scores reported by districts as proficient or better in locally defined criterion-referenced reading at the fourth-grade level increased from 76.50% in 2002 to 85.32% in 2007. The district average percent proficient at the eighth-grade level significantly increased from 79.55% in 2003 to 91.38% in 2007. The district average percent proficient at the eleventh-grade level increased from 89.22% in 2004 to 92.23% in 2007.

Writing scores increased at all grades, with grade four increasing significantly by 8.82% from 2002 to 2007, grade eight increasing significantly by 11.83%, and grade eleven increasing 3.01%. The positive perception of teachers reported in studies of the Nebraska Statewide Writing Assessment System (Anderson 2005, 2007; Gallagher, 2003) and writing gains from this study are consistent with the literature relative to the value of teacher involvement in the writing process.

## **Summary**

The purpose of this study was to examine the district achievement data available for the Statewide Writing Assessment for all students across Nebraska. Results in 2007 indicated that fourth and eighth grade made significant gains in the baseline comparisons on the Statewide Writing Assessment, while the eleventh grade displayed an increase in the baseline comparison. Nebraska's writing results are positive and would indicate support for continuation of the statewide writing assessment component of STARS. These findings provide a base of support and, along with the gains in reading and math, credibility for the general STARS process.

## **STUDY V: 2001-2007 Achievement for Special Populations**

**Shirley Mills, Ph.D, Assistant Professor, University of Texas-Pan American**  
**Jody Isernhagen, Ed.D, Associate Professor, University of Nebraska-Lincoln**

## **Introduction**

The 2007 testing year was significantly impacted by the mandates that were fully implemented this year for special populations, known as English Language Learners (ELL) and Special Education (SPED) students, across the nation. Beginning in 2007, all students were required to take their assessments at their respective grade level. This change was mandated by the No Child Left Behind Act in order to demonstrate increased academic achievement for all students.

Accommodations and modifications for students as indicated by their Individual Education Plans (IEPs) can be observed during the testing time, but the test itself must be the same as all other students. ELL students must take the English version of all tests. Students who are identified as students with special needs or are identified as having special learning problems will also take the identical test as all other students.

Although some states are faced with the possibility of lower scores for students taking tests at their assigned grade level, Nebraska is not one of them. Nebraska has been unique as students were always required to test at grade level. It is anticipated that the impact of the mandate will not significantly influence overall scores.

## **Purpose of the Study**

The purpose of this study was to examine the academic change in reading, math, and writing of special population students, in this case, ELL and SPED students.

## **Sample**

Data were included for Class III, IV, and V school districts. The districts in this study represented all public school students in Nebraska. The district data for this study were included on the state website and use of the data was facilitated by the NDE.

## **Methodology**

The criterion-referenced score (CRT) for reading and math was the district average percentage for ELL and SPED students meeting the proficiency level or better as defined by the local district for their locally developed measure in Classes III, IV, and V school districts for the state of Nebraska. Criterion-referenced scores for reading and math were unique for each district and not on a common scale. Therefore, descriptive data only was reported and discussed for reading and math scores.

The criterion-referenced score (CRT) for writing was the district average percent of ELL and SPED students rated as proficient in Classes III, IV, and V school districts for the state of

Nebraska in writing at grades four, eight, and eleven. The writing assessment was a common measure across districts and was an equal interval scale. Therefore, inferential statistics were used to examine statistical significance between pre/post scores in writing.

## **Results**

The study examined special populations' achievement data available to date for reading and math. District portfolio ratings for reading and math were also compared.

### ***ELL Reading Achievement***

The district average percent of ELL reported as proficient or higher in locally defined criterion-referenced assessments for reading at the fourth-grade level increased from 50% in 2001 to 79% in 2007. Proficiency on reading criterion-referenced measures at grade four indicated an increase of 29% from 2001 to 2007.

The district average percent of ELL scores of eighth-grade students increased from 47% in 2001 to 65% in 2007. Proficiency on reading criterion-referenced measures for eighth-grade ELL students increased by 18% from 2001 to 2007.

The district average percent of ELL scores of eleventh-grade students increased from 45% in 2001 to 57% in 2007. Proficiency on reading criterion-referenced measures for eleventh-grade ELL students increased 12% from 2001 to 2007.

Therefore, the district average percent of ELL students increased in reading at all grade levels from 2001 to 2007.

### ***ELL Math Achievement***

The district average percent of ELL reported as proficient or better in locally defined criterion-referenced assessments for math at the fourth-grade level increased from 53% in 2002 to 83% in 2007. Proficiency on math criterion-referenced measures for fourth-grade ELL students increased noticeably by 30% from 2002 to 2007.

The district average percent of ELL scores for eighth-grade students increased from 40% in 2002 to 62% in 2007. Proficiency on math criterion-referenced measures for eighth-grade ELL students increased 22 % from 2002-2007.

The district average percent of ELL scores for eleventh-grade students increased from 39% in 2002 to 61% in 2007. Proficiency on math criterion-referenced measures for eleventh-grade ELL students increased 22% from 2002 to 2007.

Therefore, the district average percent for ELL students increased in math at all grade levels from 2002 to 2007.

### ***ELL Writing Achievement***

The district average percent of ELL reported as proficient or better on the state criterion-referenced assessment for writing at the fourth-grade level increased from 49% in 2002 to 69% in 2007. Proficiency on the state writing criterion-referenced measure at fourth grade increased significantly 20% from 2002 to 2007.

The district average percent of ELL scores for eighth-grade students increased significantly from 37% in 2003 to 62% in 2007. Proficiency on the state writing criterion-referenced measure for eighth-grade ELL students increased significantly 25% from 2003 to 2007.

The district average percent of ELL scores of eleventh-grade students increased from 45% in 2004 to 56% in 2007. Proficiency on the state writing criterion-referenced measure for eleventh-grade ELL students increased 11% from 2004 to 2007.

Therefore, the district average percent for ELL students increased in writing at all grade levels.

#### ***SPED Reading Achievement***

The district average percent of SPED students reported as proficient or better on locally defined criterion-referenced assessments for reading at the fourth-grade level increased from 44% in 2001 to 81% in 2007. Proficiency of SPED fourth-grade students on reading criterion-referenced measures increased 37% from 2001 to 2007.

The district average percent of SPED eighth-grade students reported as proficient or better on locally defined criterion-referenced assessments for reading increased from 43% in 2001 to 72% in 2007. Proficiency on reading criterion-referenced measures for eighth-grade SPED students increased 29% from 2001 to 2007.

The district average percent of SPED scores of eleventh-grade students reported as proficient or better on locally defined criterion-referenced assessments for reading increased from 42% in 2001 to 65% in 2007. Proficiency on reading criterion-referenced measures for eleventh-grade SPED students increased 23% from 2001 to 2007.

Therefore, the district average percent for SPED students increased in reading at all grade levels from 2001-2007.

#### ***SPED Math Achievement***

The district average percent of SPED students reported as proficient or better on locally defined criterion-referenced assessments for math at the fourth-grade level increased from 51% in 2002 to 82% in 2007. Proficiency on math criterion-referenced measures for fourth-grade SPED students increased 31% from 2002 to 2007.

The district average percent of SPED eighth-grade students reported as proficient or better on locally defined criterion-referenced assessments for math increased from 34% in 2002 to 64% in 2007. Proficiency on math criterion-referenced measures for eighth-grade SPED students increased 30% from 2002 to 2007.

The district average percent of SPED scores of eleventh-grade students reported as proficient or better on locally defined criterion-referenced assessments for math increased from 28% in 2002 to 55% in 2007. Proficiency on math criterion-referenced measures for eleventh-grade SPED students increased 27% from 2002 to 2007.

Therefore, the district average percent for SPED students increased in math at all grade levels from 2002-2007.

#### ***SPED Writing Achievement***

The district average percent of SPED students reported as proficient or better on the state criterion-referenced assessment for writing at the fourth-grade level increased significantly from 46% in 2002 to 69% in 2007. Proficiency on the state writing criterion-referenced measure at fourth-grade increased significantly by 23% from 2002 to 2007.

The district average percent of eighth-grade SPED students increased significantly from 48% in 2003 to 55% in 2004. Proficiency on the state writing criterion-referenced measure for eighth-grade SPED students increased significantly by 19% from 2003 to 2007.

The district average percent of SPED scores of eleventh-grade students increased significantly from 55% in 2004 to 65% in 2007. Proficiency on the state writing criterion-referenced measure for eleventh-grade SPED students increased significantly by 10% from 2004 to 2007.

Therefore, the district average percent for SPED students increased in writing at all grade levels.

### **Summary**

Nebraska's special populations, ELL and SPED students, continue to demonstrate significant achievement gains. These special populations, however, continue to score lower than their fourth, eighth, and eleventh grade counterparts on group district averages. This is consistent, however, with most research in this area and, indeed, the basis for the special programs that are provided to support ELL and SPED students academically. ELL students increased their scores in reading, math, and writing from 2001 to 2007. SPED students increased their scores, as well, over the 2001-2007 time period.

### **STUDY VI: Student Mobility Effects on Achievement across Levels of Poverty**

**Jody Isernhagen, Ed.D., Associate Professor, University of Nebraska-Lincoln**

**Nino Zhvania, M.A., Graduate Assistant, University of Nebraska-Lincoln**

**Casey Tallent, M.A., Graduate Assistant, University of Nebraska-Lincoln**

### **Introduction**

Student mobility, otherwise known as “pupil turnover,” “transience,” or “turbulence” (Demie, 2002, p. 199) can be described as a “non-promotional school change” (Rumberger, Larson, Ream & Palardy, 1999, p. vi) or an “inconsistency or interruption in the educational experience” (Fisher, Matthews, Stafford, Nakagawa & Durante, 2002, p. 319). In essence, the phenomenon can be defined as an unscheduled classroom entrance or exit made by students within or between academic years (Texas Educational Agency, 1997).

The magnitude of student mobility in educational settings is a considerable problem. Mobility patterns observed across grades create an even graver picture. It was also found to have a negative impact on teachers and classrooms. Naturally, such effects on students, teachers, and classrooms will be reflected in overall school performance. Significant correlation between poverty status, school performance, and mobility were established in Offenbergs' study (2004). According to its results, students with average academic performance tended to move from low to high and from high to low achieving schools. The former pattern was especially true for students with lower achievement and high socioeconomic status while the latter was more common among students with lower achievement and lower socio-economic status (SES). Offenbergs (2004) also detected that students with higher SES and higher academic achievement tended to enter schools with the same level of achievement as the ones from which they had exited.

### **Purpose of the Study**

The purpose of this research study was to determine the influence of student mobility on the overall performance of districts. Specifically, the study examined how different levels of student mobility affect the performance of districts with varying levels of poverty. The

central question for this study was: “Does the interaction of mobility and poverty affect district achievement?”

### **Methods**

This study used quantitative data, reported by the state, to examine the impact of student mobility on district performance. Additionally, this study addresses the consistency of the claim that high student mobility persistently leads to low district achievement.

Data from 212 school districts in Nebraska were used. At the time of this study, the NDE only collected data in the form of district aggregates. The two independent variables were student mobility and poverty level. The student mobility coefficient was calculated according to the statewide formula used at the NDE. Specifically, student mobility is perceived as the ratio of all students who enter or exit any particular school/district between the last Friday in September and the last day of school in the overall school/district population. Mobility was divided into three levels based on the state average of 13.77%; districts with a mobility rate of 14% or higher were placed in the high mobility group; districts with a mobility rate between 9% and 13.99% were placed into the average mobility group; while districts with a mobility rate below 9% were placed in the low mobility group.

The district performance variable was measured according to student performance on criterion-referenced assessment (STARS Assessment) and norm-referenced (ITBS, Terra Nova, CAT, MAT etc.) in eighth-grade math. The data on both variables were analyzed longitudinally across three consecutive school years of 2003-2004, 2004-2005, and 2005-2006.

### **Results**

This research study uncovered the following findings. First, analysis of the results showed that mobility and poverty in combination have an effect on districts’ achievement. The effect of mobility or poverty alone is inconsistent. Second, high mobility rates do not necessarily lead to low achievement. The data indicated that districts with high rates of mobility might perform worse than districts with either low or average rates of mobility. However, this does not rule out the chance for these districts to perform as well as those districts with average or low mobility rates. It is maintained that even when mobility rates remain constant, achievement levels vary. Finally, the study revealed that poverty and mobility may interact differently to effect criterion-referenced achievement versus norm-referenced achievement. There are likely additional factors that account for this difference.

These findings partially corroborate the findings of previous research. Specifically, our results converge with other studies that have found the combination of mobility and poverty to have a strong negative impact on academic achievement (Ingersoll, Scamman, & Eckerling, 1989; Kerbow, 1996; Texas Education Agency, 1997).

However, the study disagrees with the evidence found in mobility research literature in two ways. First, studies have found that a combination of mobility and other factors, such as poverty persistently lead to low achievement (Kerbow, 1996). This study found that in most cases poverty had no impact on the performance of districts where mobility rates were low and average, yet found consistent results across high mobility districts. Second, previous research claims that high levels of mobility are consistently associated with low achievement (Mao, Whitsett, & Mellor, 1998; Texas Education Agency, 1997). This study has found that even districts with high rates of mobility are capable of performing as well as districts with average or low rates of mobility, especially in schools with low to average poverty levels.

There are several limitations inherent in the study. First, due to the aggregate nature of the data, it was not possible to examine the interaction of poverty and mobility on achievement with mobile students only. The study can only be generalized to district mobility, poverty, and achievement.

## **Study VII: The Effect of Nebraska's Standards & Accountability System (STARS) on School Improvement Practices**

**Jerald Riibe, Assistant Superintendent for Curriculum and Instruction  
Ralston Public Schools**

### **Introduction**

The initial emphasis for standards and accountability in Nebraska was to guide instruction and promote school improvement. The standards-wide accountability process in Nebraska is the School-based Teacher-led Assessment and Reporting System (STARS). The rationale for standards and the subsequent assessment of student achievement on those standards was to provide a catalyst for school improvement. One measure of the impact of standards and accountability in Nebraska is the role it plays in school-wide conversations regarding school improvement.

### **Purpose of the Study**

The purpose of this quantitative study was to compare the perceptions of Nebraska second-grade teachers (a non-reporting grade) with perceptions of Nebraska fourth-grade teachers (a reporting grade) regarding the effect of Nebraska's standards accountability system (STARS) on school improvement practices.

### **Methods**

A survey was used to generate quantitative data describing second and fourth-grade teachers' perceptions of Nebraska's state standards process. The goal of this research project was to evaluate the mean difference between two populations: second and fourth-grade teachers. The independent-measures  $t$  statistic was used to draw inferences between two populations (Creswell, 2002). Each research question used a  $t$  test to measure the statistical difference between the sample populations.

### **Findings**

The research questions examined the impact of an accountability system on school improvement practices. The results of the study presented two different perspectives that compared the perceptions of Nebraska second-grade teachers (a non-reporting grade) with perceptions of Nebraska fourth-grade teachers (a reporting grade) regarding the effect of Nebraska's standards accountability system (STARS) on school improvement practices.

Teachers in reporting and non-reporting grades did not have a statistically significant difference in how STARS was perceived. This would indicate that the processes in place to meet STARS requirements are not isolated to reporting grades.

### **Summary**

The finding of no statistical difference of perspectives in the research questions indicates that STARS is not limited to reporting grades. The similar perspectives of reporting and non-reporting grade teachers indicate a certain universality that is important for any accountability system. If the current system were to be replaced by a high stakes statewide test, the questions asked in this study would be just as relevant to that system.

The Nebraska standards accountability system is not without concerns. However, the positive finding is that teachers in this study rarely linked STARS to negative developments in achievement, curriculum, or assessment. STARS has moved teachers, buildings, and districts toward accountability for student achievement. That is important and provides solid footing for future school improvement efforts.

Nebraska's STARS process has been legislated to resemble other state universal testing systems. The legacy that STARS will leave is not in the assessments created or the accountability reports. STARS may be labeled successful if the idea of blending accountability and school improvement transcends grade levels and content areas. The findings of this study would indicate STARS has had an effect in moving some teachers from isolated classrooms to an improved school culture.

### **STUDY VIII: Nebraska Educators Review the Local Math Assessment Process: Reliability of Peer Review of Assessment Portfolio 2007-2008**

**John L. Moon, Assessment Coordinator, Nebraska Department of Education**

#### **Introduction**

Beginning in 2006, the evaluation of portfolios included a new District Assessment Portfolio Rubric and the utilization of on-site peer reviewers. There is a need to examine the validity and reliability of the revised system. Quality assessments are necessary for schools to report reliable and valid data on student achievement. This study will look at the impact the revised rubric and the use of on-site reviewers has had on the examination of each portfolio. The revised assessment rubric needed to be evaluated to ensure that the collection of information was applied in a consistent manner to district assessment portfolios.

#### **Purpose of the Study**

This study evaluated the Peer Review Process by establishing the reliability of reviewer decisions in collecting information from district assessment portfolios. The study will attempt to answer the following question: Is there agreement between the reviewers on rating indicators of assessment portfolios for each of the Six Quality Assessment Criteria?

#### **Methods**

Data was collected using the "assessment rubric" of the six quality criteria indicators to measure reviewer consistency in evaluating district assessment portfolios. On the third day of training the reviewers were given a sample portfolio to evaluate using the District Assessment Portfolio Rubric. Each reviewer completed a review of the six quality criteria and recorded the results on a "District Assessment Portfolio Worksheet" (Appendix G). From the data collected, an analysis was made to determine reliability of evaluations made by different reviewers on the same portfolio. Assessment Specialists along with the reviewers used the indicators (five to nine per criteria) to rate whether the portfolio met the criterion or not. A count of ratings for each indicator was used to calculate percentage of the rating agreement between the reviewers. The on-site peer review process's success was dependent on the reviewers correctly evaluating each indicator in the district's portfolio.

#### **Findings**

The results of the study indicated a high level of agreement between reviewers on evaluating the indicators of the three sample portfolios. This level of agreement provides support that the initial step in the 2007-2008 evaluation process, namely the collection of data by the peer reviewers, does provide an accurate picture of the district assessment process.

## Summary

Analysis of the data collected from “Evaluation Survey of Peer Review Training” (Appendix H) for 2007-2008 should be completed. Additional investigation of this process should focus on the different components of the peer review training and the effectiveness of the training in establishing accurate guidelines for the peer reviewers. Other uses of this process may be appropriate for educational evaluations by the state in the future. Additional study in how the process was perceived by reviewers and district staff as well as assessment specialists would help to identify how the process benefits the educators in Nebraska and eventually the students. Using evaluation rubrics similar to the District Assessment Portfolio Worksheet when evaluating local school improvement processes should be encouraged to continue the benefits of the current state level assessment process.

## CONCLUSIONS FOR STUDIES I-VIII

*As spring 2008 arrived in Nebraska and the conversations blossomed over new assessment legislation, schools in Nebraska had time to reflect on their assessment journey and its impact on student learnings . . .*

### **STARS . . . The first six years**

**Learning has been a process, an insightful process** as shared by one educator, “I think it’s a really good process. I think the State of Nebraska has been, and I know continues to be, commended. It’s not an easy process. It’s hard work. But I think it’s valuable, it’s valuable work! Putting the classroom and the teachers at the heart of the process adds the value.”

**Reflecting on the early years of this process, one recalls . . . Starting out, this was a difficult process with a great deal of change,** “So where do we begin? Do we assess every single standard? Do we clump them together?” asked a teacher. Teachers and leaders felt that they were not experts but then began to realize that others could provide guidance and training as shared by this staff developer, “Our primary role has been a facilitator of the design and refinement process. We really facilitate the process of both the development and the refinement of criterion-referenced assessments.” With some practice, and some patience, Nebraska educators began to see the possibilities and what they truly were capable of doing when acting in the best interest of students. As educators looked back six years later, they commented, “When it’s all lined up and we see the finished product, we can see that it is a good thing.”

**The six year journey of educators in Nebraska schools has been unique for each district as they journeyed to different places along the path.** While some have traveled a great distance on this learning path, others have a distance yet to go. However, for educators there have been many different stories and conversations shared. Researchers listened throughout the 650-plus interviews conducted, they listened to teacher conversations of learning, they listened to stories of newly found collaborations, they listened to stories of working together with other educators in the best interest of students . . . and they listened to stories of newly formed learning teams, with smiles on many of their faces as they reflected upon the positive impact this journey has made on student achievement.

**As the knowledge and confidence of educators grew they began to see the rewards.** Slowly, as educators collaborated, they began to realize that they were more knowledgeable and able to develop the assessments needed to measure student achievement as this superintendent shared, “I

think our staff is light years ahead of where they used to be in developing assessments. I think they're more concerned with what they're assessing and how students are progressing."

**This journey included the ability to compile, analyze, and use data to improve teaching and learning.** With practice, educators learned to use data and information to impact student achievement. Conversations emerged as they shared their new learnings, "There are three teachers at the sixth grade. We discovered someone was doing well in one section and a couple of us weren't doing as well, so we asked (him) for some ideas . . . my kids were doing very well and so they asked me for some ideas." The conversations shared while on this learning journey have been about many things: Achievement is really about making a difference; Reaching each learner is critical for meaningful achievement to occur; and recognizing that "all students" includes those in special populations, are all a part of achievement! Education can *really be about leaving no child behind.*

**Challenging students and learning new ways to use assessments in a manner that motivates learners has been a part of this journey.** Educators have walked students outside the familiar, outside of their comfort zone that extends their learning to new heights as shared by a rural teacher, "It truly is about can they learn this specific thing, and then give them a chance to learn it and assess it. That's what we're doing. It's made us better. I think that we are better, but the bottom line is, it makes our kids better . . . learning is long term, not just short term."

**Additionally, the STARS process has impacted instructional practices in schools.** It has led to a rich understanding and desire to connect best practices in teaching to student learning as shared by this superintendent, "What we've seen is that we went from the beginning of this process spending all of our time on curriculum and assessment development and nothing on the instructional piece. It was the curriculum assessment process. Now it's the curriculum, instruction, and assessment process. We've evolved to the point that we are now focusing on the instruction piece." Educators worked to broaden their curriculum teaching practices to include intervention strategies in order to reach all students. "We're [working in] learning teams. We're focusing on strategies and interventions. So we have used data to be a driving force in interventions that we are providing and will continue to provide, and also a driving force in staff development."

**New leadership roles have also emerged during this process.** There is newly found "leadership in the classroom." Many teachers today are wearing various "hats" such as classroom teacher and co-assessment coordinator. The benefits are evident as shared by this rural superintendent, "They're in the classroom, they're in the trenches, and they're in every building. They share the information with the other teachers. They coordinate, they report directly back to administration and leadership. It just seems to be a better fit."

**Teachers have not only established themselves as leaders but established ownership in this process.** For many, that was the real beauty of this process, the evolution of STARS as shared in conversations by various educators, "So, one (teacher leader) in every building who was really in charge of it, and they kept us more informed and kept us . . . more involved on a one-to-one basis so the teachers had better ownership." Another teacher echoed, "It's an exciting process . . . the exciting part to me is that teachers are designing this, teachers are the people that are using it and are really involved in the whole process of designing the questions . . . there's ownership in the whole process."

One administrator summed up the impact of STARS by stating, ". . . the knowledge that people had gained because of the process (STARS) . . . the processes that we had to go through, have

been just exceptionally good for staff at all levels.” Teachers have become skilled practitioners in their own area *and* contributors to the process. **Teachers and leaders have grown in the process together, with teacher involvement and collective collaboration being one of the essential pieces of the journey for the past six years.**

### **STARS . . . The seventh year**

**This year’s journey and the journey ahead brings with it new conversations.** New paths are yet to be taken as educators look ahead to a new era with the addition of statewide tests as a part of Nebraska’s assessment system. The past has prepared Nebraska educators for what lies before them as shared by one rural high school math teacher, “I think that through this process, we have more people that do understand what’s going on and that’s probably one of the good things about it.”

During the seventh year, educators shared new learnings from this past year, reflected upon learnings gained over the past several years, and expressed some apprehension as we move to an enhanced assessment system. During the year it was noted that the third-year STARS report (2003-2004) was also entitled “*Conversations.*” The difference between the third year study and the seventh-year study is the depth of educator conversations. Conversations this past year with the researchers included detailed discussions on new learnings as an outgrowth of STARS, the achievement of students, and the future development of the Nebraska assessment system. These new learnings included:

- CIA . . . Curriculum, Instruction, Assessment
- Instruction - Teaching and Re-teaching
- Assessment and Student Motivation
- Student Achievement and Accountability
- Teacher Knowledge and Confidence
- Teacher Collaboration and Involvement
- Integrated State Visits

The essence of the discussions and the new learnings are shared through the voices of Nebraska Educators:

#### **CIA . . . Curriculum, Instruction, Assessment!**

Educators’ conversations today are consumed with three elements: Curriculum, Instruction, Assessment *and* the connection between them. Most educators acknowledged that without partaking in this journey, they would not have produced the same conversations and ultimately impacted student learning in new ways, “The reinforcement of point of instruction was the place to go. And there’s no doubt in my mind it is the way to go.” “The research says that you have to know exactly what it is these kids need to be taught. You have to find the best way to instruct them and the best way to assess it.”

**The “right” curriculum was seen as a key to positive student progress as shared by a female elementary teacher,** “Before we started our assessments, we spent a long time redoing our curriculum. I think that was key for us working with fourth grade which is such a big assessment year. But once we got the curriculum where it flowed a little bit better, it made sense. I think it made it easier for our assessments. So everything flows a little better since we’ve gone through the curriculum first.”

**Teachers are more focused on the appropriate instruction for students to be successful as shared by this administrator,** “The important thing out of all of this is that as we learn to be better at assessments, we learn to be better at instruction. That’s really where we want everything to go back to, is the instruction.” Another rural principal shared, “Take a look at the curriculum .... OK, this is an essential skill. Let’s pull in a different lesson. Let’s pull in a unit that will make sense. Let’s bounce around (in) the book. Let’s don’t start with Chapter 1 and end at Chapter 49. Let’s see what we can do. It’s good for kids.” A high school rural educator summarized, “Our teachers became better teachers. They’re more aware of what students’ needs are at all levels. They have learned new ways to teach or new methods to teach the different levels of children or students. I just think it’s a wonderful process. It holds you accountable so you understand that you need to keep up.”

#### *Instruction: Teaching and Re-teaching*

**Not only have educators become more skilled at instructing students, but they also have become more aware of what students know and do not know** as shared by this curriculum director, “Hopefully, a lot of people have gained in their understanding that a lot of what this is about is teaching and then identifying what the students did not learn and going back and re-teaching and coming up with different strategies. Why was Tommy able to learn this, but he wasn’t able to learn that? What do we need to do differently in our instruction? Children can learn . . . sometimes not on the same day and not in the same way. It’s about the whole child and teaching that child what he or she needs and starting where that child is and going forward from there.”

#### *Assessment and Student Motivation*

**Assessments with immediate feedback and assessments used for student evaluation purposes are promising strategies to use as a motivating tool** for students as one high school math teacher shared, “It’s been wonderful. In the past, the students didn’t care what they were doing nor did they care if they passed or failed. Now it’s just built into the curriculum and the assessment process and it matters. All of a sudden, they’re concerned and trying. It’s made a huge difference. They get immediate results.” Another assessment coordinator shared the advantage of online assessments in regards to student feedback, “That’s one of the nice things that we found out about online. It’s motivating for students to get that immediate feedback and finish their assessments. Right after the assessment is taken, there are reports that they (teachers) can call up that show them visually and numerically exactly how their students did and what items they got correct.”

#### *Student Achievement and Accountability*

**It is evident that student achievement and accountability are now more than ever a part of educators’ conversations** as shared by this rural elementary teacher, “I think it’s made me more aware of what I needed to do and become accountable myself for what I’m teaching. It’s really focused my goals on . . . what I need to cover and recover or reteach to make sure my students understand..” Another rural, high school educator shared, “We . . . want to go in our room and just do our thing and say leave us alone. But we do need to be accountable to our patrons, to our students, to the parents.” Another elementary teacher echoed the connection between assessment and accountability, “It’s given me a better insight as to where my students are going to take this information and the importance of (STARS).”

#### *Teacher Knowledge and Confidence*

**As educator accountability increased, they became more knowledgeable about the process and their confidence continued to build** as indicated by this comment from a rural

superintendent, “To see not one of them lacks the confidence now . . . it’s just heartwarming to see that they all know they can.” Another administrator echoed this same comment but added, “This has been both rewarding but also placed a heavy burden on educators.” The conversations with researchers reflected the accountability “pressure” placed on educators but it also revealed that many embraced the opportunity. “The other ‘ah ha’ is that I really admire teachers that are in a small school like we are because . . . when our assessment results come back, they reflect on one person. There’s only one math teacher at our high school level. That is tremendous pressure. I’ve gained a whole lot of sensitivity about that. That’s a whole lot of pressure. Our teachers want our kids to be successful. They’ll do whatever because ultimately they realize it does reflect on them and them alone personally. Why didn’t they get that better grade? Why weren’t they proficient? They are accountable and they do take it personally. They are doing all things in their power.”

**Teacher confidence and skills continue to grow.** Teachers are not only learning new tools, new interventions and new strategies to help all children succeed; teachers are using their education expertise to “leave no child behind” as shared by this teacher, “I feel I have grown a lot, not only in the classroom but outside the classroom by being the assessment coordinator. I have a better understanding of what all of this means, and realized, we have to keep this process going . . . getting better . . . reaching what we need to for our students. Globally, we’re taking some very big steps and without this process, I don’t think we would probably keep up as well, but we know what those expectations are, what we need to meet.”

#### *Teacher Collaboration and Involvement*

**Through conversations, teachers and leaders have formed a bond that connects them to something much deeper. This personable connection has resulted in a more unified effort to improve student learning.** Lewin and Regine (2000) emphasized, “Actually, most people want to be part of their organization; they want to know the organization’s purpose; they want to make a difference. When the individual soul is connected to the organization, people become connected to something deeper – the desire to contribute to a larger purpose, to feel that they are a part of the greater whole, a web of connection” (p. 27). An assessment coordinator noted, “It’s surprising the groups of people who might be available that end up working together.” Another teacher noted that not only is the collaboration growing, but the work that is taking place in the collaboration is making a difference, “What we’re doing, it’s being valued. People are taking it seriously. It’s not just being filed away somewhere. There are many capable retired teachers, people that are even still currently teaching that do this process that have a wealth of information to share.”

**This past year, many conversations reflected the value of teacher involvement.** For some, teacher involvement meant all teachers supporting each other as shared by a female middle school teacher, “I just think it’s so important that they first invite teachers to be involved in it and secondly allow the teachers to speak to each other. . . . When we went to the fourth-grade teachers and told them we think you have way too much on your plate, they were so thankful that we recognized it and we were willing to take on more of what they had been doing. We don’t just look at it as ‘I teach sixth grade and I’m done.’ It’s I’m getting this child ready for the next level.”

Fullan (2004) emphasized, “The organization must frame the giving and receiving of knowledge as a responsibility” (p. 126). In the STARS process, teachers developed ownership from being involved. This was widely reflected in educators’ conversations regarding their involvement in the Math Portfolio Peer Review Process as reflected by this educator, “To me it was a new

learning . . . I think the teacher ownership was greater the second time around (second time for the portfolio review process). But the comfort level having been there, done that, it could have just been that we had (teacher) ownership in all areas.”

**The Portfolio process also yielded increased educator confidence in the assessment process** as shared by this teacher, “It was kind of more my attitude at first. I’m just thinking, ‘I hope I can do this all right.’ (Now) I feel more confident about giving the assessments. I don’t feel intimidated by giving assessments, and I feel I am definitely a better assessor.”

Issues that were noted as problematic in the first round of the portfolio peer review process were not noticeable in the second year and thus problematic issues appear to have been “fixed” before beginning the second year of portfolio peer reviews. Revisiting the portfolio process every year was also something that appeared to yield positive results as indicated by this rural assessment coordinator, “I think we go through a process and once it’s over, it goes away and you move on to something else. So coming back to the process each year and going through it and thinking about does this match, is it appropriate, is it reliable? It’s beneficial for all of us. Otherwise we go on auto pilot; we go through the motions instead of really thinking about and knowing what we’re doing and why we’re doing it.”

The essence of teacher collaboration is widely evidenced in Study VII. Positive findings at second and fourth-grade levels showed no differences in teacher understandings with respect to STARS, indicating that the understanding of STARS has spread to non-assessed grade levels, those outside of grades fourth, eighth and eleventh. Prior to this study being conducted, recommendations had been made in the comprehensive evaluation of Nebraska’s School-based Teacher-led Assessment and Reporting system (STARS) (Isernhagen & Mills, 2007) that all K-12 teachers should be participating in the assessment process. Although this study examined only second and fourth-grade teachers, the collaboration that has occurred across grade levels appears to have enhanced curriculum, instruction, and assessment knowledge of K-12 teachers. Furthermore, the positive findings in Study VII indicated that STARS has moved teachers, buildings, and districts toward accountability for student achievement. This is important as it provides solid footing for future continuous improvement efforts.

**Additionally, it should be noted that teacher involvement and teacher collaboration require a great deal of time.** Collaboration and involving teachers in making classrooms better learning environments must be supported by resources as shared by these rural school administrators, “You have to create time for it. We found out with all of what we’ve done with assessments and STARS, you have to create time. You have to create time to train teachers. You have to create time to teach. You have to create time to assess. You have to create time for people to have an opportunity to work with the material. You have to create time to then analyze and make another decision about what you do.” “So, it’s important to provide leadership and it’s important to provide time.”

**However educators are also realizing that the work is never really done when teacher ownership is created** as shared by this assessment coordinator, “Stiggins sticks in my head . . . It’s never a finished product. I think you have to go through the process to really understand . . . it isn’t ever finished because we can always make it better.”

**For several years a need for tracking students of mobility has been evident to ensure that assessment data stays with the student.** Throughout the state, student mobility was a continual issue reflected in prior conversations with researchers such as this statement from a rural

superintendent, “Without a job . . . they move, but they can’t be successful, so they come back. They are gone for two months. What school did they attend when they were gone? I think you’ve got to have some type of a program . . . to put them back in the classroom. When they come back, we find out where they’ve been, where they’re at.” Study VI reveals that mobility is a factor in student achievement but this can be positively influenced by other factors. Educator conversations revealed the need for a tracking system at the state level that denotes academic achievement levels for *each* student in the state, as well as when students leave their last education institution. This would allow educators to properly place students when they enter their new school district and potentially enhance the opportunity to impact student achievement.

**More parental involvement in reporting data is an issue of importance as shared by an educator,** “There’s also an option which we are not using yet. On my management system, there’s a parent portal so parents can log and see how their students have done on a certain standard assessment.”

### *Integrated State Visits*

**In 2007-08 the Nebraska Department of Education, in an effort to reduce the number of visits a school received in a calendar year, piloted an integrated visit in a limited number of schools.** The integrated visit process had multiple formats for a single visit. The initial feedback from school personnel indicated a better understanding of how various programs were connected within their school as shared by this rural high school math teacher, “It was just a really positive experience. I would highly recommend trying to give more integrated visits. I feel that you could see the bigger picture when you tie it all together. It makes you grow as an individual and as a district.”

### *Summary*

Through the multitude of conversations shared this past year (2007-08), a common echo was heard as summarized by this educator, “Well . . . I’ve been in education for quite a while. Basically, most of us when we first started teaching, we moved by the seat of our pants. Even though you might have a textbook, you might have a curriculum that the district provided; there was no assurance that curriculum was the same as another curriculum 20 miles down the road. So, as much as I probably ‘scoffed’ at looking at the standards, they gave me a basis of knowing what I needed to teach my kids, and to assure that when my kids walked out of my room, they were getting what they needed just like any other eighth grader across the state of Nebraska. I changed a lot . . . I think it (STARS process) makes you a better teacher. It makes you more accountable and it makes you more aware of what your students need. I applaud. I applaud what Nebraska has done.”

Nebraska’s assessment journey continues during the 2008-2009 school year with the implementation of **Nebraska State Accountability (NeSA)**, the new statewide accountability system featuring statewide tests in reading, math and science. This year will be the pilot for the first statewide assessment, the Reading assessment, developed by Nebraska educators.

Nebraska leaders emphasized that Nebraska’s focus will continue to be on student learning. Pat Roschewski, Director of Statewide Assessment shared that with the implementation of the new statewide assessments, the focus will continue to be on student learning while expanding the vision and finding the balance for Nebraska’s assessment system. The different assessment tools can and will provide evidence for various purposes that all come together in support of student achievement. Although educators recognize the efforts of the state department, they voiced some

apprehension moving to a statewide testing format. The following were concerns that may need to be considered as the statewide testing system is developed:

- A rural superintendent shared, “We’re all concerned with what’s going on with the legislature and the impact of throwing away what we’ve done. I do feel that the general public, and this is being reflected by the legislature, values simplicity as much as content. That’s very dangerous because if the goal is to be simple, you lose content.”
- When developing a statewide test the “quality criteria for bias” requires consideration be given to diverse student populations in Nebraska. Issues such as poverty, nontraditional guardianship, ethnic, and religious differences, etc. need to be considered. A rural superintendent shared, “You have some kids that have never actually been in a typical kind of place . . . that ever present acknowledgement of poverty. I have at least six children that are being raised in nontraditional settings with grandparents or guardians where even the mere mention of mom or dad can be a bias.” Another educator stated, “We were cognizant of the fact that we had to be careful, particularly with our Hispanic population.”
- “I don’t see a state test given once a year informing instruction the way our local assessments do.”
- “I hope it’s not that way. But I’m afraid that, looking at what other states are doing, it’s comparability, not improvement of instruction.”
- “I’m predicting that we might still use these assessments. I just don’t know if we’ll use them to the point that we do now.”
- “I’m quite resistant to it because . . . I don’t know how they’re going to produce that state test so it actually reflects what teachers do in the classroom.”
- “You lose that ownership. Nebraska teachers work hard. We’re good and our kids learn. I think that . . . we’re doing the best we can. I think it will be interesting to see what happens.”

Nebraska educators and leaders will once again be challenged to embrace the change and strike the balance between classroom based, criterion-referenced assessments, statewide tests, and norm-referenced tests. Nebraska constituents will also be challenged to boost their understanding of the nature and purposes of classroom-based, criterion-referenced assessments, and statewide tests. Educators around the state may need to give special attention to and work with local media to educate constituents about the need for a balanced assessment system.

Educators will once again be sparked by the new efforts to improve achievement for the students of Nebraska and reinforced by the knowledge and confidence gained over the past seven years. Nebraska educators believe their journey has prepared them well for what lies ahead as stated by one rural educator, “I think we’ve grown a lot. I think our teachers are way more assessment savvy than we ever were before, probably more than other states because when we are involved with it this much, you’re just going to have to learn it!” Another educator restates the value of the journey to the new learning ahead, “We have enough expertise now. We’ve developed enough expertise with our teachers that even if the state doesn’t mandate that we do local assessing, we’re still going to continue doing it because its what’s good for kids!” Finding the balance will be the next challenge for Nebraska educators as they venture into the 2008-2009 school year, but Nebraska educators and Nebraska students are up to the challenge!

## REFERENCES

- Anderson, S. (2005). *Statewide writing assessment scoring: Effective professional development for the classroom teacher. Charting STARS: Voices from the field. Year-Four Report. Comprehensive Evaluation of Nebraska's School-Based Teacher-Led Assessment and Reporting System (STARS)*. Lincoln, NE: University of Nebraska.
- Anderson, S. (2007). Statewide writing assessment scoring: Effective professional development for the classroom teacher. In J. Isernhagen & S. Mills (Eds.), *Charting STARS: Pathways to success. Year-six report. Comprehensive evaluation of Nebraska's School-Based Teacher-Led Assessment and Reporting System (STARS)* (pp. 38–40). Lincoln, NE: University of Nebraska.
- Bond, L. A. (1996). Norm-referenced and criterion-referenced testing. *Practical Assessment: Research and Evaluation*, 5(2). Retrieved on January 5, 2007, from <http://PAREonline.net/getvn.asp?v=5&n=2>
- Creswell, J.W.(2002) Educational research: Planning, conducting, and evaluating quantitative and qualitative research. New Jersey: Pearson.
- Demie, F. (2002). Pupil mobility and educational achievement in schools: An empirical analysis. *Educational Research*, 44(2), 197-215.
- Educational Quality and Accountability Act (The). (1999). State Statute 79-760. *Educational Quality and Accountability Act (The)*. (2000). LB 812 amending State Statute 79-760.
- Fisher, T. A., Matthews, L., Stafford, M. E., Nakagawa, K., & Durante, K. (2002). School personnel's perceptions of effective programs for working with mobile student and families. *The Elementary School Journal*, 102(4), 317-333.
- Fullan, M. (2004). *Leading in a culture of change: Personal action guide and workbook*. San Francisco, CA: Jossey-Bass.
- Gallagher, C. W. (2003). *Charting STARS: Sustainability as challenge and opportunity. Year two report*. Lincoln, NE: University of Nebraska.
- Gallagher, C. W. (2007). *Reclaiming assessment: A better alternative to the accountability agenda*. New Hampshire: Heinemann.
- Gallagher, C. W., & Ratzlaff, S. (December 2007/January 2008). The road less traveled. *Educational Leadership*, 65(4), 48-53.
- Ingersoll, G. M., Scamman, J. P., & Eckerling, W. D. (1989). Geographic mobility and student achievement in an urban setting. *Educational Evaluation and Policy Analysis*, 11(2), 143-149.
- Isernhagen, J.C. & Mills, S. (2007). *Charting STARS: Pathways to success. Year-Six Report. Comprehensive Evaluation of Nebraska's School-based Teacher-led Assessment and Reporting System*. Lincoln, NE: University of Nebraska-Lincoln.
- Kerbow, D. (1996). Patterns of urban student mobility and local school reform. *Journal of Education for Students Placed at Risk*, 1(2), 147-169.
- Lewin, R., & Regine, B. (2000). *The soul at work*. New York: Simon & Schuster
- Madaus, G. F. (1988). The influence of test taking on the curriculum. In L. N. Turner (Ed.), *Critical issues in curriculum Eighty-seventh yearbook of the National Society for the Study of Education, Part 1* (pp. 88–121). Chicago, IL: University of Chicago Press.
- Mao, M. X., Whitsett, M. D., & Mellor, L. T. (1998). Student mobility, academic performance, and school accountability. *ERS Spectrum*, 16(1), 3-15.
- No Child Left Behind Act of 2001. (2002). *An Act to close the achievement gap with accountability, flexibility, and choice, so that no child is left behind* (Pub. No. 0757). Washington, DC: U.S. Government Printing Office.

- Offenberg, R. M. (2004). Inferring adequate yearly progress of schools from student achievement in highly mobile communities. *Journal of Education for Students Placed at Risk*, 9(4), 337-355.
- Plake, B. S. & Impara, J. C. (2000). *Technical criteria for Nebraska school district assessment*. Lincoln, NE: Buros Center for Testing.
- Rumberger, R. W., Larson, K. A., Ream, R. K., & Palardy, G. J. (1999). *The educational consequences for California students and schools. Research Series*. (Report No. PACE-99-2). Berkeley, CA; Stuart Foundation. (ERIC Document Reproduction Service No. ED441040).
- Stiggins, R. (2007). Assessment through the student's eyes. *Educational Leadership*, 64(8), 22-26.
- Texas Education Agency. (1997). *A study of student mobility in Texas public schools. Statewide Texas educational progress study report No 3*. (Report No. RE7-601-03). Austin, TX; Texas Education Agency. (ERIC Document Reproduction Service No. ED424664).

## RECOMMENDATIONS

### *These recommendations remain from previous studies:*

1. Work collaboratively with ESUs to provide data training, aligning appropriate grading and reporting systems, and assessment knowledge for new teachers.
2. Help districts to develop quality, seamless instruction especially for students not meeting the standards.
3. Continue the state writing scoring process that enlists the participation of classroom teachers as a way to provide them with valuable training that relates positively to their classroom practices.
4. Research the assessment literacy knowledge and skills provided by higher education for teachers entering the field.

### *New recommendations based on 2007-08 study:*

5. Educate **all** constituencies about the different purposes of criterion-referenced assessments, Nebraska Statewide Accountability Tests, and norm-referenced tests.
6. Educate **all** constituencies about the different results between criterion-referenced assessments, Nebraska Statewide Accountability Tests, and norm-referenced tests.
7. Work collaboratively with ESUs to provide professional development particularly in the areas of classroom based assessment and instructional interventions for students (i.e., not mastering the standards, special populations).
8. Expand the integrated statewide visits as these visits encourage a more comprehensive vision for school improvement and minimize disruptions for school districts.
9. Continue to involve teachers in statewide initiatives to improve student learning i.e., test development, revision of standards, etc.
10. Encourage a balanced assessment system that includes classroom-based assessments, Nebraska Statewide Accountability Tests, and norm-referenced tests.
11. Ensure that every school has an appropriate curriculum based on state standards to create success for all students within our state.
12. Designate a person(s) that coordinates curriculum, instruction and assessment for continuous improvement in each district within the state.



**CHARTING STARS:**  
***Engaging Conversations***  
**Section 3: Research Studies**

**Study I: Nebraska-led Portfolio Peer  
Review Process**



**Nebraska-led Math Portfolio Peer Review Process**

**Jody Isernhagen, Ed.D., Associate Professor, University of Nebraska-Lincoln**  
**Jackie Florendo, M.Ed., Graduate Research Assistant, University of Nebraska-Lincoln**  
**Casey Tallent, M.A., Graduate Research Assistant, University of Nebraska-Lincoln**

As educators moved into a second year of the portfolio peer review process, they soon came to realize the level of confidence they now possessed because of the prior year's experience. A female rural district administrator compared last year's peer review process to this year's math portfolio peer review process by stating,

*The first one, when we did language arts, was certainly scary going in because it was different from what we had done. I had been the person who had put together the written portfolios the previous years and submitted them . . . it was a little scary because of the fact that it was going to be a verbal interview. But I can tell you that the people that we had for the first one (language arts) were just a pure delight, and very helpful. So when it was time this year to do the math portfolio, the level of concern was certainly still there because you want to do it right, but the fear factor had been eliminated because you knew that if there were concerns, they would tell you about the concerns and they would help you figure out what problems you might have and how to address those.*

## **INTRODUCTION**

The seventh-year primary study was a mixed-methods research design using both quantitative and qualitative data. The purpose of the study was to examine the Nebraska-led Math Portfolio Peer Review Process and district educator perceptions of the technical quality of their district assessments according to the six quality assessment criteria (Plake & Impara, 2000):

- The assessments reflect the state/local standards.
- Students have the opportunity to learn.
- The assessments are free of bias and insensitive situations.
- The assessments are at the appropriate level.
- The assessments are reliably scored.
- The assessment mastery levels are appropriately set.

The portfolio review consisted of teams of two state-trained assessment experts that visited each school district within the state. The assessment reviewers read the previous district portfolio and the recommended changes for improvement from past years. The visiting team then went to their assigned district, reviewed the evidence of assessment quality using the six assessment quality indicators, and provided formative feedback to the district. Two external assessment experts located in each of the regional areas assisted the peer review teams by answering questions. A summative rating will appear in the State of the Schools Report in 2008.

## **RESEARCH DESIGN**

This mixed-methods research study focused on the Nebraska-led Math Assessment Portfolio Review Process. The K-12 District Internal Math Portfolio Review Team members were surveyed prior to the Portfolio External Review. For the purpose of this research, Nebraska public school districts were divided into two classifications, non-rural and rural, based on population characteristics unique to Nebraska. Non-rural districts were defined as metro-area districts in large and mid-sized cities, large towns, and the urban fringe. All other districts are classified as rural. Of the 254 Nebraska public school districts in 2007-08, 4.3% were non-rural and 95.7% were rural.

Each school district was asked to distribute up to ten surveys to internal review team members. Of the 254 Nebraska school districts in 2008, 750 educators from 189 districts (74.4% of districts) returned surveys for the Comprehensive Evaluation Research Study. Fifty-one surveys were disallowed because they were returned after the completion of the District Portfolio Review including all surveys from one district reducing the number of participating districts to 188 total districts. Of the 188 districts responding to the survey within the prescribed timeframe, 4% were non-rural and 96% were rural.

Participants responded to the 47-item survey (Appendix C) using a five-point Likert scale for each item, with “1” representing “none of the time,” “2” “very little of the time,” “3” “some of the time,” “4” “most of the time,” “5” “all of the time.” The survey was structured to explore six themes: Alignment, Sufficiency, Clarity, Appropriateness, Scoring Procedures, and Summarizing the Review Process.

Secondly, open-ended interviews were conducted in two districts from each of four geographical areas. These four geographical areas were identified by the Nebraska Department of Education (NDE) for selection of reviewers trained for the portfolio review process. Detailed views were collected about the Nebraska-led Math Portfolio Peer Review Process in the sample districts. The interview protocol for the District Internal Math Portfolio Review Team (Appendix D) was used to gather qualitative data. Eight sample districts were purposefully selected based on geographical area, district class, and district free and reduced lunch rate. Thirty-eight individual interviews were conducted statewide during the 2007-08 school year. Two additional themes—Pilot Integrated Visit Review Process and New Learnings—emerged from the qualitative interviews.

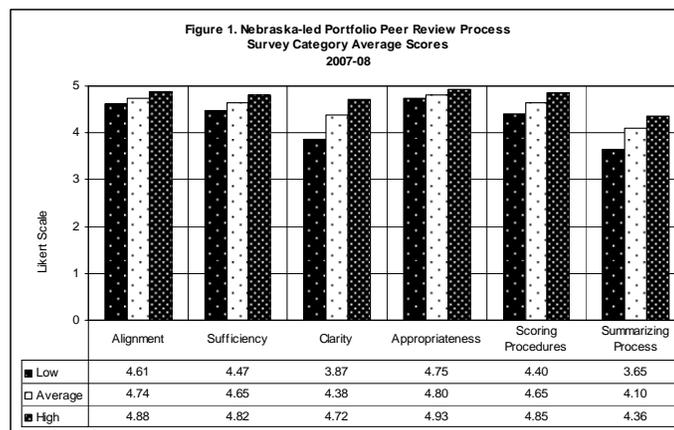
### Instruments

The STARS survey (Appendix C) was designed by the researchers to collect perceptions about the Nebraska-led Math Portfolio Review Process and the six quality criteria. The survey examined the areas of (1) Alignment, (2) Sufficiency, (3) Clarity, (4) Appropriateness, (5) Scoring Procedures, and (6) Summarizing the Review Process. Participants responded to the 47-item survey on a five-point Likert scale for each item, with “1” representing “none of the time,” “2” “very little of the time,” “3” “some of the time,” “4” “most of the time,” and “5” “all of the time.” Analysis of variance was used to compare mean scores of the survey data.

The STARS Research Interview Protocol (Appendix D) consisted of demographic information about participants and ten questions for the selected members of the Internal Portfolio Review Team. These questions targeted the participants’ perceptions of each of the six quality criteria, their preparation and initial thoughts of the review process, and any new learnings based on the process. Probes were identified for interviewers to use with each question. Interviewers were provided a STARS Interview Manual and received training to conduct the interviews.

## RESULTS

**Six categories were identified from the Nebraska-led Portfolio Review Process Survey conducted prior to the visit of the Portfolio Review Team experts. They were rated by participants on a “1” to “5” Likert scale with “5” being the highest. Noted in Figure 1 is the mean, the lowest, and the highest score per category.**



### **Theme 1: Alignment (Survey Questions 1-8)**

**In the area of Alignment for all respondents, responses ranged from 4.61 to 4.88 with an average of 4.74 on a five-point Likert scale with “1” representing “none of the time” and “5” representing “all of the time.” In the area of Alignment, female teachers rated the items significantly higher than the male teachers (p= .018).**

The strongest item rated by all respondents within the Alignment category was “our district involved staff in the alignment of the assessments to the standards” (4.88). Respondents also indicated that “there is a documentation process for alignment of assessments to standards” (4.80) and “districts’ support teachers working collaboratively to ensure assessments measure the standards” (4.79).

In the area of Alignment for all respondents, the lowest rated item was “our district had assessment items reviewed by external personnel” (4.61).

The survey reliability statistic (Cronbach’s Alpha) for the Alignment section of the survey was 0.777.

#### *Alignment Discussion*

**The strongest perception from all respondents indicated that “our district involved staff in the alignment of the assessments to the standards” (4.88). Middle school principals rated this item higher than non-middle school principals (p=.039).**

- A female rural elementary teacher stated how staff were selected for aligning standards, “Initially we wanted to align with federal standards and state standards and they had people on teams. They selected people from different grade levels and different types of teachers—like the amount of experience—because you get a lot of interesting insights.”
- A female rural middle school math teacher shared how their standards team developed knowledge about alignment, “As a team, we worked once a week. I think just all of our backgrounds with being a small district and really having our hands dirty in it, the whole process from evolution in ’98-’99 till now, we just have that common knowledge.”

**Another item rated high by respondents indicated that “there is a documentation process for alignment of assessments to standards” (4.80).**

- A middle school rural math teacher indicated that they had a documentation process by stating, “We just took the standards and then at each grade level we went through what we taught and . . . it definitely evolved. We started with a big document where we listed all of our objectives by grade level under the standard that we felt that they met and then from that we started creating tests.”

**A third item rated high by survey participants indicated that “districts support teachers working collaboratively to ensure assessments measure the standards” (4.79).**

- A female rural high school math teacher shared, “The administration really supported us and gave us about one hour per week for our assessment time to get together with the coordinators of our district and work on our portfolio.”
- A male rural elementary teacher spoke of the time provided by the district stating, “We have been given time to go to the ESU. If we went in the summer, we were

compensated for that time. My colleague and I also had time within the school year and the school day to work on assessments. We were given professional leave and we had a sub provided by the district.”

**Principals rated the question, “our district assessment items/tasks reflect a match to the appropriate standards” higher than teachers (p=.030). High school principals also rated the item higher than high school teachers (p=.032). Middle school principals rated the item “our district involved staff in the alignment of the assessments to standards” significantly higher than elementary and high school principals (p=.039).**

- A curriculum director shared how all staff were involved in the alignment process by stating, “Everybody is going to have to be on board. We even pull in our ELL teachers and our resource teachers to help with the standards. They bring different insight. It was frustrating when . . . we were reworking a test, but it does pay off.”

**Elementary school teachers rated the item, “our district has a list of specifications mapping the assessment items to the standards in order to show which items assess which standards” higher than elementary school principals (p=.033).** A rural superintendent shared how they align standards, curriculum and assessments, “They meet and look at those assessments and continue to talk about . . . how it aligns with the new books. They’re also looking at our other assessments, and not just the online assessments, so I would say that it’s stronger in reading simply because we’ve been doing it longer. But the opportunity for them to meet monthly and have those K-12 conversations is invaluable to us.”

**All females completing the survey rated the question, “our district assessment items/tasks reflect the content and skills found within the standards” higher than their male counterparts (p=.008). Female teachers rated the question, “our district assessment items/tasks reflect the content and skills found within the standards” higher than male teachers (p=.048).**

- A female assessment coordinator shared how the teachers in her school worked on alignment, “The math subject area committee really went through them with a fine-tooth comb and matched them to our high achievement outcomes in our curriculum and also compared them to the assessments that we were currently giving.”

**All females surveyed rated the question, “our district had assessment items reviewed by district personnel” higher than males (p=.036). Additionally, female teachers rated this question higher than male teachers (p=.044). Principals also rated this item higher than teachers (p=.046). Additionally, high school principals rated the item significantly higher than high school teachers (p=.045). Women working in middle schools rated the item significantly higher than men working in middle schools (p=.040).**

- A female curriculum director shared, “I just think those cross grade level teams are critical to that along with looking at different samples of materials at different grade levels. Seeing what those question items look like and continually looking at balance for our grade levels so that you’re balanced within a grade level but you’re also balanced across grade levels and you’re assuring that the rigor is there for students on the high end as well as giving students opportunities to assess at the low end. I just think the only way to do that is to cross grade levels.”

**In the area of Alignment for all respondents, the lowest rated area was “our district had assessment items reviewed by external personnel” (4.61). Females teachers rated this question higher than male teachers (p=.026).**

- A female district administrator shared, “We’re fortunate in that we are big enough that we’ve been able to pretty much do this within our district. Like I said, early on we did some things with the ESU, and it was good. I mean, it’s a good foundation, good learning but we did reach a point in each of the areas that we realized the other school districts at the ESU are all considerably smaller than we are, so they approach things a little differently, not wrong, just simply differently.”

**Principals rated the item, “our district supports teachers working collaboratively to ensure assessments measure the standards” higher than teachers (p=.002). Elementary school principals rated this item higher than elementary school teachers (p=.033). High school principals also rated the item higher than high school teachers (p=.013). Women working in middle schools rated the item higher than men working in middle schools (p=.048).**

- A female rural curriculum director indicated that teachers are working collaboratively to make revisions and check for alignment by stating, “Principals aren’t heavily involved, it’s mainly been teacher groupings put together across grade levels groups that have gone through revisions looking at the existing curriculum and checking for alignment.”
- A rural superintendent shared, “I think knowing that they were going to do the portfolio from the beginning helped and it was ongoing training and knowing that they could ask and get assistance, but I think they felt like they were supported. They also were given time during the day to work.”

### *Alignment Recommendations*

**During the interviews, participants were asked if they had recommendations to make to other school districts regarding each of the survey category areas. Alignment recommendations included:**

- A female rural assessment coordinator shared alignment recommendations by stating, “I think you have to have a really good curriculum. I think that’s the place you start with the curriculum piece and the standards are within that curriculum. Our math team set up high achievement outcomes and then embedded the standards into those high achievement outcomes and had them aligned at each grade level so there’s a scope and sequence. I think that’s the key. But it starts with a good solid curricular foundation.”
- A female rural high school math teacher shared, “It’s an ongoing process that needs to be done continually every year to make sure that that curriculum is matched up.”
- A female rural middle school math teacher shared, “Get the input of your teachers because they are the ones that have to give the test. If they don’t understand, it’s not going to work.”
- A rural curriculum director recommended, “I think you’ve got to have cross grade level groups that can talk about how things flow from grade level to grade level. You’ve got to get those out there where everybody can see it.”

- A female district assessment coordinator recommended, “Most schools are driven by expert curriculum, and you need to have a very careful team looking at any purchasing of new curriculum to see if it aligns to the standards.”

### **Theme 2: Sufficiency (Survey Questions 9-13)**

**In the area of Sufficiency for all respondents, responses ranged from 4.47 to 4.82 with an average of 4.65 on a five-point Likert scale with “1” representing “none of the time” and “5” representing “all of the time.” Women responded to the items in the Sufficiency portion of the survey significantly higher than men ( $p=.004$ ). The total area of Sufficiency was not significant for leaders or teachers.**

The strongest item rated by all respondents within the Sufficiency category was “our district reviewed assessment items/tasks for sufficiency results” (4.82). Respondents also indicated that “our assessment items/tasks are distributed across all performance levels” (4.75).

In the area of Sufficiency for all respondents, the lowest rated area was “our assessment items/tasks use a variety of appropriate formats” (4.47).

The survey reliability statistic (Cronbach’s Alpha) for the Sufficiency section of the survey was 0.790.

### *Sufficiency Discussion*

**The strongest perception from all respondents indicated that “our district reviewed assessment items/tasks for sufficiency results” (4.82). Females also rated this item higher than males ( $p=.004$ ).**

- A female rural high school math teacher explained, “At first we were thinking at the very, very top level of expectancy for a student. Therefore, we wrote questions that were extremely difficult. Our test was tough. But that’s the way we were looking at it at the time. In retrospect, we really needed to think about all of the levels of beginning and progressing so we could see where kids were. It wasn’t fair to assume that everybody was going to be at this advanced level.”

**Females rated the item, “our district measures all academic content standards in the assessment items/tasks” higher than males ( $p=.007$ ). Female principals rated this item higher than male principals ( $p=.035$ ).**

- A female rural elementary math teacher emphasized when examining proficiency levels, “They sat down and actually looked at what does beginning, proficient, progressing, and advanced look like. When you get to math it gets so hard ‘What makes it advanced?’ I know the process they went through at different grade levels. They were all people that had already dealt with the standards and so they were pretty familiar with how the kids had done on the assessments.”

**Respondents also indicated that “our assessment items/tasks are distributed across all performance levels” (4.75).**

- A male rural middle school math teacher shared how they distributed items across performance levels, “We went through descriptors and were told what a question would look like that was a more advanced question. Sometimes we didn’t always agree. But the curriculum director would either come in or clarify it for us or we

would talk between sixth, seventh, and eighth grade teachers and decide why it would be considered a more difficult question or not such a difficult question.”

- A female rural curriculum director shared how they developed performance level descriptors (PLDs), “Teachers had developed very complete PLDs at the ESU and those were extremely helpful to us. They actually looked at those PLDs and tried to make sure that they had assessment items at every level.”

**Females rated the question, “our assessment items/tasks include higher order thinking skills” higher than males (p=.022). Females teachers rated the question, “our district had assessment items reviewed by district personnel” higher than male teachers (p=.033).**

- A female rural assessment coordinator shared how PLDs were used to examine the level of the question, “They look at how the students actually responded to the question. If they thought it was a higher level question but all of the students that responded got it correct, then they had to go back and say ‘well, I don’t think this is a higher level question. Let’s look at this again or rewrite it in such a way that it is.’”

### *Sufficiency Recommendations*

**During the interviews, participants were asked if they had recommendations to make to other school districts regarding each of the survey category areas. Sufficiency recommendations included:**

- A male rural elementary teacher recommended, “I recommend that they work as a whole team at a grade level. One person or two people cannot do it all. It really helps to have everyone involved.”
- A female rural assessment coordinator shared, “Don’t overdo it. The belief that more is better is not true because the system has to be able to sustain itself and if more is better, that’s not always possible.”
- A female rural assessment coordinator urged, “They have to have really strong PLDs. That’s the first thing we looked at was the PLDs. Making sure that there’s agreement on those and making sure they’re strong statements.”
- A rural superintendent suggested, “I think small districts absolutely have to work cooperatively and work in consortiums. I don’t think you can go it alone.”
- A female rural assessment coordinator emphasized, “Don’t do it alone. I would say you need to work with other districts and see what they’ve done. I think you need to have feedback from what is happening across the state.”

### **Theme 3: Clarity (Survey Questions 14-21)**

**In the area of Clarity for all respondents, responses ranged from 3.87 to 4.72 with an average of 4.38 on a five-point Likert scale with “1” representing “none of the time” and “5” representing “all of the time.” In the area of Clarity, elementary school teachers rated the items significantly higher than non-elementary school teachers (p=.032).**

The strongest item rated by all respondents within the Clarity category was “our assessment directions for teachers are standardized across the district” (4.72). Respondents also indicated that “our assessment directions for students are standardized across the district” (4.69), “our

assessment directions for students are clear” (4.68), and “our assessment directions for teachers are clear” (4.68).

In the area of Clarity for all respondents, the lowest rated area was “our district provides parents with reports that give an appropriate explanation of assessments results” (3.87). Additionally, respondents also rated this item low, “our district sends individual reports each school year” (3.93).

The survey reliability statistic (Cronbach’s Alpha) for the Clarity section of the survey was 0.795.

### *Clarity Discussion*

**The strongest perception from all respondents indicated that “our assessment directions for teachers are standardized across the district” (4.72).**

- A female rural elementary teacher shared, “Teacher wise, directions come from the curriculum director’s office. It’s a big sheet of paper and everyone gets the same one and we read through it. Typically it’s the same thing every year, but it’s good to refresh what it is we’re doing and how to do the assessment. If there’s anything specific for your grade level, it’s usually attached to the back of it. Everyone gets the same thing.”
- A female rural assessment coordinator, “We wrote assessment protocols for the district for each type of assessment that we give. The teachers have a document that says when you give an online assessment, here’s what you do. Here are the accommodations, here are the instructions, and here are your responsibilities as a teacher. So they have that up front as an umbrella document. But each time they give the assessment, the directions are there for them.”

**Respondents also indicated strongly that “our assessment directions for students are standardized across the district” (4.69), “our assessment directions for students are clear” (4.68), and our assessment directions for teachers are clear”(4.68). High school teachers rated this item “our assessment directions for students are standardized across the district” significantly lower than non-high school teachers (p=.022).**

- A female rural middle school math teacher shared, “We have in our testing standard’s booklet a page where directions are clearly typed out. We wrote the directions for that page and we also include the point value, whether they can use scratch paper, calculators, and technology.”
- A female rural high school math teacher stated, “We’re the ones that give the tests. So the students know what our rules are every time. They just don’t feel threatened going into a new environment. They know everything is constant and consistent.”
- A rural high school math teacher shared about writing directions for students and teachers, “Even writing out the most mundane tasks is important to specify what the students are to do. Are they to answer with a label? It just needs to be precisely written so that both the teacher and the student understand.”

**For all respondents, the lowest rated area was “our district provides parents with reports that give an appropriate explanation of assessments results” (3.87). Elementary school teachers rated it significantly higher than non-elementary school teachers (p=.039). Elementary school teachers rated the item “all district/school reports are**

**appropriately disaggregated” significantly higher than non-elementary school teachers (p=.005).**

- A female rural elementary teacher shared, “There are at least two reports, one that states the standard and what each student scores. I have another form that I give them that shows all of the tests that they are responsible for that I call a summary sheet. At the end of the school year, I have a letter that I send home to the parents about the standards.”
- A middle school math teacher shared, “At the end of the year we send out a parent report that shows the performance level of each standard and each subject.”
- A high school math teacher shared, “We really haven’t done a good job on the parents’ part yet. I usually try to tell the students as soon as I grade them, how many problems they got right.”

**Additionally, respondents also rated this item low, “our district sends individual reports each school year” (3.93).**

- An assessment coordinator shared, “The one thing we’re doing this year is a standards-based elementary report card. We are moving into creating a report that’s more standards based for our middle school and high school.”

#### *Clarity Recommendations*

**During the interviews, participants were asked if they had recommendations to other school districts regarding each of the survey category areas. Clarity recommendations included:**

- A male rural middle school math teacher recommended, “They need to make sure, especially when they’re doing new teacher training that they don’t take it for granted (that new teachers understand assessment instructions).”
- A female rural elementary teacher emphasized the importance of individual conferences with kids to talk about assessment results, “I think having the individual conferences with kids really help. Kids then know where they stand.”
- A female rural assessment coordinator advised, “I would just say that you need to have a standard procedure established district-wide. Whoever is administering the tests needs to be trained in that standard procedure so that every testing environment is equal and every testing environment is the best for the kid.”

#### **Theme 4: Appropriateness (Survey Questions 22-28)**

**In the area of Appropriateness for all respondents, responses ranged from 4.75 to 4.93 with an average of 4.80 on the Likert scale with “1” representing “none of the time” and “5” representing “all of the time.”**

The strongest item rated by all respondents within the Appropriateness category was “our assessments were screened for fairness, bias, and sensitivity” (4.93). Respondents also indicated that “our assessments were reviewed by internal or external groups” (4.84).

In the area of Appropriateness for all respondents, the lowest rated areas were “our assessments are appropriate for the assessed grade level” (4.75) and “our assessments indicate our expectations for our students” (4.75).

The survey reliability statistic (Cronbach's Alpha) for the Appropriateness section of the survey was 0.858.

### *Appropriateness Discussion*

**The strongest perception from all respondents indicated “our assessments were screened for fairness, bias, and sensitivity” (4.93). High school principals rated the item “our assessments were screened for fairness, bias, and sensitivity” significantly higher than high school teachers (p=.040).**

- A male rural high school principal shared, “Bias in language, bias in vocabulary. Math is not as critical except when it is a story problem but we read through every question to make sure.”
- A rural superintendent shared, “Teachers when giving the tests keep notes and bring them back to the assessment team. You have some kids that have never actually been in a typical kind of place . . . that ever present acknowledgement of poverty. I have at least six children that are being raised in nontraditional settings with grandparents or guardians where even the mere mention of mom or dad can be a bias.”
- A female rural high school math teacher shared, “Every summer we go through our questions and we look for bias. We have found out that some school districts don't think certain questions are biased but then they may have a certain student move in and that question then becomes biased.”

**Respondents also indicated that “our assessments were reviewed by internal or external groups” (4.84). Females working at the elementary level rated this item significantly higher than male counterparts (p=.028).**

- A female rural elementary teacher shared, “A lot of times we have other people review it because we may miss it. Teachers that give the assessment, we have their input as well. We had ESU come in and check them out.”
- A male rural high school teacher, “You always worry about the bias and we had other people checking to make sure. Then you had the peer review do that as well. So, we were cognizant of the fact that we had to be careful about that.”

**In the area of Appropriateness for all respondents, the lowest rated items were “our assessments are appropriate for the assessed grade level” (4.75) and “our assessments indicate our expectations for our students” (4.75). Females working in elementary schools rated the item “our assessments indicate our expectations for our students” higher than males working in elementary schools (p=.046).**

- A male rural high school principal explained, “They've been trained in new techniques which I think is helpful and what we're doing is meaningful. It's a challenge, but it's going to show up for students on their national assessments and tests as they go forward in education. Education will be easier, not because they were successful at what they did, but because they were challenged in what they did. I think it is a struggle, but it's been a good one.”
- A female rural middle school math teacher shared how they ensured appropriateness across all grade levels, “We do have the luxury of having three grade levels in the building. As a math department we sat down and checked to see where things are taught, reinforced, instructed and mastered. That was very valuable because our 8th grade teachers are saying, well if it's taught in 6th grade and reinforced in 7th grade,

what needs to be mastered in 8th grade. Our curriculum is pretty much aligned to our standards.”

- A rural superintendent shared, “Sometimes, the content is not developmentally appropriate, so you’re challenged by the fact that even though the kid could read it, the way the question is written on the assessment itself . . . we find that it’s just not age appropriate. Is it really a true measure of what the kid is able to know and be able to do?”
- A male rural high school math teacher addressed the appropriateness and expectations of assessments for special education and ELL students, “We give the student the option of having the test read to them, especially special education students. We have paraprofessionals who can translate if there’s something the student doesn’t quite fully understand in English.”

**Elementary teachers rated the item “our assessments demonstrate an increase of expectation from one grade level to the next” significantly higher than middle & high school teachers combined (p=.027). In addition, middle school teachers rated this item higher than other elementary and high school teachers combined (p= .021).**

- An elementary teacher shared, “We did a lot . . . in our curriculum groups. We had the standards and then we would work our way backwards so then in third grade you need to be to this point, and in second grade, you need to be to this point. We did that together.”
- A male rural superintendent shared, “There’s been a conversation and a review between the different grade levels to see that they’re building upon one another. That’s happened both in house and it’s been also somewhat orchestrated by our ESU, too, with different schools coming together at different grade levels.”
- A rural curriculum director stated, “I just think those cross grade level teams are critical to looking at different grade levels and seeing what those question items look like. You’re assuring that the rigor is there for students. I just think the only way to do that is to cross grade levels.”

#### *Appropriateness Recommendations*

**During the interviews participants were asked if they had recommendations to make to other school districts regarding each of the survey category areas. Appropriateness recommendations included:**

- A female elementary teacher offered, “I think, the more eyes that can see the test and go through it is probably the best. There were five of us on the committee, but there were things that got by us that we didn’t see.”
- A female rural assessment coordinator emphasized, “It’s important to provide leadership and it’s important to provide time. In-service is the time to do it. Summertime is a wonderful time to do it. It’s surprising the groups of people who might be available that end up working together. It’s a much more relaxed atmosphere and it’s just positive for the school.”
- A rural curriculum director suggested, “You have to look at the reliability of your results and see if it’s time to raise the bar or do we have this too high. Balance is really important.”

- A female rural middle school math teacher, “I guess if they were a smaller district, maybe pair up with another district and trade. I would say use other districts in other schools to get a different perspective.”
- A female middle school math teacher advised, “Sitting down together, looking at the 6th and 7th grade, when it was taught, when it was tested, when it was mastered, at what level and so forth and carry that on to the high school. The 8th and 9th worked together well. Really hit those transitions.”
- A rural curriculum director, “When you’re writing assessments, you always want to make sure you have special education teachers involved. You want to have ELL teachers and various grade levels.”

### **Theme 5: Scoring Procedures (Survey Questions 29-42)**

**In the area of Scoring Procedures for all respondents, responses ranged from 4.40 to 4.85 with an average of 4.65 on the Likert scale with “1” representing “none of the time” and “5” representing “all of the time.”**

The strongest item rated by all respondents within the Scoring Procedures category was “our participation rates are documented” (4.85). Respondents also indicated that “our assessments have established scoring guidelines and directions” (4.79) and “our district has local assessment policies in place to assure comparability and consistency across the district” (4.76).

In the area of Scoring Procedures for all respondents, the lowest rated areas were “our students are given instruction about behavioral objectives during the assessments” (4.40) and “our district provides training for those administering the assessments” (4.48).

The survey reliability statistic (Cronbach’s Alpha) for the Scoring Procedures section of the survey was 0.883.

### *Scoring Procedures Discussion*

**The strongest perception from all respondents indicated that “our participation rates are documented” (4.85). Respondents also indicated that “our assessments have established scoring guidelines and directions” (4.79). Elementary school principals rated the item “our assessments have established scoring guidelines and directions” higher than non-elementary principals (p=.044). Additionally, middle school principals rated this item higher than non-middle school principals (p=.050).**

- A female rural high school math teacher disclosed the need for a scoring guide, “A scoring guide, because that’s another issue for teachers. If different people are scoring, you are doing the same thing.”
- A female rural assessment coordinator shared, “Our scoring procedures are being standardized by having strong keys, having no blanks in those keys. I think that the administrative guides also help on the scoring procedures.”

**Elementary female educators rated the item “our performance level descriptors are clear and specific for each assessment” significantly higher than males working in elementary schools (p=.031). Elementary school principals rated this same item significantly higher than elementary school teachers (p=.020). Elementary teachers rated the item “our district consistently applies performance level descriptors to the cut**

**scores for each assessment or standard” significantly higher than elementary principals (p=.037). Middle school teachers rated this item higher than non-middle school teachers (p=.003).**

- A female assessment coordinator revealed, “We need to rewrite our performance level descriptors, mostly for clarity. They’re pretty vague and pretty broad right now. They could be more descriptive.”
- A female rural high school math teacher shared, “We will look at the cut score process, not only the consortium but we will look at our individual school results to see what we feel we need to change in the cut score area when we’re doing the first round of putting in our Angoff methods.”
- A female rural high school math teacher imparted, “We make sure that after we change them, we still have the right amount of questions at the proficiency level . . . and due to our changes, then we have to change our cut scores also. We go through the Angoff method with our cut scores. After we have those scores in, then we look at the consortium cut scores.”

**Survey participants also rated this item high, “our district has local assessment policies in place to assure comparability and consistency across the district” (4.76).**

- A female rural assessment coordinator shared, “We use the Angoff method. That gives us our level of difficulty. You can take that information and use it to really look at your cuts and make some very strong decisions about whether they’re appropriate keeping in mind that the better you teach something, even if it’s a difficult concept, more kids are going to get it right.”
- A rural superintendent advised, “When you have small group sizes . . . the consortium method, I think is the only way to go. The ability to dialogue with other people outside, then rely on those people. (This) can be very good.”
- A female rural elementary math teacher offered, “The assessment team decided to have our teachers do DCMs. It’s called Decision Consistency Model. It is a list of students and the teacher predicts on this form how each student is going to perform. We do this as a district.”
- A female elementary teacher emphasized, “We all get the same exact test. For those students that don’t speak any English, it is translated into Spanish for them. But otherwise, it’s the same test for all fifth graders even ELL and SPED.”
- A female district administrator indicated that they use the KR21 method, “I know there were other methods but I will tell you from the very beginning, because we’re large enough and we have a large number of students at every grade level we made everything very objective. We just simply used the KR21. It was black and white. Then I could go back to the teachers and say, ‘This assessment didn’t meet the reliability.’”

**In the area of Scoring Procedures for all respondents, the lowest rated items were “our students are given instruction about behavioral objectives during the assessments” (4.40) and “our district provides training for those administering the assessments” (4.48). Elementary teachers rated the item “our students are given instruction about behavioral objectives during the assessments” significantly higher than elementary principals (p=.021). Principals rated the item “our district provides training for those administering the assessments” higher than teachers (p=.043).**

- A rural curriculum director indicated that, “As far as the administration of the assessments, we met with grade level teams at the beginning of the year and went through that whole process and what’s expected. We’re developing a kind of quick reference guide to the assessments and things to remember, a check list to be distributed.”
- A female rural assessment coordinator shared how their school has an assessment team administer all assessments, “Well, our assessment teams are in charge of the assessment process. We decided to go that way because what we were finding is that if teachers were in charge of the process, even if you had directions on how to handle the process, they handled it differently which we thought skewed the validity of the test.”

### *Scoring Procedures Recommendations*

**During the interviews participants were asked if they had recommendations to make to other school districts regarding each of the survey category areas. Scoring Procedures recommendations included:**

- A male rural high school math teacher indicated, “Just re-evaluate every year, see where you’re at. Sometimes it can be just the clientele you’re dealing with that year. We all know that some years you have classes that are fantastic students and some years you have classes that that’s not necessarily the case.”
- A rural superintendent recommended, “I think that if they watch how that process works with scoring and double scoring, those are invaluable opportunities for staff. I think those are healthy opportunities. I would encourage that double scoring opportunity so people understand how that works.”
- A female rural elementary teacher advised, “I think putting a table team together to do it is helpful and we have other people to bounce things off of. I think the peer review is really good because when you write it, it looks really good to you but when someone else looks at it, it may not be so clear. So I think . . . peer editing is important.”
- A female rural middle school teacher recommended, “I just think they have to make sure when they get those results back that they don’t become shelf material. That information is important to you as a teacher and to your kids. Now they realize, I’m supposed to have all these areas at the top that are mastered. If I have things down at the bottom that say I didn’t master them, then we’ve got work to do.”
- A female rural high school math teacher shared, “First of all, in reliability, they need to make sure that they’re a big enough school district that they can do the reliability on their own. If not, they need to make sure that they find somebody that they can go together to make the sample size large enough that the reliability is realistic.”
- A female rural elementary teacher advised about the value of having assessment facilitators, “I like the consistency it brings to kids so that when they go in and test, they know it’s going to be the same way every time.”

**Theme 6: Summarizing the Review Process (Survey Questions 43-47)**  
**In the area of Summarizing the Review Process for all respondents, responses ranged from 3.65 to 4.36 with an average of 4.10 on the Likert scale with “1” representing “none of the time” and “5” representing “all of the time.” In the area of Summarizing**

**the Review Process, principals rated the items significantly higher than teachers (p=.010). High school principals also rated the Summarizing the Review Process items higher than high school teachers (p=.043).**

The strongest item rated by all respondents within the Summarizing the Review Process category was “I have the necessary information to prepare the district assessment portfolio” (4.36). Respondents also indicated that “I feel prepared to present my district portfolio to my peer reviewers” (4.32) and “I have had adequate help in preparing the district assessment portfolio” (4.26).

In the area of Summarizing the Review Process for all respondents, the lowest rated areas were “Compensation is provided to prepare the district assessment portfolio when completed outside of the regular school day” (3.65) and “I was provided time within the teaching day to prepare the district assessment portfolio” (3.85).

The survey reliability statistic (Cronbach’s Alpha) for the Summarizing the Review Process section of the survey was 0.801.

#### *Summarizing the Review Process Discussion*

**The strongest perception from all respondents indicated “I have the necessary information to prepare the district assessment portfolio” (4.36). Principals rated the item “I have the necessary information to prepare the district assessment portfolio” significantly higher than teachers (p=.014). High school principals also rated the item higher than high school teachers (p=.006).**

- A female rural high school math teacher shared, “I felt very confident. We thought we were prepared and apparently we were. Things were in order. We didn’t have anything to worry about.”
- A female rural high school math teacher, “Well, my initial thoughts were different from the initial thoughts the first time. I think it’s a good process. I can see the value in it. You are a different person at different stages and the process you go through helps you value it even more. It is a lot of work and it is a process but I think it’s a process that makes people really think about what they’re really doing now more than ever before. The accountability piece of it I like, if you felt there might not have been people before who were too aware of what was going on, they pretty much need to be aware now.”
- A female rural assessment coordinator, “I felt prepared. I would have been really surprised if we hadn’t received the ratings that we did.”
- A female assessment coordinator compared the reading and math portfolio processes, “Well, having gone through it with reading, I had two feelings about it. One, I was familiar with the process so it felt comfortable to me. On the other hand I felt like *Oh No*, this is so much work to go through. But, not that it was easier with joining a consortium, but we could pool the resources and then be better prepared. Our teachers had done a lot of the work on district assessment, so they were familiar with the quality criteria and the processes that we take each assessment through to be of high quality. So it wasn’t like we were getting out of work. But, ah, I think it’s a good process for all of us to go through. I think our assessment literacy is much higher than it was prior to going through this process.”

Respondents also indicated “I feel prepared to present my district portfolio to my peer reviewers” (4.32) and “I have had adequate help in preparing the district assessment portfolio” (4.26). Principals rated the item “I feel prepared to present my district portfolio to my peer reviewers” significantly higher than teachers ( $p < .001$ ).

Additionally, females rated the item significantly higher than their male counterparts ( $p = .003$ ). High school principals, in particular, rated the item higher than high school teachers ( $p = .004$ ). Middle school principals rated the item significantly higher than middle school teachers ( $p = .048$ ). High school teachers rated the item lower than the non-high school teachers ( $p = .048$ ).

- A female rural high school math teacher indicated a level of confidence with the process, “I felt very confident especially after we heard our results. We thought we were prepared and, apparently we were. Things were in order. We didn’t have anything to worry about.”
- A female rural middle school math teacher shared, “I actually enjoyed it quite a bit. I thought the two reviewers made us feel very comfortable. They were very clear on how the day would move and what they would do and what we would do. I enjoyed it quite a bit.”
- A male rural elementary teacher shared the support he has experienced, “We have been given time to go to the ESU. If we went in the summer, we were compensated for that time. My colleague and I also had time within the school year and the school day to work on assessments. We were given professional leave and we had a sub provided by the district. Since then, we’ve also had time with our own grade level staff to assess and work on the process.”

In the area of Summarizing the Review Process for all respondents, the lowest rated areas were “Compensation is provided to prepare the district assessment portfolio when completed outside of the regular school day” (3.65) and “I was provided time within the teaching day to prepare the district assessment portfolio” (3.85). Principals rated the item “I was provided time within the teaching day to prepare the district assessment portfolio” significantly higher than teachers ( $p < .001$ ). High school principals also rated the item higher than high school teachers ( $p = .003$ ). Additionally, middle school principals rated the item significantly higher than middle school teachers ( $p = .009$ ).

- A female rural high school math teacher indicated, “The administration really supported us and gave us time . . . about one hour per week for our assessment time to get together with the four coordinators of our district and work on our portfolio, plus we had to work on it outside at non contract time. I had been through the process where the other three had not, but . . . having that experience from the reading portfolio really helped with the math portfolio this year.”

Elementary school principals rated the item “our performance level descriptors are clear and specific for each assessment” significantly higher than elementary school teachers ( $p = .020$ ).

- A female rural assessment coordinator shared how PLDs (performance level descriptors) were used to examine the level of the question, “They look at how the students actually responded to the question. If they thought it was a higher level question but all of the students that responded got it correct, then they had to go back

and say ‘well, I don’t think this is a higher level question. Let’s look at this again or rewrite it in such a way that it is.’”

### *Summarizing the Review Process Recommendations*

**During the interviews, participants were asked if they had recommendations to make to other districts regarding each of the survey category areas. Overall, the participants were complimentary of the process and offered the following recommendations to enhance the review process.**

#### **Keep up Bias Training**

- A female rural assessment coordinator emphasized the importance of bias training, “I just would like to say you need to keep up your training on bias. If you go to the state department trainings or you go to their workshops every once in awhile they say something that makes me go oh, yeah, I can see that.”

#### **Teachers as Peer Reviewers**

- A female rural middle school math teacher emphasized the need for teachers to serve as peer reviewers, “I would have the teachers - they’re the ones in the trenches - so they should be ones offering the suggestions.”

#### **Time Required**

- A female rural district administrator commented on the time required for quality assessment, “I guess I would tend to bet that no matter what district you’re talking to, if there’s an improvement that can be made, it’s to have more time to be able to bring people together and not to be taking away from their time to do what they normally do. It’s always been a time issue.”

#### **Same Reviewers Every Year**

- A female rural assessment coordinator emphasized the need for consistency with peer review, “If the same people could come back to your school year after year, you have a better ability to review what improvements have been made. I’ve been on several external evaluation teams and when I’m able to visit the same school I think it’s the most powerful tool to help schools improve. Consistency in reviewers will help us to have better accountability. I think there needs to be a follow-up process—something in the middle of the year.”

#### **Getting the Teachers Involved in the Process**

- A female high school math teacher talked about getting teachers involved in the development of state tests, “I’d really like them to realize that the teachers need to be involved in the assessments. I appreciate what the state has done the last two years making the teachers be involved and have a part of the say about assessment.”
- A male rural assessment coordinator echoed the importance of teacher involvement, “Actually, the involvement of teachers are the key to that.”
- A female rural middle school math teacher emphasized the importance of ownership, “They have to get the teachers involved in this because we take ownership of it.”

### **Assessment Team**

- A female rural assessment coordinator shared the importance of working as an assessment team, “We have an assessment team that goes to all of those trainings. I have been involved in some of the training on a one-to-one basis with the ESU, but basically, information brought back by our assessment team that tells us what’s going on, what needs to be done, and what we can do as a school district to help them. In every building an assessment team member kept us more informed and kept us, I think, more involved on a one-to-one basis so the teachers had better ownership.”
- A female rural assessment coordinator stated that having an assessment team helps with professional collaboration, “If you can have data teams and professional learning communities that meet on a regular basis by content area, especially if you’re a larger district, that communication process between teachers and between grade levels is essential. Professional collaboration is important.”
- A rural curriculum director shared the positive experience that the portfolio review process had for their district, “In hearing about it over the years you have your big vision of what’s going to happen to you but it was a very pleasant experience. It was highly professional. We think some of that depends on the people that come to do your review but the people we had did a nice job and one was more knowledgeable than the other and I think that they pair you that way on purpose which works well. I like how they gave everyone an opportunity to share what they knew about the process and how they had been involved and in doing so you pick up some unique pieces of information from different participants and so that’s good. I think it makes their time more valuable, too, when they’re actually asked to participate.”

### **Theme 7: The Pilot Integrated Visit Review Process**

**In 2007-08, the NDE, in an effort to reduce the number of visits a school received in a calendar year, piloted an Integrated Visit Review Process in a limited number of school districts. The integrated visit process had multiple formats for a single visit. Some schools connected the portfolio review process with the school improvement process in a single visit while other districts connected Title I with the Portfolio Review Process in a single visit. Initial feedback from pilot schools was positive and districts elaborated on the success of the integrated visit during the Portfolio Math Review Process qualitative study interviews:**

- A female rural middle school math teacher made the following recommendation regarding an integrated visit in her district, “Discussion was a lot about how can we bring all these groups together and not as single entities, rather . . . have one group and we can have a common goal over all. I sit on school improvement, on assessment, and all those teams. It was very interesting.”
- A rural superintendent expanded on the recommendation for an integrated visit by stating, “It’s very different in terms of an integrated site visit. I think that is important to note because I didn’t know what to expect. When you look at the whole school as an umbrella and you look at that portfolio, I liked the feel of that meeting much more than when NCA people come from the outside. It just . . . accomplished what we really want to think about with school improvement.”
- A female rural elementary teacher further reinforced the integrated concept by stating, “Well, I think the state department is coming for a lot of different things at one time.

The less visits you can do the better. It was kind of stressful for the staff that day because we had all these different teachers that had to leave and go out but at least it was just a day and then it was done. It wasn't quite so hard. It seems like you're constantly being taken out of the classroom. In our district—when you have a small district—everybody does everything. I think the more they can integrate it, the better it will be.”

- A female rural high school math teacher shared that their integrated visit was a positive experience “I would highly recommend trying to give more integrated visits like we received. I feel that you could see the bigger picture when you tie it all together. Again, it makes you grow, as an individual and as a district.”

### **Theme 8: New Learnings**

**Conversations with teachers and leaders emphasized the success of the Portfolio Peer Review Process and provided insights into the growth of educators through the use of STARS. The STARS process, as illustrated in the comments, honors teachers and allows them to participate in conversations about their daily assessment efforts to improve learning for students and educators:**

#### **The Portfolio Peer Review Process**

- A female rural high school math teacher shared, “I think now the system is in place, I don't see the confusion for science and social studies because they were informed from the start of what was expected and people now know the process.”
- A male rural high school teacher shared, “I think through this process, we have more people that do understand what's going on. That's probably one of the good things about it.”
- A rural superintendent compared the reading and math portfolio reviews, “I think that the second time it was much more manageable. We have a tendency perhaps to be a little overly confident in terms of what we think we're doing at times, but I think having a hard copy of it was easier. I think it looked more authentic.”
- A female rural elementary teacher, “After being a part of it for two years, I really was not near as nervous this year. You do really know that they are not coming to criticize you and they're not coming to look at what you're doing wrong. They're there to make suggestions and to help and they're teachers like we are.”
- A female rural assessment coordinator reflected on the process, “There's a team of people that do that and as an administrator, do I really reflect on it? Do I really dig in to it? Well no, not unless something like this comes up and it gives me the opportunity.”

#### **Accountability through the STARS Process**

- A female rural assessment coordinator shared the importance of standards, “When looking at the standards, they gave me a basis of knowing what I needed to teach my kids, and to assure that when my kids walked out of my room, they were getting what they needed just like any other eighth grader across the state. So I felt like my teaching quality improved. I always made sure that whatever I was teaching I also had standards. I also made sure my kids knew it. This is the standard that we're covering right now. I applaud what Nebraska has done.”

- A female rural elementary teacher shared, “It has given me better insight as to where my students are going to take this information. I knew it was important but it just kind of reinforced how important it is so that by the time they have to get up and have to take that assessment they’ll have that knowledge.”

### **Valuing Teachers**

- A female rural elementary teacher shared the importance of being valued, “Just the fact that it’s given me a chance to see that this is not for naught. What we’re doing; it’s being valued. People are taking it seriously. It’s not just being filed away somewhere. There are many capable retired teachers, people that are even still currently teaching that do this process that have a wealth of information to share with us by doing this.”

### **Learning Improvement for Teachers and Students**

- A male rural high school principal emphasized, “I’m not so sure it’s about us being better but I think our kids are better prepared. I hope they’re taking away learning that is long-term, not just short-term. I mean, these are the things that we expect of students when they graduate.”
- A female rural high school math teacher indicated, “I feel both processes have made our teachers become better teachers. They’re more aware of what student’s needs are at all levels. I feel they understand their curriculum better. They have learned new ways to teach or new methods to teach the different levels of children or students. I just think it’s a wonderful process. It’s just made us all become better teachers. It holds you accountable so you understand that you need to keep up on the latest.”
- A female rural assessment coordinator shared, “I feel I have grown a lot, not only in the classroom but outside the classroom by being the assessment coordinator, I have a better understanding of what all of this means, and realized, we have to keep this process going to make us keep getting better, to keep . . . reaching what we need to for our students . . . be able to keep up with all the technology and the way the world’s changing. I mean, globally, we’re taking some very big steps and without this process, I don’t think we would probably keep up as well.”
- A female elementary teacher recognizes that change is constant, “I say we’ve made the perfect assessments. There’s no such thing. I think it benefits our students in the fact that there’s a consistency as far as standards.”

### **Assessment Viewed as Best Practice**

- A male rural high school teacher shared, “The reinforcement at point of instruction was the place to go. And there’s no doubt in my mind it is the way to go. Again, I really feel confident we are finding out what these kids know.”
- A female elementary teacher reflects on best practice, “I think the benefit from this as a teacher is that you have to know exactly what it is these kids need to be taught. You have to find the best way to instruct and the best way to assess it to know, are they getting this? I think that this has been a process that helps us. I think it’s raised an awareness that a lot of people didn’t necessarily have. You never stop learning.”
- A male rural superintendent shared how best practices become a part of classroom instruction, “That was another sort of evolution of how can we make this all part of what we do. Rather than an ‘add on,’ how can we integrate it into what takes place in

the classroom. Now we really want to use these results to direct us to make student gains.”

### **The Importance of Having a Curriculum Director**

- A female rural middle school math teacher, “We have a great curriculum director that keeps us on our toes. In the program that I’m in at UNL right now, I have colleagues from all across the state of Nebraska, all different grade levels, all different school sizes, and they don’t (all) have that (a curriculum director). So that was my ‘ah ha.’”
- A female rural middle school math teacher stated, “Our curriculum director just makes us all know that we may not enjoy some of the tasks. But when it’s all lined up and we see the finished product, we can see that it is a good thing to do. But, I like knowing all the steps. I like knowing why I have to do it. Probably the disaggregating of the data was probably a big ‘ah ha.’”
- A female rural elementary teacher emphasized the role of a curriculum director, “Our curriculum director had us all sit down. He went through what the process was going to be with us just to kind of calm our nerves. It was going to be real intense. He did go through everything with us so we had an idea what was going to happen. Just having the director ask me out of all the other elementary teachers, I felt like he thought I knew what I was doing and felt I knew the process. So, I was honored.”

### **As teacher ownership and involvement increases over time, it helps to grow teacher assessment literacy, confidence, and comfort level with the Portfolio Review Process and STARS overall:**

- A female rural elementary teacher, “I think one of the things it’s done for me is I have gained more confidence.”
- A male rural superintendent indicated growth for educators, “When you talk about proficient, and the cut line for proficient, at first educators think of proficient as being you are fine, you don’t need any more help. That’s not really an accurate definition. It’s a minimum expectation, not something to shoot for. We should be shooting much higher. That was a struggle for all of us.”
- A female rural assessment coordinator shared her learning, “You know, I didn’t really understand and I even probably fought reliability a little bit in the beginning. But I see where they’re coming from and I understand about having a variety of levels on your assessments.”
- A female rural elementary teacher reflected upon her learning, “I think it’s made me more aware of what I needed to do and become accountable myself for what I’m teaching. It’s really focused my goals in what I need to cover and re-teach to make sure my students understand. I don’t think I was maybe doing it as well before STARS came.”
- A rural superintendent, “Our teachers never asked those questions ten years ago. Informally they did. But they couldn’t sit and talk with one another about it. When I have a teacher ask me, why do we do science fair with 1st graders? Is it because we want them to have the opportunity to learn science? Do we teach them? Do we assess it? How do we know? Those are thoughtful questions and the courage to say to me, I’m not sure we did this right. I love that. I love that they will say, we don’t have time for this, and I’m hoping, that we will embrace this as an opportunity, not to

think it was a conflict, but just simply that we have to be accountable and will do our best to let people know how our kids are doing.”

- A female rural elementary teacher, “It just gave me a little more confidence that I could handle something like that (portfolio peer review). It was pretty intimidating at first. It was a good learning experience for me. I really enjoyed it and I learned a lot. It wasn’t as intimidating as I thought it was going to be. It boosted my confidence a little bit about what we were doing is right and we’re doing a good job of it.”
- A female rural elementary teacher shared, “I’m proud of what I have learned. I’m proud of our school. I’m proud of our district, what we’ve done. I am very proud of our assessments that we do.”

### **Assessment Literacy**

- A female rural middle school math teacher indicated, “I think that being involved in the process makes you much more knowledgeable and confident with each of the criteria. If you’re not involved in the actual portfolio, you know those criteria, but they’re not something that you can just spit out.”
- A female rural middle school math teacher pondered, “I think probably my involvement in this process enhanced my passion for assessments. Stiggins sticks in my head ‘it’s never a finished product’. There’s always the latest revision or whatever and teachers don’t always understand it isn’t ever finished because we can always make it better.”
- A female rural district administrator, “We’ve certainly all grown in our assessment literacy. . . . Well, that in and of itself is certainly a good thing. The important thing out of all of this is that as we learn to be better at assessment, we learn to be better at instruction.”
- A male rural high school principal pointed out, “Well, you have to create time for it. We found out with all of what we’ve done with assessments and STARS that you have to create time to train teachers. You have to create time to teach. You have to create time to assess. You have to create time for people to have an opportunity to work with that material. You have to create time to then analyze that and make another decision about what you do. You’ve got to realize that it’s meaningful enough that you pay attention to it. It’s not . . . you just don’t go through the process and do it and put it away.”
- A male rural high school principal shared the time and energy required by teachers, “It is a burden on teachers. It takes time. It takes energy. Teachers that take it seriously pay a cost for that and so we’ve tried to find ways of rewarding teachers, not just financially, not just time-wise, but other ways, too. For instance, every one of our math teachers are involved in some kind of training they’ve identified that they’ve wanted.”

### **Accountability**

- A male rural high school principal offered, “The ‘ah ha’ is that I really admire teachers in a small school when assessment results come back, they reflect on one person. There’s only one math teacher at our high school level . . . that is tremendous pressure. I’ve gained a whole lot of sensitivity about that. Our teachers want our kids to be successful. They’ll do whatever because ultimately they realize it does reflect on them and them alone. Why weren’t they proficient? In larger districts, there’s a little

more anonymity. They are accountable and they do take it personally. They are doing all things in their power.”

- A female rural high school math teacher referring to accountability and to both the STARS process and the portfolio review process shared, “Yes, I have found through the peer review by having people come out, we have shared ideas and things that their district is doing and our district is doing. So we have learned from them but they have also learned from us. I think it’s been a very positive factor.”
- A male rural assessment coordinator shared, “It’s tough with staff if you don’t have continuity in staff. We have had high staff turnover and it’s really been a struggle to continually update staff on tests and testing procedures. It takes a lot of time.”

### **Stakeholders Accountability**

- A female rural assessment coordinator shared on accountability, “We kind of want to go in our room and just do our thing and say, leave us alone. But we do need to be accountable to our patrons, to our students, to the parents.”
- A female rural elementary teacher shared, “I have a child now at the high school and another one that will go to middle school next year and it’s given me a better insight as a parent as to what these teachers are having to do and what that looks like.”

## **SUMMARY**

The recommended changes to the portfolio peer review process that were made last year appear to have been implemented successfully. Some of the educators interviewed had been a part of the reading peer review process (the first year) and felt confident and less anxious with this process the second time around. Educators interviewed felt comfortable and at ease. Many expressed enjoyment in being a part of the process, felt they benefited from the experience, and believed it was a good process to go through. It was obvious to the researchers that the portfolio peer review process was a success!

## **REFERENCES**

- Plake, B. S. & Impara, J. C. (2000). *Technical criteria for Nebraska school district assessment*. Lincoln, NE: Buros Center for Testing.

# **CHARTING STARS: *Engaging Conversations***

## **Section 3: Research Studies Study II: STARS Enhancement: The Impact of Revisions to the Quality Accountability Act**



### **STARS Enhancement: The Impact of Revisions to the Quality Accountability Act**

**Jody Isernhagen, Ed. D., Associate Professor, University of Nebraska-Lincoln  
Casey Tallent, M. A., Graduate Research Assistant, University of Nebraska-Lincoln**

#### **INTRODUCTION**

The seventh-year study was an initial exploration of the impact of the revisions to the Quality Accountability Act enacted by Nebraska Legislative Bills #653 and #1157. These bills initiated a revision of the state content standards and following the completion of this study the development of state tests. The purpose of the study was to examine participants perceptions about the changes to STARS based on the enhancement criteria. After LB #653 was passed, a newly formed STARS Enhancement Design Team began work to enhance STARS through a revision of content standards. Seven enhancement criteria were arrived upon by consensus of the STARS design team. The Enhancement Criteria are:

- The system is in the best interest of ALL students.
- The system promotes best practices for teaching and learning.
- The system meets federal requirements.
- The system meets state requirements.
- The system includes a balance of classroom-based and large-scale assessment.
- The system is manageable.
- The system is fair, equitable and accurate.

The Enhancement Design team also commissioned three other teams, the Standards Advisory Team, the Assessment Advisory Team, and the Reporting Advisory Team. The teams consisted of NDE officials, Policy Partners, and school district personnel. The teams consisted primarily of members who were familiar with the history of STARS.

## **RESEARCH DESIGN**

The purpose of the study was to explore the impact of the changes to the Quality Accountability Act and to examine the participants' perceptions about the changes to STARS based on the enhancement criteria. Members from the four divisions (i.e., enhancement design team, standards advisory team, assessment advisory team, and reporting advisory team) of the enhancement project were asked to participate in an online survey about the project. The STARS Enhancement project members were surveyed following their second team meeting to assess their experiences as team members and part of the STARS Enhancement project. Of the 83 members who were asked to participate in the survey, 40 members began the survey; however one participant did not consent to take the survey. Of the 39 members who consented to begin the survey, 25 completed the survey in its entirety (62.5%). Of the 25 members responding all four divisions were represented; Assessment Advisory team (38%), Enhancement Design team (29%), Standards Advisory team (16%), and Reporting Advisory team (17%).

Participants responded to a 33-item survey (Appendix E). Participants used a five-point Likert scale, with "1" representing "strongly disagree," "2" "disagree," "3" "neutral," "4" "agree," "5" "strongly agree," for 22 of the questions. Participants also responded to ten open-response questions and one multiple-choice question. The survey was structured to explore overall opinions as well as the perceptions of the four teams: Enhancement Design, Standards Advisory, Assessment Advisory, and Reporting Advisory.

### **Instruments**

The STARS Enhancement survey (Appendix E) was designed by the researchers to collect perceptions about the STARS Enhancement project. The survey examined the participants' feelings about the changes to STARS based on the Enhancement Criteria. Participants responded to a 33-item survey. Participants used a five-point Likert scale, with "1" representing "strongly disagree," "2" "disagree," "3" "neutral," "4" "agree," "5" "strongly agree," for 22 of the questions. Participants also responded to ten open-response questions and one multiple-choice question. Participants responded to seven demographic questions, 23 general questions, and team specific questions that were automatically given to participants based on their team membership.

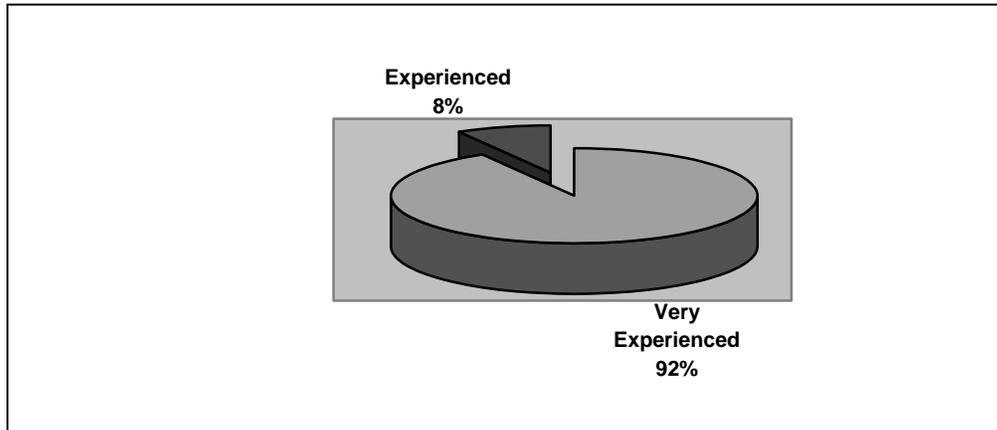
## **RESULTS**

### **Demographics**

Thirty-nine people began the survey, while only 25 (62.5%) completed the survey. The majority of the respondents were female (66.7%), while 33.3% of the respondents were male. In terms of employment of the team members who completed the survey, 77.8% of the respondents were employed by a school district, and 22.2% were employed by a policy partner. Primary roles of the respondent varied widely, with the majority working in curriculum and assessment (32.4%), administration (13.5%), and teaching (10.8%). The

majority, 25 respondents, indicated that they were experienced or very experienced with STARS as indicated in Figure 2.

Figure 2. STARS Enhancement Survey Respondents Level of Experience  
2007-08



### Enhancement Criteria

The nine items were developed to measure the level to which the respondents agree or disagree with the enhancement criteria. To assess participant's perception of the proposed STARS enhancements, nine items on the survey asked them to rate their level of agreement on a five-point Likert scale; from 1-*Strongly Disagree* to 5-*Strongly Agree* with the following statements:

- 1) The purpose of the standards, assessment, and reporting system is to support accountability for continuous improvement ( $M=4.45$ ;  $SD=.67$ ).
- 2) The changes to STARS are in the best interest of ALL students ( $M=3.45$ ;  $SD=1.15$ ).
- 3) The changes to STARS promote best practices for teaching and learning ( $M=3.18$ ;  $SD=1.26$ ).
- 4) The changes to STARS meet federal regulations ( $M=3.52$ ;  $SD=.83$ ).
- 5) The changes to STARS meet state requirements ( $M=3.94$ ;  $SD=.79$ ).
- 6) The changes to STARS include a balance of classroom-based and large-scale assessments ( $M=3.43$ ;  $SD=1.00$ ).
- 7) The changes to STARS are still manageable for schools/districts ( $M=3.67$ ;  $SD=.78$ ).
- 8) After the changes to STARS, the system remains fair, equitable, and accurate ( $M=3.42$ ;  $SD=1.12$ ).
- 9) STARS is enhanced by the changes made by my team ( $M=3.85$ ;  $SD=.94$ ).

The mean scores suggest that the participants are, overall, in agreement with the enhancement criteria. The highest rated item "the purpose of the standards, assessment, and reporting system is to support accountability for continuous improvement" provides evidence that the participants support STARS. However, the lowest rated items, "the changes to STARS promote best practices for teaching and learning" and "after the changes to STARS, the system remains fair, equitable, and accurate" display a lower degree of confidence in the participants that the enhancements will improve teaching and learning while remaining fair and equitable. However, the mean scores for the lowest items are still in the positive range, so any negative interpretations should be made with hesitancy. Additionally, participants were asked six questions in order to ascertain their understanding of the enhancement process as well as their experience in the enhancement process.

Participants selected their level of agreement on a five-point Likert scale, from 1=*Strongly Disagree* to 5=*Strongly Agree* on the following statements:

- 1) I discussed the Enhancement of STARS with others (M=4.36; SD=.54);
- 2) I understand the new requirements that will be implemented by the enhancement of STARS (M=3.97; SD=.88);
- 3) I understand how the enhancement of STARS will impact my school/district (M=3.45; SD=.94);
- 4) I feel that I have a responsibility to help improve STARS with the new enhancements (M=4.33; SD=.69);
- 5) The members of my team worked well together (M=4.67; SD=.48);
- 6) I feel free to express my opinions about the enhancement of STARS with my team (M=4.67; SD=.48).

The responses to this section of the survey suggest that the teams worked well together and that participants felt free to express their opinions with their teams. Additionally, group members appear to have taken their responsibility as a team member seriously and consulted with others about the enhancements. Respondents appear to be somewhat unclear in their understanding of the new requirements and the impact that the enhancements will have on their school or district. The short answer section of the survey allowed participants to state exactly how they felt about how they will be impacted by the enhancements to STARS as well as their opinions on the enhancements and the process. From the short answer questions, several themes emerged across the participant responses. The majority of the participants felt that they will be able to use the skills that they have acquired in the STARS process over the years, particularly the knowledge of assessments. One participant stated, “The assessment literacy that I have learned from implementing STARS will be useful in implementing the enhancement.” Others commented that implementing the enhancements will allow them to “further refine our current practice and skills we’ve already developed.” Additionally, respondents believed that their knowledge of the six quality criteria developed by the Buros Center for Testing will continue to be useful. Overall, respondents appeared to feel confident in their abilities to adjust to the changes to STARS due to the knowledge that they have as one participant stated, “working knowledge of STARS will help with the transition.”

Respondents indicated that they viewed both strengths and limitations in the enhancements to STARS. Clarity of expectations, consistency, comparability, and time management were commonly viewed as strengths of the enhancements. Many of the respondents felt that it was a good time to review STARS and the standards, specifically in reading. Others mentioned that the enhancements could put new energy and “re-focus energy” into high quality assessments. Whereas, a few respondents didn’t feel that there were any benefits to the enhancements, a few also felt that there were no limitations to the enhancements. Some common limitations expressed were the influence of legislatures, decreased “local control,” comparison across districts, and increased work for teachers and districts. Summing up the thoughts of many on the negative impact of the legislative influence on STARS, one respondent stated, “Not a good thing!”

Respondents reflected upon how the changes will influence their professional roles. Several administrators commented that their new role will be learning about the new system and disseminating information to their teachers. Overall, the largest impact that the respondents believe the changes will have on their job is time; many respondents believe that the enhancements will

take time that is already sparse. When considering the impact of the enhancements on the schools/districts, respondents provided mixed responses. Several respondents believed that teachers will be angry and the process will take time away from the classrooms and students, while others believe that the enhancements will increase time for “quality instruction” and professional development. Overall, respondents were unsure of the impact that the enhancements will have on their schools and appeared to be opposed to many of the possible changes that could occur. Respondents were also asked what impact they believe the enhancement to STARS will have on standards and assessments. The majority of respondents believe that the enhancements will improve and revise the standards, resulting in a need for assessment revisions. However, many respondents believe that, although standards will be impacted positively, the move to statewide tests rather than local assessments will have a negative impact.

The impact of legislation was continually expressed by the participants as a negative impact on the educational process in Nebraska. Respondents appear to be reluctant to have the legislature involved in a process that they take much ownership over. Several respondents also stated that they still do not fully understand the impact that the enhancements will have and so their answers may be different as they develop an understanding of the enhancements. Additionally, a few participants commented that the enhancements are dynamic and, therefore, it is difficult to know the impact at the current time. Overall, participants appear to believe that there are several benefits and liabilities in the proposed enhancements to STARS.

### **Design Teams**

Each participant was asked to identify which team of the four design teams that they belonged to during the enhancement process. Each team was then asked a series of five-point Likert scale questions and short answer questions.

### **Enhancement Design Team (N=7)**

The purpose of the Enhancement Design Team was to develop the design criteria to guide the work of the advisory and work groups and to provide input into the overall STARS process. Participants from this team were asked seven questions to specifically examine the work of the team.

Participants rated their answers to these statements on a five-point Likert scale similar to that used in early parts of the survey with 1=*Strongly Disagree* and 5=*Strongly Agree*. The statements were:

- 1) I agree with the STARS Enhancement Criteria that our team came up with. (M=4.86, SD=0.38)
- 2) My expertise was put to good use in developing the STARS Enhancement Criteria. (M=4.86, SD=0.38)
- 3) I feel our team understands the impact that LB 653 will have on Nebraska schools. (M=4.43, SD=0.53)
- 4) Our team meetings were productive. (M=4.71, SD=0.49)
- 5) Our team meetings were convenient for me to attend. (M=4.71, SD=0.49)
- 6) The time that I spent working with my team was worthwhile. (M=4.86, SD=0.38)

Participants were asked to comment on the impact they believed that the Enhancement Design Team would have on STARS. Two participants saw the team as the forum to provide the “big picture.” Many also commented that the team was setting standard criteria to aid the work of other teams and the STARS process as a whole. Participants were also given the opportunity to offer suggestions for improvements for the Enhancement Design team. The majority of people were

pleased with the team. A few people suggested that the team meet more often in order to provide more feedback and to further encourage inter-team communication.

### **Standards Advisory Group (N=3)**

The purpose of the Standards Advisory Group was to create a set of recommendations for standards work groups based upon the design team criteria. Participants from this team were asked seven questions to specifically examine the work of the team. Participants rated their answers to these statements on a five-point Likert scale similar to that used in early parts of the survey with 1=*Strongly Disagree* and 5=*Strongly Agree*. The statements were:

- 1) I feel that the design criteria provided guidance to my team. (M=4.33, SD=0.58)
- 2) I agree with the Standards Revision Criteria that our team came up with. (M=4.67, SD=0.58)
- 3) My expertise was put to good use in developing the Standards Revision Criteria. (M=4.67, SD=0.58)
- 4) I feel that the revision of the standards will enhance STARS. (M=5.00, SD=0)
- 5) Our team meetings were productive. (M=4.67, SD=0.58)
- 6) Our team meetings were convenient for me to attend. (M=4.67, SD=0.58)
- 7) The time that I spent with my team was worthwhile. (M=5.00, SD=0)

Participants were asked to comment on the impact they believed that the Standards Advisory Group would have on STARS. One participant felt that the teams' recommendations would allow for an easier implementation process of STARS revisions in school districts. All commented on the positive nature of the group's accomplishments. Participants were also given the opportunity to offer suggestions for improvements for the Standards Advisory Group. The majority of people were pleased with the team. They commented that more meetings and input from more people on the committee would have improved the team.

### **Assessment Advisory Group (N=9)**

The purpose of the Assessment Advisory Group was to review the work done by the Enhancement Design Team and Standards Advisory Group, reflect on the application of the work to assessment, and to recommend future plans for work groups for assessment. Participants from this team were asked seven questions to specifically examine the work of the team. Participants rated their answers to these statements on a five-point Likert scale similar to that used in early parts of the survey with 1=*Strongly Disagree* and 5=*Strongly Agree*. The statements were:

- 1) I feel that the design criteria provided guidance to my team. (M=3.78, SD=0.97)
- 2) I feel that the Standards Revision Criteria provided guidance to my team. (M=3.67, SD=1.12)
- 3) My expertise was put to good use in developing the Assessment Advisory Group recommendations. (M=3.89, SD=0.93)
- 4) Our team understands the impact that LB 653 will have on assessment. (M=3.67, SD=1.32)
- 5) Our team meetings were productive. (M=4.22, SD=0.44)
- 6) Our team meetings were convenient for me to attend. (M=4.22, SD=0.83)
- 7) The time that I spent with my team was worthwhile. (M=4.22, SD=0.44)

Participants were asked to comment on the impact they believed that the Assessment Advisory Group would have on STARS. Several participants indicated a desire to move forward with an assessment process that would continue to highlight best practices and criterion-based tests.

Participants were also given the opportunity to offer suggestions for improvements for the Assessment Advisory Group. The only suggestion for change in the group was a need for more time to meet to accomplish the goals of the group.

### **Reporting Advisory Group (N=4)**

The purpose of the Reporting Advisory Group was to advise the NDE in how data are reported. Participants from this team were asked seven questions to specifically examine the work of the team. Participants rated their answers to these statements on a five-point Likert scale similar to that used in early parts of the survey with 1=*Strongly Disagree* and 5=*Strongly Agree*. The statements were:

- 1) I feel that the design criteria provided guidance to my team. (M=4.50, SD=0.58)
- 2) I feel that the Standards Revision Criteria provided guidance to my team. (M=4.25, SD=0.50)
- 3) My expertise was put to good use in developing the Reporting Advisory Group recommendations. (M=4.25, SD=0.96)
- 4) Our team understands the impact that LB 653 will have on assessment. (M=4.24, SD=0.50)
- 5) Our team meetings were productive. (M=4.25, SD=0.50)
- 6) Our team meetings were convenient for me to attend. (M=4.25, SD=0.50)
- 7) The time that I spent with my team was worthwhile. (M=4.25, SD=0.50)

Participants were asked to comment on the impact they believed that the Reporting Advisory Group would have on STARS. Many participants indicated that their team would create a foundation of reporting and assist in creating a more seamless process of reporting student performance data. Participants were also given the opportunity to offer suggestions for improvements for the Reporting Advisory Group.

## **DISCUSSION**

As a result of the revisions to the Quality Accountability Act enacted by the Nebraska Legislature, an enhancement process for STARS was developed which is in the best interest of all students, promotes best practices for teaching and learning, meets federal requirements for assessment, meets state requirements, includes a balance of classroom-based and large-scale assessment, is fair, equitable and accurate, and is manageable in Nebraska. This study was designed to examine the impact of this bill on STARS and to examine the participants' perceptions about the changes to STARS based on the enhancement criteria.

Among the participants surveyed, most agreed with the enhancement criteria designed by the STARS team. However, the lower average scores on items about improvement in teaching practice indicated a lower degree of confidence that the enhancements will improve teaching and learning while maintaining fair and equitable practices. Participants did express understanding of the changes to the standards and indicated a personal responsibility to help improve STARS. Participants recognized both strengths and limitations of the STARS enhancement process. Many believed that the revised standards will be beneficial in the long term. Some also favored the ability to compare student performance across districts; however, some saw this as a limitation of the enhancement design. The increased influence of the legislature and the increased time spent outside of the classroom that will be required in the STARS enhancement process were two of the main limitations of the enhancement process expressed by participants.

Participants were also asked to evaluate their experience within their team. The majority felt that the teams worked well together and that the team atmosphere was a forum to openly express opinions about the enhancement of STARS. Team members felt that their expertise was used within the team and that the time spent working in the team was worthwhile. Many of the teams expressed a need for additional time to accomplish the assigned tasks.

### **CONCLUSION**

Overall, participants appear to agree with the enhancement criteria for STARS and to understand the enhancement process developed by the NDE as a result of the revisions to the Quality Accountability Act. Participants expressed a belief in both benefits and liabilities in the proposed enhancements to STARS. In addition, participants favored the enhancement teams and felt that they were able to share their expertise and make an impact in the team environment.

**CHARTING STARS:**  
***Engaging Conversations***  
**Section 3: Research Studies**

**Study III: Reading and Math Achievement**



**2001 to 2007 Reading and Math Achievement  
Grades 4, 8, and 11**

**Shirley Mills, Ph.D., University of Texas-Pan American  
Jody Isernhagen, Ed.D., University of Nebraska-Lincoln**

**INTRODUCTION**

Gallagher and Ratzlaff (2008) referred to Nebraska’s School-based Teacher-led Assessment and Reporting System (STARS) as the “road less traveled.” Nebraska was the only state that opted to develop their own system of local assessments based on six quality criteria developed by the Buros Center for Testing (Plake & Impara, 2000).

1. Assessments align to state or local standards.
2. Students have an opportunity to learn the content that they will be tested.
3. Assessments will be free of bias or offensive language.
4. The level is developmentally appropriate for all students.
5. Scoring is consistent.
6. The mastery levels are appropriate to subject and grade level.

The Nebraska School-based Teacher-led Assessment and Reporting System (STARS) is based on the philosophy described by the National Research Council (2001) that the effectiveness of a state assessment system must be judged by the extent to which it promotes student learning. In search of evidence of the positive effects of high-stakes tests on student achievement, Stiggins (2004), a notable assessment researcher, conducted a study searching for evidence of the positive effects of high stakes tests on student achievement and found only one study that indicated small gains. STARS, Nebraska’s system, was described by the Partnership for the 21<sup>st</sup> Century Skills (2005) as “...the nation’s most innovative assessment

system” (p.13). Sternberg (2008) suggested that assessments should reflect the skills that matter in school as well as life. Conducting high stakes tests written by outside test writers, dominated with multiple choice answers did not seem to meet the goal of improving student achievement in Nebraska (Sternberg, 2008). Who better to assess students of Nebraska than the teachers who are in their classroom on a daily basis? Nebraska honored those teachers in the STARS process and encouraged all teachers to work collaboratively to build and administer district based assessments that were aligned with Nebraska state standards.

Nebraska’s STARS requires each district to either adopt state standards or develop local standards that are at least equal to or exceed the state standards. Each district then developed a plan for assessing their standards. The plan was based primarily on locally developed criterion-referenced tests (CRT’s), which were unique to that district. The STARS assessment results are reported at fourth, eighth, and eleventh grades. Districts also report Average Yearly Progress (AYP) at grades three through eight and one year in high school.

In Nebraska, districts are also required to administer a norm-referenced test (NRT) of their choosing (e.g., Terra Nova, Stanford Achievement Test). NRT’s are perceived by many as reliable indicators of student achievement. As Stiggins (2007) indicated, “a major role of assessment has been to detect and highlight differences in student learning in order to rank students according to their achievement” (p. 22).

## **PURPOSE OF THE STUDY**

The purpose of this study was to examine STARS data available to date for reading and math for all students. An average of the percentage of students in Nebraska school districts demonstrating proficiency in these areas was provided. The report included locally developed criterion-referenced data, norm-referenced data, and District Assessment Portfolio data.

The research questions were:

1. What was the district average percent of students rated as proficient or advanced in reading on their locally developed criterion-referenced measure and the norm-referenced measure used in that district for 2001, 2003, 2005, 2006, and 2007 (years tested to date for reading)?
2. What changes occurred in the district average percent of students rated as proficient in the criterion and norm-referenced data in reading over these years?
3. What was the district average percent of students rated as proficient in math on their locally developed criterion-referenced measure and the norm-referenced measure used in that district for 2002, 2004, 2005, 2006, and 2007 (years tested to date for math)?
4. What changes occurred in the district percentage of students rated as proficient or advanced in the criterion-referenced and norm-referenced data in math over these years?
5. What was the average rating for the District Assessment Portfolios in reading and math over the years of available data?
6. What changes occurred in District Assessment Portfolios in reading and math over these years?
7. What were the implications for the STARS program?

## RESEARCH DESIGN

### Sample

Data was included for Class III, IV, and V school districts. Class III school districts are represented by any school district with territory having a population of more than 1000 but less than 150,000 inhabitants; Class IV school districts (Lincoln only) with a territory having a population of 100,000 or more with a city of the primary class; and Class V school districts (Omaha only) within the territory having a population of 200,000 or more inhabitants with a city of the metropolitan class within the territory (Nebraska Education Directory, 2006-2007). The districts in this study represented all public school students in Nebraska. The district data for this study were included on the state website and use of the data was facilitated by the Nebraska Department of Education (NDE).

### Score Definitions

The criterion-referenced score used was the district average percent of students meeting the proficiency level or better defined by the local district for their locally developed measure. The norm-referenced score was the district average percent of students scoring in the top two quartiles on the nationally standardized test used by that district (e.g., California Achievement Test, Iowa Test of Basic Skills, and Terra Nova). While the norm-referenced measure used will vary, the data reported (percent of students in the top two quartiles) was constant for all districts.

Since tests used to measure standards are a mix of locally developed criterion-referenced measures and may include sections of district specific norm-referenced tests, there are few common measures to all districts. It must be remembered that STARS was designed to support instruction in local classrooms, not to facilitate ranking of schools. This strong reliance on district developed criterion-referenced measures challenged traditional validity and reliability views. Therefore, the primary measure of credibility for assessments was a District Assessment Portfolio that is submitted annually to the NDE.

The Portfolio included school district ratings on six Assessment Quality Criteria that were identified by the Buros Center for Testing (Plake & Impara, 2000), the technical advisors to the STARS program as mentioned in the introduction. Portfolios were rated by an independent measurement expert specifically trained in the rubrics of each of the six Quality Criteria. Locally trained educators assisted in the evaluation of each district in the state.

### Data Analysis

The data for this study were described as unconventional. Traditional inferential statistics, therefore, were not appropriate measures to use. Instead the researchers conducted a study of descriptive data. Researchers averaged the district averages for each district and reported the changes in the total “average of the averages” each year.

## RESULTS

### STARS Reading Achievement

As shown in Table 1, the district average percent of student scores reported by districts as proficient or better in locally defined criterion-referenced reading at the fourth-grade level increased from 74.99% in 2001 to 79.39% in 2003, to 87.20% in 2005, to 90.70% in 2006, and increased to 92.30% in 2007. The district average percent proficient at the eighth-grade

level increased from 73.67% in 2001 to 74.78% in 2003, to 84.49% in 2005, to 87.70% in 2006, and to 89.79% in 2007. The district average percent proficient at the eleventh-grade level increased from 73.54% in 2001 to 74.74% in 2003, to 82.26% in 2005, to 86.10% in 2006, and finally increased to 87.48% in 2007. Table 1 indicates the percent of students who are proficient and/or the changes indicated on the district average criterion-referenced tests in reading district scores across Nebraska.

Proficiency on criterion-referenced measures increased at all grade levels each year; the average district gain from 2001 to 2007 was 17.31% at fourth grade, 16.12% at eighth grade, and 13.94% at eleventh grade.

Table 1.  
*Percent Proficient or Higher (Change) on Criterion-Referenced Tests in Reading<sup>a</sup>*

Grade	2001	2003	2005	2006	2007	Total Change
4	74.99%	79.39% (+ 4.40)	87.20% (+ 7.81)	90.70% (+3.50)	92.30% (+1.60%)	+17.31%
8	73.67%	74.78% (+ 1.11)	84.49% (+ 9.71)	87.70% (+3.21)	89.79% (+2.09%)	+16.12%
11	73.54%	74.74% (+ 1.20)	82.26% (+ 7.52)	86.10% (+3.84)	87.48% (+1.38%)	+13.94%

<sup>a</sup> Percent of students scoring proficient or higher was calculated for each district and then all districts averaged across the state.

Table 2 reports the district average percent of students in the top two quartiles on the norm-referenced reading test used by districts at the fourth grade. The fourth grade increased from 64.93% in 2001 to 66.75% in 2003, to 67.59% in 2005, to 69.42% in 2006, and decreased to 69.25% in 2007. The eighth grade declined from 62.85% in 2001 to 62.56% in 2003, increased to 63.01% in 2005, to 63.24% in 2006, and increased to 63.61% in 2007. The eleventh grade increased from 59.87% in 2001 to 61.44% in 2003, to 63.67% in 2005, but decreased to 63.59% in 2006, and again decreased to 62.05% in 2007. Proficiency, as determined by the percent of students in districts in the top two quartiles on norm-referenced measures, also increased from 2001 to 2007 with a 4.32% increase at fourth grade, a 0.76% increase at eighth grade, and 2.18% increase at eleventh grade.

Table 2  
*Percent Proficient or Higher (Change) on Norm-Referenced Tests in Reading<sup>a</sup>*

Grade	2001	2003	2005	2006	2007	Total Change
4	64.93%	66.75% (+ 1.82)	67.59% (+ 0.84)	69.42% (+1.83%)	69.25% (-0.17)	+4.32%
8	62.85%	62.56% (- 0.29)	63.01% (+ 0.45)	63.24% (+0.23%)	63.61% (+0.37)	+0.76%
11	59.87%	61.44% (+ 1.57)	63.67% (+ 2.23)	63.59% (-0.08%)	62.05% (-1.54)	+2.18%

<sup>a</sup> Percent of students scoring in the top two quartiles was calculated for each district and then averaged for all districts across the state.

### STARS Math Achievement

As shown in Table 3, the district average percent of students reported by districts as proficient or better in locally defined criterion-referenced math at the fourth-grade level increased from 78.29% in 2002 to 85.16% in 2004, to 89.00% in 2005, to 90.90% in 2006, and increased to 92.83% in 2007. The district percent proficient at the eighth-grade level increased from 68.58% in 2002 to 75.34% in 2004, to 80.27% in 2005, to 82.90% in 2006, and increased to 86.04% in 2007. The district percent proficient at the eleventh-grade level increased from 66.22% in 2002 to 72.20% in 2004, to 76.24% in 2005, to 80.30% in 2006, and increased to 84.20% in 2007. Proficiency on criterion-referenced measures increased at all grade levels each year, the increase from 2002 to 2007 at fourth grade was 14.54%, at eighth grade was 17.46%, and at the eleventh grade was 17.98%.

Table 3

*Percent Proficient or Higher (Change) on Criterion-Referenced Tests in Math<sup>a</sup>*

Grade	2002	2004	2005	2006	2007	Total Change
4	78.29%	85.16% (+ 6.87)	89.00% (+ 3.84)	90.90% (+1.9%)	92.83% (+1.93%)	+14.54%
8	68.58%	75.34% (+ 6.76)	80.27% (+ 4.93)	82.90% (+2.63%)	86.04% (+3.14%)	+17.46%
11	66.22%	72.20% (+ 6.98)	76.24% (+ 4.04)	80.30% (+4.06%)	84.20% (+3.90%)	+17.98%

<sup>a</sup> Percent of students scoring proficient or higher was calculated for each district and then averaged across the state.

Table 4 reports the district average percent of students in the top two quartiles on the norm-referenced math test used by districts at the fourth grade. The fourth grade increased from 68.12% in 2002 to 71.31% in 2004, to 72.05% in 2005, to 73.83% in 2006, and decreased to 70.48% in 2007. The eighth grade declined from 67.34% in 2002, to 66.67% in 2004, increased to 73.67% in 2005, declined to 67.83% in 2006, and increased to 68.60% in 2007. The eleventh grade increased from 67.49% in 2002 to 67.90% in 2004, declined to 67.13% in 2005, increased to 67.62% in 2006, and decreased to 66.49% in 2007. Proficiency on norm-referenced measures increased and decreased slightly from 2002-2007 at all grade levels. However, the overall gain from 2002 to 2007 was a slight gain at fourth (2.36) and eighth (1.26) grades, with a slight decrease at grade eleven (-1.00).

Table 4

*Percent Proficient or Higher (Change) on Norm-Referenced Tests in Math<sup>a</sup>*

Grade	2002	2004	2005	2006	2007	Total Change
4	68.12%	71.31% (+ 3.19)	72.05% (+ 0.74)	73.83% (+1.78%)	70.48% (-3.35%)	+2.36%
8	67.34%	66.67% (+0.33)	73.67% (+ 7.0)	67.83% (-5.84%)	68.60% (+0.77%)	+1.26%
11	67.49%	67.90% (+0.41)	67.13% (- 0.77)	67.62% (+.49%)	66.49% (-1.13%)	-1.00%

<sup>a</sup> Percent of students scoring in the top two quartiles was calculated for each district and then averaged across the state.

### District Assessment Portfolio Ratings

As reflected in Table 5, the total district average of Reading Assessment Portfolio ratings across grades four, eight, and eleven, on the “1” to “5” Likert scale, increased at grade four from 3.57 in 2001 to 4.34 in 2003, to 4.55 in 2005, decreased to 4.54 in 2006, and increased to 4.83 in 2007. This represented a total increase of 1.26 from 2001 to 2007. Portfolio ratings at grade eight increased from 3.48 in 2001 to 4.35 in 2003, to 4.56 in 2005, remained the same (4.56) in 2006, and increased to 4.83 in 2007. This represented a total increase of 1.35 from 2001 to 2007. Portfolio ratings at grade eleven increased from 3.46 in 2001 to 4.35 in 2003, to 4.55 in 2005, remained the same (4.55) in 2006, and increased to 4.76 in 2007. This was a total increase of 1.30 from 2001 to 2007.

The total district average of Reading Assessment Portfolio rating across grades four, eight and eleven increased from 3.50 in 2001 to 4.35 in 2003, to 4.55 in 2005, declined to 4.48 in 2006, and increased to 4.81 in 2007. This was a total increase of 1.30 from 2001 to 2007.

Table 5

*Reading District Average Portfolio Ratings (Gain/Loss) 2001-2007*

Grade	2001	2003	2005	2006	2007	Total Change 2001-2007
4	3.57	4.34 (+.077)	4.55 (+0.15)	4.54 (-.01)	4.83 (+0.29)	+1.26
8	3.48	4.35 (+0.87)	4.56 (+0.21)	4.56 (0)	4.83 (+0.27)	+1.35
11	3.46	4.35 (+0.89)	4.55 (+0.20)	4.55 (0)	4.76 (+0.21)	+1.30
Total Average Rating (4, 8, 11)	3.50	4.35 (+2.53)	4.55 (+0.56)	4.48 (-0.01)	4.81(+0.33)	+1.31

As reflected in Table 6, the total district average for Math Assessment Portfolio ratings across grades four, eight, and eleven, on the “1” to “5” Likert scale increased at grade four from 3.98 in 2002, to 4.67 in 2004, declined to 4.57 in 2005, remained the same (4.57) in 2006, and increased to 4.75 in 2007. This reflected an increase of 0.77 from 2002 to 2007. Portfolio ratings at grade eight increased from 3.96 in 2002 to 4.77 in 2004, declined to 4.66 in 2005, declined to 4.65 in 2006, and increased to 4.85 in 2007 reflecting a total increase of 0.89 from 2002 to 2007. Grade eleven portfolio ratings increased from 3.96 in 2002 to 4.77 in 2004, declined to 4.60 in 2005, increased to 4.61 in 2006, and increased to 4.85 in 2007.

The total district average of the Math Assessment Portfolio rating across grades four, eight, and eleven increased from 3.97 in 2002, increased to 4.74 in 2004, declined to 4.61 in 2005, remained at 4.61 in 2006, and increased to 4.82 in 2007. This was a total increase of 0.85 from 2002 to 2007.

Table 6

*Math District Average Portfolio Ratings (Gain/Loss) 2002-2007*

Grade	2002	2004	2005	2006	2007	Total Change
4	3.98	4.67 (+0.69)	4.57 (-0.10)	4.57 (0)	4.75 (+0.18)	+0.77
8	3.96	4.77 (+0.81)	4.66 (-0.11)	4.65 (-0.01)	4.85 (+0.20)	+0.89
11	3.96	4.77 (+0.81)	4.60 (-0.17)	4.61 (+0.01)	4.85 (+0.24)	+0.89
Total Average Rating (4, 8, & 11)	3.97	4.74 (+2.31)	4.61 (-0.38)	4.61 (0)	4.82 (+0.62)	+0.85

**SUMMARY**

District criterion-referenced measures for grades four and eight continue to show growth over time in the areas of reading and math from 2001 to 2007. District norm-referenced measures have generally increased in reading and math for fourth and eighth grades with a small decline for eleventh grade from 2001 to 2007. The district assessment portfolio ratings have increased for both reading and math for all grades from 2001 to 2007.

School improvement with student academic achievement as the goal was not intended to be a short-term process. Nebraska is in its seventh year of full implementation of the STARS program and variability still exists in achievement in some areas; however, Nebraska educators have made strides towards improving all students' math and reading scores on the district average achievement scores. Generally, criterion-referenced and norm-referenced assessment scores have improved. The portfolio scores for reading and math have consistently improved over time.

**REFERENCES**

- Gallagher, C. W., & Ratzlaff, S. (Dec., 2007, Jan. 2008). The road less traveled. *Educational Leadership*, 65(4), 48-53.
- National Research Council. (2001). *Knowing what students know. The science and design of educational assessment*. Washington, DC: National Academy Press.
- Nebraska Department of Education. (n.d.). *Nebraska standards in reading/writing*. Retrieved March 3, 2006 from <http://www.nde.state.ne.us/ndestandards/AcademicStandards.htm>
- Nebraska Department of Education. (2006-2007). *Nebraska Department of Education Directory*. Lincoln, NE. 109<sup>th</sup> Edition, 158.
- Partnership for 21<sup>st</sup> Century Skills. (2005). *The Road to 21<sup>st</sup> century learning: A policymaker's guide to 21<sup>st</sup> century skills*. Retrieved on May 4, 2005 from <http://www.21stcenturyskills.org/route/>
- Plake, B. C., & Impara, J. C. (2000). *Technical Quality Criteria for Nebraska's District Assessments*. Lincoln, NE: Buros Center for Testing.

- Sternberg, R. J. (Dec., 2007, Jan. 2008). The road less traveled. *Educational Leadership*, 65(4), 20-26.
- Stiggins, R. (2004). New assessment beliefs for a new school mission. *Phi Delta Kappan*, 86(1), 22-27.
- Stiggins, R. (2007). Assessment through the student's eyes. *Educational Leadership*, 64(8), 22-26.

**CHARTING STARS:  
*Engaging Conversations*  
Section 3: Research Studies**

**Study IV: 2002-2007 Writing Achievement**



**2002-2007 Nebraska STARS Statewide  
Writing Achievement**

**Shirley Mills, Ph.D., University of Texas-Pan American  
Jody Isernhagen, Ed.D., University of Nebraska-Lincoln**

**INTRODUCTION**

Conversations with educators across the state of Nebraska have focused on writing skills since the early 2000's. Nebraska Legislative Bill 812, which amended State Statute 79-760 (Educational Quality and Accountability Act, 1999), required district involvement in a statewide assessment of writing for all students in grades four, eight, and eleven as a part of Nebraska School-based Teacher-led Assessment and Reporting System (STARS). This bold step required that each district either adopt the state writing standards or develop writing standards equal to or more rigorous than the state standards.

The purpose of the assessment is to measure student achievement in meeting the state's writing standards which address proficiencies in standard English conventions, text structures, revision and editing, and use of writing in multiple formats including descriptive, persuasive, and narrative writing for different audiences and purposes. The Nebraska Department of Education (NDE) releases results for the Statewide Writing Assessment and all Nebraska STARS assessments on the State of the Schools website each fall. Local district data shared on the NDE web site is provided by school districts and includes the percentage of students meeting proficiency at fourth, eighth, and eleventh- grade levels in designated content areas for each district. The statewide writing assessment score is the average percentage of students rated proficient from each district.

## **PURPOSE OF THE STUDY**

The purpose of this study was to examine the district achievement data available for the Statewide Writing Assessment for the Nebraska STARS program. The research questions for this study were:

1. What was the district average percent of students rated proficient or better in grades four, eight, and eleven for years 2002-2007 on the Nebraska Statewide Writing Assessment?
2. What were the changes over the years 2002-2007 in the district average percent of students rated as proficient or better in district writing scores at grades four, eight, and eleven?
3. What were the teacher perceptions of the writing assessment?
4. What were the implications for the Nebraska STARS program?

## **RESEARCH DESIGN**

### **Sample**

Data was included for Class III, IV, and V school districts. Class III school districts are represented by any school district with territory having a population of more than 1000 but less than 150,000 inhabitants; Class IV school districts (Lincoln only) with a territory having a population of 100,000 or more with a city of the primary class; and Class V school districts (Omaha only) within the territory having a population of 200,000 or more inhabitants with a city of the metropolitan class within the territory (Nebraska Education Directory, 2006-2007). The districts in this study represented all of the public school students in Nebraska. The district data for this study were included on the state website and use of the data was facilitated by the Nebraska Department of Education (NDE).

### **Statewide Writing Assessment Prompt Development**

The process of the development of writing prompts for use in the Statewide Writing Assessment relied on the involvement of Nebraska classroom teachers. Participating teachers were recommended by their district superintendent or assessment contact person and selected by the NDE each year to take part in a writing development task force. The task force consisted of three panels, each consisting of 10-15 teachers representing grades four, eight, and eleven from a variety of school sizes and geographic regions. The task force was convened for a one-day workshop facilitated by the NDE for the purpose of:

- Reviewing the characteristics of mode-specific writing.
- Learning the criteria for effective writing prompts.
- Reviewing and examining areas of bias to be avoided.
- Creating writing prompts for field testing.

During the workshop, participants read and discussed examples of current research related to best practices in the teaching and assessment of student writing. In addition, they read about and discussed criteria for effective writing prompts as well as issues related to bias that should be avoided when creating writing prompts. A number of examples of writing prompts including those that had been used in previous Nebraska statewide writing assessments were also reviewed.

### **Statewide Writing Assessment Prompts Field Testing Process**

From information gathered at the Writing Prompt Development workshop, school districts representing various sizes and geographic locations were selected to field test the writing prompts with students in grades four, eight, and eleven before the end of the current school year. Multiple prompts were field tested across multiple school districts.

Participants in the field-testing were at the appropriate grade level and completed assessments according to standard administration procedures. The size of the student sample selected for the field-testing was adequate to provide responses sufficient for scoring and anchoring purposes. At the conclusion of the field-testing, the NDE conducted a review to “fine-tune” the Statewide Writing Assessment scoring process.

### **Statewide Scoring**

Nebraska teachers are recruited by the NDE to score the writing assessment each year. The scorer qualifications included:

- The teacher was currently teaching or had taught at or near the grade-level being assessed.
- The teacher was familiar with student writing at the grade-level being assessed.
- The teacher had basic knowledge of the Six Trait writing assessment model.

### **Statewide Writing Assessment Scoring Process**

Scoring of the state assessment was held at a central location in the state and scorers came to the site for three days during which training and the scoring occurred. The scoring process of Nebraska’s Statewide Writing Assessment required each sample of student writing to be read and scored by two trained teacher raters who assigned a single holistic score within allowable ranges as prescribed by the rubric. The rubric criteria were identified as ideas and content, organization, voice or tone, word choice, sentence fluency, and conventions as identified in the Nebraska Content Standards (NDE, 2008). Raters assigned a score based on how the writing met these criteria overall. If there was more than a two-point difference, a third scoring was done. The scoring range was from one to four in + and – intervals resulting in a ten-point scale. The final score was the composite of the two individual scores. The NDE contracted with the Buros Center for Testing to establish the statewide cut-score.

In the first three years, scoring was done at three sites across the state. To improve reliability, scoring is now done at one site. A sample was sent out of state for scoring by an independent contracted testing company. The NDE released results for the statewide writing assessment and all Nebraska STARS assessments on their website each fall. Local district and individual school data shared included the district average percent of students meeting proficiency or better on the Statewide Writing Assessment.

### **Data Analysis**

The unit of analysis for this study was the district average percent of students rated as proficient in Class III, IV, and V school districts for the State of Nebraska in writing at grades four, eight, and eleven. While this statewide assessment took on some formal technical assessment characteristics that would more characterize norm-referenced tests (statewide common administration and scoring, common cut-score) than many criterion-referenced assessments, it was clearly not a comparison with a separate norm group.

Descriptive data was, therefore, reported and discussed. However, because the assessment was a common measure across districts and was an equal interval scale, inferential statistics were also used to examine statistical significance between pre/post scores from inception to last scoring. All significance tests were two-tailed.

## RESULTS

### Ratings of Writing Proficiency

As shown in Table 7, the district average percent of student scores reported by districts as proficient or better in locally defined criterion-referenced reading at the fourth-grade level increased from 76.50% in 2002 to 80.89% in 2004, to 84.57% in 2005, decreased to 83.41% in 2006, and increased to 85.32% in 2007. The district average percent proficient at the eighth-grade level significantly increased from 79.55% in 2003 to 85.44% in 2004, to 86.31% in 2005, to 87.00% in 2006, and increased to 91.38% in 2007. The district average percent proficient at the eleventh-grade level increased from 89.22% in 2004 to 91.02% in 2005, to 91.91% in 2006, and increased to 92.23% in 2007.

Writing scores increased for all grades, with grade four increasing significantly by 8.82% from 2002 to 2007, grade eight increasing significantly by 11.83%, and grade eleven increasing 3.01% during the same period as indicated in Table 7

Table 7  
*Statewide Writing Assessment 2002-2007: Mean District Percent of Student Scores at the Proficient Level or Higher<sup>a</sup>*

Grade	2002	2003	2004	2005	2006	2007	Change 2007 vs. Baseline
4	76.50%		80.89% <sup>**</sup>	84.57% <sup>**</sup>	83.41%	85.32%	+8.82% <sup>††</sup>
8		79.55%	85.44% <sup>**</sup>	86.31%	87.00%	91.38% <sup>**</sup>	+11.83% <sup>††</sup>
11			89.22%	91.02% <sup>*</sup>	91.91%	92.23%	+3.01%

\*p<.05 compared to the previous year

\*\*p<.001 compared to the previous year

†p<.05 compared to the baseline year

††p<.001 compared to the baseline year

Note: 2002, 2003, and 2004 were baseline years for grades 4, 8, and respectively

<sup>a</sup> Percent of students scoring proficient or higher was calculated for each district and then averaged across the state

### Teacher Perceptions

In “Charting STARS – Sustainability as Challenge and Opportunity,” Gallagher (2003) reported the results of a second year research study and comprehensive evaluation of Nebraska’s School-based Teacher-led Assessment and Reporting system (STARS). Among the major findings of the analyses of a survey administered to teachers on their perceptions and classroom practices related to their involvement with the state writing assessment were that 69% of all teachers placed more emphasis on practice writing assessments, 73% placed more emphasis on sharing assessment criteria in class, and 73% placed more emphasis on explicit instruction in six trait writing.

In addition to these findings, Gallagher (2003) reported that 88% of teachers agreed or strongly agreed that the six traits scoring rubric used to score the state writing assessment was useful for instruction; 75% agreed or strongly agreed that the state writing assessment supported learning objectives they have for their students; 72% agreed or strongly agreed that

the results of the state writing assessment were useful for teachers; and 65% agreed or strongly agreed that the six traits were the most important features of writing. The positive perception of teachers reported in studies of the Nebraska Statewide Writing Assessment System (Anderson 2005, 2007; Gallagher, 2003) and writing gains from this study are consistent with the literature relative to the value of teacher involvement in the writing process.

### SUMMARY

The purpose of this study was to examine the district achievement data available for the Statewide Writing Assessment for all students across Nebraska. Results indicated that fourth and eighth grades made significant gains from the baseline comparisons to 2007 on the Statewide Writing Assessment while the eleventh grade displayed a non-significant increase in the baseline comparison to 2007.

Nebraska's writing results show positive student gains and would indicate support for continuation of the statewide writing assessment component of STARS. These findings provide a base of support and, along with the gains in reading and math, credibility for the general STARS process. The Nebraska Statewide Writing Assessment is a relatively new model and while it has undergone improvements and has been receiving good responses over the years of implementation, evaluative comments will remain guarded until replications of this model by other states are made and further longitudinal evaluations are completed.

Finally, the researchers urge readers to use caution when reviewing the statistics. While this writing assessment has characteristics that enable inferential statistical analysis to be used, there may be some question from the traditional measurement community concerning this practice.

### REFERENCES

- Anderson, S. (2005). *Statewide writing assessment scoring: Effective professional development for the classroom teacher. Charting STARS: Voices from the field. Year-Four Report. Comprehensive Evaluation of Nebraska's School-Based Teacher-Led Assessment and Reporting System (STARS)*. Lincoln, NE: University of Nebraska.
- Anderson, S. (2007). *Statewide writing assessment scoring: Effective professional development for the classroom teacher. Charting STARS: Pathways to Success. Year-Six Report. Comprehensive Evaluation of Nebraska's School-Based Teacher-Led Assessment and Reporting System (STARS)*. Lincoln, NE: University of Nebraska.
- Educational Quality and Accountability Act (The). (1999). State Statute 79-760.
- Gallagher, C. W. (2003). *Charting STARS: New conversations. Year three report to the Nebraska Department of Education*. Lincoln, NE: University of Nebraska.
- Nebraska Department of Education. (2006-2007). *Nebraska Department of Education Directory*. Lincoln, NE. 109<sup>th</sup> Edition, 158.
- Nebraska Department of Education. (2008). Nebraska standards in reading/writing. Retrieved from [http://www.nde.state.ne.us/ndestandards/academic\\_standards.htm](http://www.nde.state.ne.us/ndestandards/academic_standards.htm)



**CHARTING STARS:**  
***Engaging Conversations***  
**Section 3: Research Studies**

**Study V: 2001-2007 Achievement for  
Special Populations**



**2001-2007 Achievement for Special Populations**

**Shirley Mills, Ph.D., University of Texas-Pan American**  
**Jody Isernhagen, Ed.D., University of Nebraska-Lincoln**

**INTRODUCTION**

The 2007 testing year was significantly impacted by the mandates that began this year for the special populations known as English Language Learners (ELL) and Special Education (SPED) students across the nation. This testing year, all students were required to take assessments at their respective grade level. This change was mandated by the No Child Left Behind Act in order to demonstrate increased academic achievement for all students.

Accommodations and modifications for SPED students, as indicated by their Individual Education Plans (IEP), can be observed during the testing time, but the test itself must be the same as all other students. ELL students must take the English version of the tests.

In this pilot year, many educators across the nation expect lower scores for special population students. However, Nebraska has been unique as students were always required to test at grade level. It is anticipated that the impact of the mandate will not significantly influence overall scores in the state.

## **PURPOSE OF THE STUDY**

The purpose of this study was to examine academic change of special populations, in this case, ELL and SPED students. The questions for this study were:

1. What were the changes in district average percent of students rated as proficient or higher in reading for 2001-2007 on their locally developed criterion-referenced tests for ELL and SPED students?
2. What were the changes in district average percent of students rated as proficient or higher in math for 2002-2007 on their locally developed criterion-referenced tests for ELL and SPED students?
3. What were the changes in district average percent of students rated as proficient or higher in writing on the statewide criterion-referenced writing assessment for ELL and SPED students?
4. What were the implications to the Nebraska STARS program from these findings?

## **RESEARCH DESIGN**

### **Districts Included**

Data for this study were included for Class III, IV, and V school districts. Class III school districts are represented by any school district with territory having a population of more than 1000 but less than 150,000 inhabitants; Class IV school districts (Lincoln only) with a territory having a population of 100,000 or more with a city of the primary class; and Class V school districts (Omaha only) within the territory having a population of 200,000 or more inhabitants with a city of the metropolitan class within the territory (Nebraska Education Directory, 2006-2007). The districts in this study represented all public school students in Nebraska. The district data for this study were included on the state website and use of the data was facilitated by the Nebraska Department of Education (NDE).

### **Score Definitions**

The criterion-referenced score (CRT) for reading and math was the district average percentage for ELL and SPED students meeting the proficiency level or better defined by the local districts for their locally developed measure in classes III, IV, and V school districts for the state of Nebraska.

The criterion-referenced score (CRT) for writing was the district average percent of ELL and SPED students rated as proficient in classes III, IV, and V school districts for the state of Nebraska in writing at grades four, eight, and eleven.

### **Data Analysis**

Criterion-referenced scores for reading and math were unique for each district, therefore not on a common scale and did not support common and inferential statistics. Descriptive data only was reported and discussed for reading and math scores by averaging each district score across the state.

The writing assessment was a common measure across districts and was an equal interval scale. Therefore, inferential statistics were used to examine statistical significance between pre/post scores in writing. All significance tests were two-tailed.

## RESULTS

### STARS Achievement for English Language Learners

#### *ELL Reading Achievement*

The district average percent of ELL reported as proficient or higher on locally defined criterion-referenced assessments for reading at the fourth-grade level decreased from 50% in 2001 to 49% in 2003, increased to 67% in 2005, increased to 72% in 2006, and then to 79% in 2007. Proficiency on reading criterion-referenced measures at grade four indicated an increase of 29% from 2001 to 2007 as shown in Table 8.

The district average percent of ELL students reported as proficient or higher on locally defined criterion-referenced assessments for reading for eighth-grade students decreased from 47% in 2001 to 42% in 2003, increased to 57% in 2005 to 60% in 2006, and increased to 65% in 2007. Proficiency on reading criterion-referenced measures for eighth-grade ELL students increased by 18% from 2001 to 2007 as shown in Table 8.

The district average percent of ELL reported as proficient or higher on locally defined criterion-referenced assessments for reading for eleventh-grade students decreased from 45% in 2001 to 32% in 2003, increased to 47% in 2005, to 53% in 2006, and increased to 57% in 2007. Proficiency on reading criterion-referenced measures for eleventh-grade ELL students increased 12% from 2001 to 2007 as shown in Table 8.

Therefore, district average percent of ELL students reported as proficient or higher in locally defined criterion-referenced assessments for reading at all grade levels increased from 2001 to 2007.

Table 8

*District Average Reading Criterion-Referenced Assessments English Language Learners (ELL) 2001-2007<sup>a</sup> (District Sample Size – Number of districts reporting from population)*

ELL Grade	District Mean Percent Proficient					Change in District Mean Percent Proficient				
	2001	2003	2005	2006	2007	2001-03	2003-05	2005-06	2006-07	2001-07
4	50%	49%	67%	72%	79%	-1%	+18%	+5%	+7%	+29%
8	47%	42%	57%	60%	65%	-5%	+15%	+3%	+5%	+18%
11	45%	32%	47%	53%	57%	-13%	+15%	+6%	+4%	+12%

<sup>a</sup> Percent of students scoring proficient or higher was calculated for each district and then averaged across the state.

All percentages were rounded to whole percents.

#### *ELL Math Achievement*

The district average percent of English Language Learners (ELL) reported as proficient or better on locally defined criterion-referenced assessments for math at the fourth-grade level increased from 53% in 2002 to 70% in 2004, to 72% in 2005, to 80% in 2006, and increased to 83% in 2007. Proficiency on math criterion-referenced measures for fourth-grade ELL students increased noticeably by 30% from 2002 to 2007 as shown in Table 9.

Table 9

*District Average Math Criterion-Referenced Assessments English Language Learners (ELL) 2002-2007<sup>a</sup> (District Sample Size - Number of districts reporting from population)*

ELL	District Mean Percent Proficient					Change in District Mean Percent Proficient				
	Grade	2002	2004	2005	2006	2007	2002-04	2004-05	2005-06	2006-07
4	53%	70%	72%	80%	83%	+17%	+2%	+8%	+3%	+30%
8	40%	43%	59%	61%	62%	+3%	+16%	+2%	+1%	+22%
11	39%	48%	51%	48%	61%	+9%	+3%	-3%	+13%	+22%

<sup>a</sup> Percent of students scoring proficient or higher was calculated for each district and then averaged across the state.

All percentages were rounded to whole percents.

The district average percent of ELL reported as proficient or better on locally defined criterion-referenced assessments for math for eighth-grade students increased from 40% in 2002 to 43% in 2004, to 59% in 2005, to 61% in 2006, and increased to 62% in 2007. Proficiency on math criterion-referenced measures for eighth-grade ELL students increased 22 % from 2002-2007 as also shown in Table 9.

The district average percent of ELL reported as proficient or better on locally defined criterion-referenced assessments for math for eleventh-grade students increased from 39% in 2002 to 48% in 2004, to 51% in 2005, then decreased to 48% in 2006, and increased to 61% in 2007. Proficiency on math criterion-referenced measures for eleventh-grade ELL students increased 22% from 2002 to 2007 as shown in Table 9.

Therefore, the district average percent of ELL students reported as proficient increased in math at all grade levels from 2001 to 2007.

***ELL Writing Achievement***

The district average percent of ELL reported as proficient or better on the state criterion-referenced assessment for writing at the fourth-grade level increased from 49% in 2002 to 52% in 2004, to 64% in 2005, to 66% in 2006, and to 69% in 2007. Proficiency on the state writing criterion-referenced measure significantly increased 20% from 2002 to 2007 as shown in Table 10.

The district average percent of ELL reported as proficient or better on the state criterion-referenced assessment for writing for eighth-grade students increased significantly from 37% in 2003 to 56% in 2004, increased to 60% in 2005, decreased to 56% in 2006, and increased to 62% in 2007. Proficiency on the state writing criterion-referenced measure for eighth-grade ELL students increased significantly 25% from 2003 to 2007 as shown in Table 10.

The district average percent of ELL reported as proficient or better on the state criterion-referenced assessment for writing for eleventh-grade students decreased from 45% in 2004 to 44% in 2005, increased to 53% in 2006, and increased to 56% in 2007. Proficiency on the state writing criterion-referenced measure for eleventh-grade ELL students increased 11% from 2004 to 2007 as shown in Table 10.

In summary, the district average percent proficient for ELL fourth-grade students increased significantly by 20% from 2002-2007. The district average percent proficient for ELL eighth-grade students increased significantly by 25% from 2003-2007. The district average

percent for ELL eleventh-grade students increased by 11% from 2004-2007 as shown in Table 10.

Table 10

*District Average Writing Criterion-Referenced Assessments English Language Learners (ELL) 2002-2007<sup>a</sup> (District Sample Size - Number of districts reporting from population)*

ELL Grade	District Mean Percent Proficient						Change in District Mean Percent Proficient					
	2002	2003	2004	2005	2006	2007	02-04	03-04	04-05	05-06	06-07	2002-07
4	49%		52%	64%	66%	69%	+3%		+12%	+2%	+3%	+20% <sup>♦♦</sup>
8		37%	56%*	60%	56%	62%		+19%	+4%	-4%	+6%	+25% <sup>♦♦♦</sup>
11			45%	44%	53%	56%			-1%	+9%	+3%	+11%

\*p<.05 compared to the previous year

♦ p<.05 compared to the baseline year

\*\* p<.01 compared to the previous year

♦♦ p<.01 compared to the baseline year

\*\*\* p<.001 compared to the previous year

♦♦♦ p<.001 compared to the baseline year

Note: 2002, 2003, and 2004 were baseline years for grades 4, 8, and 11 respectively

<sup>a</sup> Percent of students scoring proficient or higher was calculated for each district and then averaged across the state.

All percentages were rounded to whole percents.

## STARS Achievement for Special Education Students

### *SPED Reading Achievement.*

The district average percent of SPED students reported as proficient or better on locally defined criterion-referenced assessments for reading at the fourth-grade level increased from 44% in 2001 to 52% in 2003, to 67% in 2005, to 74% in 2006, and increased to 81% in 2007. Proficiency of SPED fourth-grade students on reading criterion-referenced measures increased 37% from 2001 to 2007 as shown in Table 11.

The district average percent of SPED eighth-grade students reported as proficient or better on locally defined criterion-referenced assessments for reading remained the same from 2001 to 2003 (43%), increased to 59% in 2005, to 66% in 2006, and increased to 72% in 2007. Proficiency on reading criterion-referenced measures for eighth-grade SPED students increased 29% from 2001 to 2007 as also shown in Table 11.

The district average percent of SPED scores for eleventh-grade students reported as proficient or better on locally defined criterion-referenced assessments for reading decreased from 42% in 2001 to 37% in 2003, increased to 54% in 2005, to 61% in 2006, and increased to 65% in 2007. Proficiency on reading criterion-referenced measures for eleventh-grade SPED students increased 23% from 2001 to 2007 as shown in Table 11.

Therefore, the district average percent of SPED students' scores reported as proficient or better on locally defined criterion-referenced assessment increased in reading at all grade levels from 2001-2007.

Table 11

*District Average Reading Criterion-Referenced Assessments Special Education (SPED) Students 2001-2007<sup>a</sup> (District Sample Size - Number of districts reporting from population)*

SPED	District Mean Percent Proficient					Change in District Mean Percent Proficient				
	Grade	2001	2003	2005	2006	2007	01-03	03-05	05-06	06-07
4	44%	52%	67%	74%	81%	+8%	+15%	+7%	+7%	+37%
8	43%	43%	59%	66%	72%	+0%	+16%	+7%	+6%	+29%
11	42%	37%	54%	61%	65%	-5%	+17%	+7%	+4%	+23%

<sup>a</sup> Percent of students scoring proficient or higher was calculated for each district and then averaged across the state.

All percentages were rounded to whole percents.

***SPED Math Achievement***

The district average percent of SPED students reported as proficient or better on locally defined criterion-referenced assessments for math at the fourth-grade level increased from 51% in 2002 to 62% in 2004, to 72% in 2005, to 75% in 2006, and increased to 82% in 2007. Proficiency on math criterion-referenced measures increased 31% from 2002 to 2007 as shown in Table 12.

The district average percent of SPED eighth-grade students reported as proficient or better on locally defined criterion-referenced assessments for math increased 34% from 2002 to 44% in 2004, to 54% in 2005, to 56% in 2006, and increased to 64% in 2007. Proficiency on math criterion-referenced measures for eighth-grade SPED students increased 30% from 2002 to 2007 as also shown in Table 12.

Table 12

*District Average Math Criterion-Referenced Assessments Special Education (SPED) Students 2002-2007<sup>a</sup> (District Sample Size- Number of districts reporting from population)*

SPED	District Mean Percent Proficient					Change in District Mean Percent Proficient				
	Grade	2002	2004	2005	2006	2007	02-04	04-05	05-06	06-07
4	51%	62%	72%	75%	82%	+11%	+10%	+3%	+7%	+31%
8	34%	44%	54%	56%	64%	+10%	+10%	+2%	+8%	+30%
11	28%	36%	43%	46%	55%	+8%	+7%	+3%	+9%	+27%

<sup>a</sup> Percent of students scoring proficient or higher was calculated for each district and then averaged across the state.

All percentages were rounded to whole percents.

The district average percent of SPED scores for eleventh-grade students reported as proficient or better on locally defined criterion-referenced assessments for math increased from 28% in 2002 to 36% in 2004, to 43% in 2005, to 46% in 2006, and increased to 55% in 2007. Proficiency on math criterion-referenced measures for eleventh-grade SPED students increased 27% from 2002 to 2007 as shown in Table 12.

Therefore, the district average percent of SPED students' scores reported as proficient or better on locally defined criterion-referenced assessments increased in math at all grade levels from 2002-2007.

### ***SPED Writing Achievement***

The district average percent of SPED students reported as proficient or better on the state criterion-referenced assessment for writing at the fourth-grade level increased significantly from 46% in 2002 to 55% in 2004, increased significantly to 65% in 2005, decreased to 64% in 2006, and increased significantly to 69% in 2007 as shown in Table 13. Proficiency on the state writing criterion-referenced measure increased significantly by 23% from 2002 to 2007 as shown in Table. 13.

The district average percent of SPED students reported as proficient or better on the state criterion-referenced assessment for writing for eighth-grade students increased significantly from 48% in 2003 to 55% in 2004, then increased to 61% in 2005 and to 63% in 2006, and increased to 67% in 2007. Proficiency on the state writing criterion-referenced measure for eighth-grade SPED students increased significantly by 19% from 2003 to 2007 as shown in Table 13.

Table 13

*District Average Writing Criterion-Referenced Assessments Special Education (SPED) Students 2002-2007<sup>a</sup>(District Sample Size- Number of districts reporting from population)*

SPED Grade	District Mean Percent Proficient						Change in District Mean Percent Proficient					
	2002	2003	2004	2005	2006	2007	02-04	03-04	04-05	05-06	06-07	2002-07
4	46%		55%*	65%***	64%	69%*	+9%		+10%	-1%	+5%	+23%***
8		48%	55%**	61%	63%	67%		+7%	+6%	+2%	+4%	+19%***
11			55%	63%**	65%	65%			+8%	+2%	0%	+10%***

\*p<.05 compared to the previous year

\*\* p<.01 compared to the previous year

\*\*\* p<.001 compared to the previous year

♦ p<.05 compared to the baseline year

♦♦ p< .01 compared to the baseline year

♦♦♦ p<.001 compared to the baseline year

Note: 2002, 2003, and 2004 were baseline years for grades 4, 8, and 11 respectively

<sup>a</sup> Percent of students scoring proficient or higher was calculated for each district and then averaged across the state.

All percentages were rounded to whole percents.

The district average percent of SPED students reported as proficient or better on the state criterion-referenced assessment for writing for eleventh-grade students increased significantly from 55% in 2004 to 63% in 2005, increased to 65% in 2006, and stayed the same at 65% in 2007. Proficiency on the state writing criterion-referenced measure for eleventh-grade SPED students increased significantly by 10% from 2004 to 2007 as shown in Table 13.

In summary, the district average percent proficient for SPED fourth-grade students on the state criterion-referenced assessment for writing increased significantly by 23% from 2002-2007. The district average percent proficient for SPED eighth-grade students increased significantly by 19% from 2003-2007. The district average percent for SPED eleventh-grade students increased significantly by 10% from 2004-2007 as shown in Table 13.

### **SUMMARY**

Nebraska's special populations, ELL and SPED students, continue to demonstrate significant gains. These special populations, however, continue to score lower than total group district

averages. This is consistent with most research in this area and, indeed, the basis for the special programs that are provided to support ELL and SPED students academically. It is important to note that both Nebraska ELL and SPED students increased their scores in reading, math, and writing from 2001 to 2007, contrary to national predictions for these populations.

STARS has impacted the teaching of all students in Nebraska, especially those known as English Language Learners. Between 1990 and 2000, the Hispanic population increased by approximately 10 million people, accounting for 38 percent of the nation's overall population growth during the decade (National Center for Educational Studies, 2008). The issue surrounding Hispanic students, who comprise the largest population to immigrate to Nebraska, is that of the language spoken in the homes of school age children. The parents speak Spanish while the children are asked to learn and speak in English. With the landscape of ELL students changing radically in the last five years, districts have struggled to provide quality education for these students. However, Nebraska educators are meeting the challenges of the education needs for ELL and Special Education students as evidenced by the criterion-referenced assessment data.

## REFERENCES

- National Center for Education Statistics (2003-2008). *Status and trends in the education of the Hispanics*. Retrieved from <http://nces.ed.gov/pubs2003/2003008.pdf>
- Nebraska Department of Education, (2006-2007). *Nebraska Department of Education Directory*, Lincoln, NE. 109<sup>th</sup> Edition, 158.

**CHARTING STARS:**  
***Engaging Conversations***  
**Section 3: Research Studies**

**Study VI: Student Mobility Effects on Achievement  
Across Levels of Poverty**



**Student Mobility Effects on Achievement Across Levels of Poverty**

**Jody Isernhagen, Ed.D., University of Nebraska-Lincoln**

**Nino Zhvania, M.A., University of Nebraska-Lincoln**

**Casey Tallent, M.A., Graduate Student, University of Nebraska-Lincoln**

**INTRODUCTION**

Student mobility, otherwise known as “pupil turnover,” “transience,” or “turbulence” (Demie, 2002, p 199) can be described as a “non-promotional school change” (Rumberger, Larson, Ream & Palardy, 1999, p vi) or an “inconsistency or interruption in the educational experience” (Fisher, Matthews, Stafford, Nakagawa & Durante, 2002, p 319). In essence, however, the phenomenon can be defined as an unscheduled classroom entrance or exit made by students within or between academic years (Texas Educational Agency, 1997).

The magnitude of student mobility in educational settings is a considerable problem. Long (1992) found that students in the United States are more mobile than their peers in well-industrialized Western European countries and in Japan. In its report to the House of Representatives, the General Accounting Office (1994) uncovered that approximately one out of six third-graders had changed schools more than three times since the beginning of the first grade. In the most recent study, Rumberger (2003) reported that nationwide a considerable number of fourth, eighth, and twelfth graders had changed schools at least once within a two-year span.

Mobility patterns observed across grades create even a graver picture. Studying school movements of elementary students in Chicago, Kerbow (1996) unveiled that mobility

disrupted initial cohorts of students over time. Examining student transience in another elementary school in Los Angeles, Bruno, and Isken (1996) projected that only 38% of the initially enrolled students would be able to “survive from Year One to Year Six and receive the full instructional program offered by the school site (p 245).” Though yielding controversial results, research has well documented the impact of mobility on student achievement. Looking into the effect of mobility on student performance, a large number of studies have found this effect to be negative (Brent & Diobilda 1993; GAO, 1994; Mao, Whitsett, & Mellor, 1998; Reynolds, 1991). Contrastingly, other studies have discovered that mobility had no significant independent effect on students’ academic performance (Alexander, Entwisle, & Dauber, 1996; Heinlin & Shinn, 2000). According to the evidence from this research, it was low socio-economic status and a gap in prior achievement that negatively influenced mobile students’ academic performance.

Mobility was also found to have a negative impact on teachers and classrooms. Research indicates that teachers perceive mobility as a major barrier that prevents students from succeeding (Bruno & Isken, 1996; Kerbow, 1996; Lash & Kirpatrick, 1990). Teachers in highly mobile classes blamed mobility for their inability to effectively preserve the learning environment and deliver quality instruction (Bruno & Isken, 1996; Lash & Kirpatrick, 1990; Kerbow, 1996).

Naturally, such effects on students, teachers, and classrooms will be reflected in overall school performance. Several research reports and findings paint a portrait of schools afflicted with high rates of mobility. They serve predominantly disadvantaged students – those from minority groups and those from a low socio-economic background. Classrooms in such schools are overcrowded and the majority of students fail to perform satisfactorily in their academic subjects (Mao, Whitsette, & Mellor, 1998; Nakagawa, Stafford, Fisher, & Matthews, 2002; Rumberger, Larson, Ream, & Palardy, 1999; Texas Education Agency, 1997).

Significant correlation between poverty status, school performance and mobility were established in Offenbergs’ study (2004). According to its results, students with average academic performance tended to move from low to high and from high to low achieving schools. The former pattern was especially true for students with lower achievement and high socioeconomic status (SES) while the latter was more common among students with lower achievement and lower SES. Offenbergs (2004) also detected that students with higher SES and higher academic achievement tended to enter schools with the same level of achievement as the ones from which they had exited.

Based on these results and the premise that mobility can increase an inflow of high achieving and an outflow of low achieving students, Offenbergs (2004) assumes that any given level of school performance might not be necessarily attributed to school characteristics (e.g. highly qualified teachers, well developed teaching and learning programs, school policies, etc.) but be a direct result of student mobility. To demonstrate the implications of this theory, Offenbergs provides a possible scenario of a small school that serves predominantly poor students but which, contrary to expectations, demonstrates an unusually high level of academic success. It is not absolutely groundless, maintains researcher Offenbergs, to think that the school is achieving because

its low performing students leave. Though assumed, such a correlation between the school's overall performance and mobility lacks empirical documentation.

## **PURPOSE OF THE STUDY**

The purpose of this research study was to determine the influence of student mobility on the overall performance of districts. Specifically, the study looks into how different levels of student mobility affect the performance of districts with varying levels of poverty.

### **Research Questions**

*Central Question:* Does the interaction of mobility and poverty affect district achievement?

*Primary Question:* What impact does student mobility have on districts across levels of poverty?

*Secondary Questions:*

1. How much of district performance is explained by the interaction of mobility and poverty?
2. What effect does mobility have on district performance?
3. What effect does poverty have on district performance?
4. Do high mobility rates persistently lead to low district performance across all levels of poverty?

## **REVIEW OF THE LITERATURE**

### **Magnitude of Student Mobility**

The phenomenon of student mobility has been a notorious problem in the world of schooling in the United States. National reports and research on the performance of national samples of students have long been alerting parents, educators, and policy makers about its seriousness. Based on the 1981 census data and population surveys, Long (1992) maintained that students in the United States were more mobile than their peers in other industrialized countries as Great Britain, Ireland, Belgium, and Japan. Young students were particularly unstable when they experienced a larger number of changes than their peers in the elementary years. Using the 1987 census data from the United States Bureau of the Census, Berg-Cross and Flanagan (1988) reported that nearly all the 45 million residential movers, comprising 20% of the country's population, 23 million (approximately 51%) were school-aged students. Based on the data collected in 1991, the General Accounting Office (1994) found that in the national sample of 15,000 third graders, nearly one out of six students had changed schools more than three times since the beginning of the third (first) grade. The data from the National Assessment of Educational Progress (NAEP) reported by Rumberger (2003) in his most recent study that nationwide 34% of fourth graders, 10% of eighth graders and 10% of twelfth graders transferred at least once during the two-year period prior to the NAEP math testing in 1998.

Longitudinal studies give even more detailed patterns of student mobility. Having observed the performance of 767 randomly selected elementary school students in Baltimore for five years, Alexander, Entwisle, and Dauber (1996) reported an escalating pattern of student mobility across grade levels. Specifically, 3% (22 students) of the original sample of 767 students left their schools before the second grade, 7% (52 students) left before the third

grade and 9% (62 students) left before the fourth grade. The tendency towards a gradual meltdown of an original class of students was also detected in the study of Chicago elementary schools conducted by Kerbow (1996). It was uncovered that approximately 19.5% of first graders would no longer continue to the second grade. By the end of the fourth grade, there would be only 46% of the students initially enrolled in the schools. Studying the same schools, but focusing on a cohort of 1,087 African American students from the first through the seventh grades, Temple and Reynolds (1999) identified 73% of their sample as mobile and only 23% as stable.

### **Mobility Effect on Students' Academic Achievement**

Research reports of the effect of mobility on student achievement, mostly measured in test scores in math and reading, are most controversial. Some of them testify that the impact is negative. Others reveal that it is either insignificant or positive. Adding to the controversy of evidence, some other studies prove that the effect of mobility either positive or negative depends on various interfering factors.

A significant number of research studies have documented that mobility has a negative effect on students' academic success (Audette, Algozzine, & Warden, 1993; Benson, Haycraft, Steyaert, & Weigel, 1979; Benson & Weigel, 1981; Brent & Diobilda, 1993; Bruno & Isken 1996; General Accounting Office, 1994; Levine, Wesolowski, & Corbett, 1966; Mao, Whitsett, & Mellor, 1998; Rumberger, Larson, Ream, & Palardy, 1999; Schuller, 1990). This effect is so strong that it impacts students regardless of their socio-economic status, minority status, prior achievement, and parental support (Engec, 2006; Ingersoll, Scamman, & Eckerling, 1989; Kerbow, 1996; Reynolds, 1991; Texas Education Agency, 1997).

Most notably, research reports testify that an achievement gap between mobile and non-mobile students is irreparable (Texas Education Agency 1997). It was found that the cause of this problem lies in the fact that mobile students make academic progress slower and lose knowledge quicker than their non-mobile peers (Mao, Whitsett, & Mellor, 1998; Texas Education Agency, 1997). Studies have also revealed that students are highly unlikely to compensate for their knowledge gap because their knowledge deficiency increases every consecutive year (Reynolds, 1991).

Contradicting these findings, other studies have failed to find the negative effect of mobility on student achievement. They argue that mobile students failed academically not because of their mobility but because of something else. Some found students' IQ, socio-economic and minority status to be the main culprits of mobile students' underachievement (Alexander, Entwisle, & Dauber, 1996; Morris, Pestaner, & Nelson, 1967). Others maintained that the problem of mobile students' underperformance was caused by these students' underachievement that truly existed before the move and only aggravated afterwards (Blane, 1985; Heinlein & Shinn, 2000; Temple & Reynolds, 1999).

### **Mobility Effect on Teachers and Classrooms**

Students are not the only subjects of harmful mobility consequences. Teachers experience the negative impact as well. Disappointingly, they demonstrate frustration and hopelessness while teaching mobile students. There are "no benefits of working with children who move," they say (Lash & Kirkpatrick, 1990, p 185). Exploring the problem of mobility from

teachers' perspectives, studies have found that teachers rarely know in advance about how many new students will enter their classrooms during a school year and how many more will exit them before the last day of school (Bruno & Isken, 1996; Lash & Kirkpatrick, 1990). Such unexpected classroom changes make it difficult for teachers to adjust and deliver quality instruction. As teachers admitted, the most frequently used instructional technique was a constant and frequent revision of covered material. This allowed them to compensate for a difference in their mobile and non-mobile students' knowledge. However, such an approach was not effective when covered content was extensive in its volume. In this situation, leveling the diversified knowledge was hopeless. Adding to the difficulty, teachers also had to "build a new classroom environment and community . . . several times each year with the net result that instructional time and instructional continuity for all children were lost" (Bruno & Isken, 1996, p 247).

### **Mobility Effect on Schools**

Needless to say, having a dire effect on students and teachers, student mobility will have a considerable effect on schools. Trying to uncover the factors that explain achievement differences between schools, Alspaugh (1992) found that student mobility explained 40% of achievement in math and 50% of achievement in reading. Other studies have revealed that mobility negatively effects school achievement – the higher the mobility rate, the greater the possibility a school will show low achievement data (Audette, Algozzine, & Warden, 1993; Texas Education Agency, 1997).

### **Context of the Problem**

Today the average rates of student mobility in Nebraska public schools have slowly decreased from 13.89% in 2003-2004 to 13.82% in 2004-2005, and further to 13.77% in 2005-2006 (Nebraska Department of Education Website, 2008). These statistics reveal that the decrease is incremental statewide and the rates remain at a stable high roughly approaching an average of 14%. However, a large number of Nebraska schools report mobility at a much higher percentage than the state average.

It is natural that the high rates of student mobility put the issue of student transience on agenda in many schools and districts. As it has already been mentioned in the Literature Review of the study, there is empirical evidence revealing a negative influence of mobility on student achievement. However, this evidence is contradictory as some other research studies have failed to find such a negative effect of mobility. In order to determine the true nature of the relationship between mobility and student achievement, more empirical evidence is required. The present research study was undertaken in order to identify evidence of the effect of mobility on student achievement in Nebraska.

The research attempts to examine the relationship between mobility, poverty and district performance. The effects of district reported mobility and poverty on both criterion-referenced tests and norm-referenced tests of achievement are examined over a three year time period, due to the dynamic nature of this phenomenon. At the time that this study was conducted, only district aggregate data was available for poverty, mobility, and achievement in the state. Thus, the research aims at studying the influence of student mobility on the achievement of districts with varying levels of poverty.

## Definitions

- *Student Mobility*  
Student mobility is a “non-promotional school change” (Rumberger, Larson, Ream, & Palardy, 1999, p vi) or an “inconsistency or interruption in the educational experience” (Fisher, Matthews, Stafford, Nakagawa, & Durante, 2002, p 319). In essence, however, the phenomenon can be defined as an unscheduled classroom entrance or exit made by students within or between academic years (Texas Educational Agency, 1997).
- *Student Mobility Rates*  
The research uses the formula applied by the Nebraska Department of Education (2001). Specifically, the rates are calculated according to the following definition: ‘Any child who enters or leaves school between the last Friday in September and the last day of school is counted in the mobility rate. This would include students who transfer into a district and within a district, homebound students, students contracted to other agencies, etc. An individual child is counted only once’.
- *School/District Performance*  
According to the Nebraska Department of Education, school district overall performance is measured by student achievement on both standards-based criterion-referenced assessment (STARS Assessment) and a norm-referenced instrument. According to the demonstrated student proficiency of schools/districts, school performance is rated on a five-point scale of Excellent, Very Good, Good, Needs Improvement, and Unacceptable. Schools/Districts are rated as Excellent, Very Good, and Good if they meet the standard performance expectations set by the state. Schools/Districts are rated as Needs Improvement and Unacceptable if demonstrated student achievement is below the state standards.
- *Criterion-referenced Tests (CRT)*  
An achievement test is regarded to be criterion-referenced if it measures students’ knowledge of a subject matter. Performance on these tests demonstrates how well students have mastered content. As an assessment tool, the State of Nebraska has adopted STARS Assessment – criterion-referenced tests developed by Nebraska teachers. The criterion-referenced score was the district average percent of students meeting the proficiency level or better defined by the local district for their locally developed measure. These tests are administered in grades four, eight, and eleven and encompass reading, writing, and math (Gallagher, 2007).
- *Norm-referenced Tests (NRT)*  
An achievement test is regarded to be norm-referenced if it measures students’ achievement relative to other students. Performance on these tests aims at ranking students according to their achievement scores. In the State of Nebraska, schools administer one of five tests (e.g., Terra Nova, Iowa Test of Basic Skills (ITBS), the Metropolitan Achievement Test (MAT), etc.).

## METHODS

### Data Collection

This study will use quantitative data, reported by the state, to examine the impact of student mobility on district performance. Additionally, this study addresses the consistency of the claim that high student mobility persistently leads to low district achievement. Also, the data

will serve to reveal the statistical significance of the difference in performance between districts with the same high level of student mobility. The data will be used to explore factors other than student mobility that might possibly contribute to overall district performance under the conditions of the same high rates of student mobility. Specifically, poverty will be chosen as such a factor. By using poverty level as a second variable, the effect on mobility across different poverty levels on achievement will be examined. By adding a second variable we will be able to reduce the error variance.

### **Sampling**

Data from 212 school districts in Nebraska were used. Data were gathered by the Nebraska Department of Education (NDE) and was provided to the researchers to conduct the analysis. At the time of this study, the NDE only collected data in the form of district aggregates. The two independent variables are student mobility and poverty level. Student mobility coefficient is calculated according to the statewide formula used at the NDE. Specifically, student mobility is perceived as the ratio of all students who enter or exit any particular school/district between the last Friday in September and the last day of school and the overall school/district population. Mobility was divided into three levels based on the state average of 13.77 %; districts with a mobility rate of 14 % or higher were placed in the high mobility group, districts with a mobility rate between 9 % and 13.99 % were placed into the average mobility group, where districts with a mobility rate below 9 % were placed in the low mobility group.

Poverty was measured using the reported percentage of students in each district who qualify for free-and-reduced lunch benefits (FRL). Poverty was broken into four levels (high, high average, low average, and low) based on the state average for FRL benefits. High poverty districts were those with at least 50% of the student population was eligible for FRL benefits; high-average poverty districts were those where 35% to 49.99% of the student population qualified for FRL benefits; low-average poverty districts were those where 20% to 34.99% of the student population qualified for FRL benefits; and low poverty districts were those with fewer than 19.99% of students qualifying for FRL benefits. The district performance variable was measured according to student performance on criterion-referenced assessment (STARS Assessment) and norm-referenced (ITBS, Terra Nova, CAT, MAT) tests in 8<sup>th</sup> grade math. The data on both variables was analyzed longitudinally across three consecutive school years of 2003-2004, 2004-2005, and 2005-2006.

### **Data Analysis**

The descriptive analysis is used to determine whether high mobility rates persistently cause low academic achievement in schools/districts across levels of poverty. It also determines whether districts with the same high rates of student mobility demonstrate different levels of overall performance on criterion-referenced assessment and norm-referenced tests across levels of poverty. The quantitative analysis is conducted through ANOVA and SPSS computer programs.

## **RESULTS**

A series of 3x4 fixed-effect ANOVA's were used to examine the effects of the two independent variables: mobility (high, average, and low) and poverty (high, high average, low average, and low) on achievement during a three year time span. Poverty data was

collected using district averages for free-and-reduced lunch rates. District performance on both norm-referenced tests and criterion-referenced tests were examined. Six separate studies were conducted to measure the effects of mobility on norm-referenced achievement scores and criterion-referenced achievement scores across three consecutive years. Out of the six ANOVA's, statistically significant interactions were observed between mobility and poverty during four of the conditions. Mobility and poverty had a significant interaction on criterion-referenced achievement during the 2003-2004 school year, but not for the 2004-2005 school year or the 2005-2006 school year. Mobility and poverty had a significant interaction on norm-referenced achievement for all three years.

### **ANOVA I**

A study measuring the effects of mobility across levels of poverty on achievement, measured using a criterion-referenced test in 2003-2004, was conducted. Prior to inspecting the ANOVA results for the 2003-2004 school year for criterion-referenced based achievement, it was determined whether or not the homogeneity of variance assumption was tenable. This was accomplished by examining Levene's test of homogeneity of variance. Levene's test indicated a statistically non-significant difference between the three group variances,  $F = .816, p = .624$ , which suggests that the assumption has been met. A statistically significant interaction was observed between mobility and school size on criterion-referenced achievement in 2003-2004,  $F = 2.650, p = .017$ . The partial omega squared effect size for the interaction effect ( $\omega^2_p = .05$ ) indicated that approximately 5% of the variability in achievement can be explained by the interaction between the two independent variables. Using Cohen's (1988) benchmark values, this represents a small effect size, but is very close to a medium effect size. Following the significant interaction, simple main effects tests were conducted to examine differences among the three mobility groups for each level of poverty. These analyses indicated a statistically significant difference among poverty levels only for the high mobility group,  $F = 5.745, p = .001, \omega^2_p = .07$ . No significant differences were observed for the low and average mobility groups:  $F = 1.147, p = .331, \omega^2_p = 0.00$ , and  $F = .216, p = .885, \omega^2_p = 0.01$ , respectively.

Tukey's HSD *post hoc* tests were conducted and yielded two statistically significant pairwise differences. Within the high mobility group, districts reporting high poverty performed significantly worse than those districts reporting high-average poverty ( $p = .005$ ) and those reporting low-average poverty ( $p < .001$ ).

### **ANOVA II**

A study measuring the effects of mobility across levels of poverty on achievement, measured using a criterion-referenced test in 2004-2005, was conducted. Prior to inspecting the ANOVA results for the 2004-2005 school year for criterion-referenced based achievement, it was of interest to determine whether or not the homogeneity of variance assumption was tenable. This was accomplished by examining the Levene's test of homogeneity of variance. Levene's test indicated a statistically non-significant difference between the three group variances,  $F = 1.320, p = .215$ , which suggests that the assumption has been met. No significant interactions were observed between mobility and school size on criterion-referenced achievement in 2004-2005,  $F = .755, p = .606, \omega^2_p = 0.00$ . However, a significant main effect for poverty was found,  $F = 3.241, p = .023, \omega^2_p = 0.03$ . Follow-up pairwise

comparisons then revealed statistically significant differences between the low-average poverty groups and the high-average ( $p = .005$ ) and high poverty groups ( $p = .017$ ). This means that districts in the low-average range performed better on criterion-referenced assessments than districts in the high-average and high poverty groups. However, there was no significant difference between low and high poverty.

### **ANOVA III**

A study measuring the effects of mobility across levels of poverty on achievement, measured using a criterion-referenced test in 2005-2006, was conducted. Prior to inspecting the ANOVA results for the 2005-2006 school year for criterion-referenced based achievement, it was of interest to determine whether or not the homogeneity of variance assumption was tenable. This was accomplished by examining the Levene's test of homogeneity of variance. Levene's test indicated a statistically significant difference between the group variances,  $F = 2.321$ ,  $p = .011$ , which suggests that the assumption has not been met. However, ANOVA is robust to homogeneity of variance problems and, therefore, the results can still be interpreted. No significant interactions were observed between mobility and school size on criterion-referenced achievement in 2005-2006,  $F = 1.977$ ,  $p = .071$ ,  $\omega_p^2 = 0.03$ . However, a significant main effect for mobility was found,  $F = 4.754$ ,  $p = .010$ ,  $\omega_p^2 = 0.04$ . Follow-up pairwise comparisons then revealed statistically significant differences between the low mobility groups and the average ( $p = .023$ ) and high mobility groups ( $p = .012$ ). This means that districts in the low mobility group performed better on criterion-referenced achievement assessments than districts in the average and high mobility groups.

### **ANOVA IV**

A study measuring the effects of mobility across levels of poverty on achievement, measured using a norm-referenced test in 2003-2004, was conducted. Prior to inspecting the ANOVA results for the 2003-2004 school year for norm-referenced based achievement, it was of interest to determine whether or not the homogeneity of variance assumption was tenable. This was accomplished by examining the Levene's test of homogeneity of variance. Levene's test indicated a statistically significant difference between the three group variances,  $F = 2.586$ ,  $p = .004$ , which suggests that the assumption has not been met and interpretations should be made with caution. However, ANOVA is robust to homogeneity of variance errors and, therefore, interpretations will still be presented. A statistically significant interaction was observed between mobility and school size on norm-referenced achievement in 2003-2004,  $F = 4.041$ ,  $p = .001$ . It should be noted that the main effect for poverty was also statistically significant ( $F = 8.149$ ,  $p < .001$ ,  $\omega_p^2 = .10$ ) and the main effect for mobility was also statistically significant ( $F = 8.279$ ,  $p < .001$ ,  $\omega_p^2 = .08$ ), but these main effects were not interpreted due to the significant interaction.

The partial omega squared effect size for the interaction effect ( $\omega_p^2 = .09$ ) indicated that approximately 9% of the variability in achievement can be explained by the interaction between the two independent variables. Using Cohen's (1988) benchmark values, this represents a medium effect size. Following the significant interaction, simple main effects tests were conducted to examine differences among the three mobility groups for each level of poverty. These analyses indicated a statistically significant difference among poverty levels only for the high mobility group,  $F = 13.203$ ,  $p < .001$ ,  $\omega_p^2 = .15$ . No significant

differences were observed for the low and average mobility groups:  $F = 0.475, p = .700, \omega^2_p = 0.01$ , and  $F = .372, p = .773, \omega^2_p = 0.01$ , respectively.

Tukey's HSD *post hoc* tests were conducted and yielded three statistically significant pairwise differences. Within the high mobility group, districts reporting high poverty performed significantly worse than those districts reporting high-average poverty ( $p < .001$ ), those reporting low-average poverty ( $p < .001$ ), and those reporting low poverty ( $p = .002$ ). This means that districts with high mobility and high poverty had significantly worse achievement scores than districts with lower poverty rates.

## ANOVA V

A study measuring the effects of mobility across levels of poverty on achievement, measured using a norm-referenced test in 2004-2005, was conducted. Prior to inspecting the ANOVA results for the 2004-2005 school year for norm-referenced achievement, it was of interest to determine whether or not the homogeneity of variance assumption was tenable. This was accomplished by examining the Levene's test of homogeneity of variance. Levene's test indicated a statistically significant difference between the three group variances,  $F = 4.958, p < .001$ , which suggests that the assumption has not been met and interpretations should be made with caution. However, ANOVA is robust to homogeneity of variance errors; therefore, interpretations will still be presented. A statistically significant interaction was observed between mobility and school size on norm based achievement in 2004-2005,  $F = 2.712, p = .015$ . It should be noted that both the main effect for poverty was statistically significant ( $F = 9.341, p < .001, \omega^2_p = .11$ ) and the main effect for mobility was statistically significant ( $F = 7.734, p = .001, \omega^2_p = .06$ ), but these main effects were not interpreted due to the significant interaction.

The partial omega squared effect size for the interaction effect ( $\omega^2_p = .05$ ) indicated that approximately 5% of the variability in achievement can be explained by the interaction between the two independent variables. Using Cohen's (1988) benchmark values, this represents a small effect size; however this is close to a medium effect size. Following the significant interaction, simple main effects tests were conducted to examine differences among the three mobility groups for each level of poverty. These analyses indicated a statistically significant difference among poverty levels for the high mobility group,  $F = 6.542, p < .001, \omega^2_p = .08$ , and average mobility group,  $F = 4.124, p = .007, \omega^2_p = .04$ . No significant differences were observed for the low mobility group:  $F = 0.907, p = .439, \omega^2_p = 0.00$ .

Tukey's HSD *post hoc* tests were conducted and yielded three statistically significant pairwise differences. Within the high mobility group, districts reporting high poverty performed significantly worse than those districts reporting high-average poverty ( $p = .001$ ), those reporting low-average poverty ( $p < .001$ ), and those reporting low poverty ( $p = .006$ ). This means that districts with high mobility and high poverty had significantly worse achievement scores than districts with lower poverty rates. Within the average mobility group, districts reporting high poverty performed significantly worse than those districts reporting high-average poverty ( $p = .002$ ), those reporting low-average poverty ( $p = .001$ ), and those reporting low poverty ( $p = .025$ ). This means that districts with average mobility

and high poverty had significantly worse achievement scores than districts with lower poverty rates.

## **ANOVA VI**

A study measuring the effects of mobility across levels of poverty on achievement, measured using a norm-referenced test in 2005-2006, was conducted. Prior to inspecting the ANOVA results for the 2005-2006 school year for norm-referenced achievement, it was of interest to determine whether or not the homogeneity of variance assumption was tenable. This was accomplished by examining the Levene's test of homogeneity of variance. Levene's test indicated a non-statistically significant difference between the three group variances,  $F = 1.664$ ,  $p = .084$ , which suggests that the assumption has been met and interpretations can be made. A statistically significant interaction was observed between mobility and school size on norm based achievement in 2005-2006,  $F = 4.792$ ,  $p < .001$ . It should be noted that both the main effect for poverty was statistically significant ( $F = 5.931$ ,  $p = .001$ ,  $\omega^2_p = .07$ ) and the main effect for mobility was statistically significant ( $F = 5.971$ ,  $p = .003$ ,  $\omega^2_p = .05$ ), but these main effects were not interpreted due to the significant interaction.

The partial omega squared effect size for the interaction effect ( $\omega^2_p = .10$ ) indicated that approximately 10% of the variability in achievement can be explained by the interaction between the two independent variables, mobility and poverty. Using Cohen's (1988) benchmark values, this represents a medium effect size. Following the significant interaction, simple main effects tests were conducted to examine differences among the three mobility groups for each level of poverty. These analyses indicated a statistically significant difference among poverty levels for the high mobility group,  $F = 12.047$ ,  $p < .001$ ,  $\omega^2_p = .14$ , which indicates that approximately 14% of the variability in achievement can be explained by the interaction between high mobility and poverty. Using Cohen's (1988) benchmark values, this represents a large effect size. No significant differences were observed for the low or average mobility group:  $F = 0.439$ ,  $p = .725$ ,  $\omega^2_p = 0.01$  and  $F = 0.869$ ,  $p = .458$ ,  $\omega^2_p = 0.00$ , respectively.

Tukey's HSD *post hoc* tests were conducted and yielded five statistically significant pairwise differences. Within the high mobility group, districts reporting high poverty performed significantly worse than those districts reporting high-average poverty ( $p < .001$ ), those reporting low-average poverty ( $p < .001$ ), and those reporting low poverty ( $p < .001$ ). This means that districts with high mobility and high poverty had significantly worse achievement scores than districts with lower poverty rates. Additionally, within the high mobility group, districts reporting high-average poverty performed significantly worse than those districts reporting low-average poverty ( $p = .049$ ) and those reporting low poverty ( $p = .041$ ). This means that districts with high mobility and high-average poverty had significantly worse achievement scores than districts with lower poverty rates, but performed better on norm-referenced achievement tests than districts with high poverty.

## **Summary of Findings**

In summary, results from the effects of mobility on criterion-referenced achievement across levels of poverty varied greatly from year to year. During 2003-2004, a significant interaction was found between high mobility and poverty, with high mobility, high poverty districts performing worse than schools falling in the high-average and low-average poverty

ranges. During 2004-2005, there was no significant interaction, but rather a significant main effect for poverty with districts in the low-average poverty group performing better on criterion-referenced achievement assessments than districts with high-average or high poverty. For 2005-2006, no significant interaction was found, but rather a significant main effect for mobility, with low mobility schools performing better on criterion-referenced achievement tests than districts in the average and high mobility groups.

Studies IV-VI displayed more similarities for the effects of mobility on norm-referenced achievement tests across levels of poverty. Significant interactions for high mobility and high poverty were found each year from 2003-2004 to 2005-2006. Districts with high mobility and high poverty performed worse on norm-referenced achievement tests than districts with lower poverty all three years. Additionally, in 2004-2005 districts with average mobility and high poverty performed worse on norm-referenced achievement tests than districts with lower rates of poverty.

## **DISCUSSION**

Based on the above given statistical results, the findings of the study can be explained as follows. The effect of mobility and poverty on districts' performance on criterion-referenced assessments varies. In 2003-2004 there was an interaction between mobility and poverty on achievement. However, in the consecutive years of 2004-2005 and 2005-2006, there was no interaction between the two variables, but were differing significant main effects on poverty or mobility on achievement. Specifically, in 2004-2005 achievement was influenced by only poverty while in 2005-2006 it was affected only by mobility. Further, in 2003-2004 regardless of poverty levels, districts with low mobility rates performed equally well on the achievement tests. The same was true for districts with average levels of mobility – notwithstanding different levels of poverty they all displayed the same level of achievement. The only difference in achievement according to the level of poverty was demonstrated by districts with high rates of mobility. Here, districts with high poverty levels performed worse than those with low-average or high-average poverty levels. Additionally, in those districts with high rates of mobility, districts with low, low-average and high-average levels of poverty demonstrated similar achievement scores. There was no significant difference between the high poverty and low poverty districts within the high mobility group. The patterns changed in 2005-2006, districts with low mobility rates performed better than districts with average and high mobility rates. Interestingly, there was no significant difference in the performance of districts with average and high mobility rates.

Mobility and poverty demonstrated a more persistent pattern of influence regarding districts' achievement on the norm-referenced tests. However, despite this relative consistency of the effect, mobility still revealed variability. Specifically, in 2003-2004 regardless of poverty levels, there was no significant difference in the achievement of districts within either low or average mobility rate groups. The only difference in achievement was demonstrated by districts with high mobility rates. In this group, districts with low, low-average and high-average poverty levels performed equally well on the achievement test. Only the districts with high levels of poverty performed significantly worse than districts with lower levels of poverty. In 2004-2005 only the districts with low mobility rates managed to maintain its persistence of performance across poverty levels. Districts with average and high levels of mobility differed in their achievement according to poverty levels. In the group of districts

with average mobility rates, districts in low, low-average or high-average poverty level did not differ significantly in achievement. During the academic year 2005-2006, schools with low and average mobility rates demonstrated a consistency of achievement scores across level of poverty. Within these groups, districts with various levels of poverty demonstrated similar achievement. More diversified performance was displayed by districts with high mobility rates. In high mobility districts, the only districts that performed equally well were those with low and low-average levels of poverty. In other cases, high mobility districts' achievement displayed a decline with an escalation of poverty level. Specifically, low poverty districts performed better than districts with a high-average poverty level; high-average poverty districts performed better than districts with high levels of poverty; high poverty districts displayed the worst achievement scores.

Thus, our research has detected the following. First, analysis of the results showed that the mobility and poverty in combination have an effect on districts' achievement. The effect of mobility or poverty alone is inconsistent. Second, high mobility rates do not necessarily lead to low achievement. The data demonstrate that districts with high rates of mobility might perform worse than districts with either low or average rates of mobility; however, this does not rule out the chance for these districts to perform as well as those districts with average or low mobility rates. It is maintained that even when mobility rates remain constant, achievement levels vary. Finally, the study has revealed that poverty and mobility may interact differently to effect criterion-referenced achievement versus norm-referenced achievement. There are likely additional factors that account for the difference in interaction with criterion-referenced achievement and norm-referenced achievement.

These findings partially corroborate the findings of previous research. Specifically, our results converge with other studies that have found the combination of mobility and poverty to have a strong negative impact on academic achievement (Ingersoll, Scamman, & Eckerling, 1989; Kerbow, 1996; Texas Education Agency, 1997). The findings are also in accordance with another pool of research that found mobility as an independent factor effecting academic success (Audette, Algozzine, & Warden, 1993; Benson, Haycraft, Steyaert, & Weigel, 1979; Benson & Weigel, 1981; Brent and DiObilda, 1993; Bruno & Isken 1996; General Accounting Office, 1994; Levine, Wesolowski, & Corbett, 1966; Mao, Whitsett, & Mellor, 1998; Rumberger, Larson, Ream, & Palardy, 1999; Schuller, 1990). Finally, some findings of our research correspond with the body of research suggesting that mobility has little to no effect on academic performance (Alexander, Entwisle, & Dauber, 1996; Blane, 1985; Heinlein & Shinn, 2000; Temple & Reynolds, 1999; Nelson, Simoni, & Adelman, 1990).

However, the present research disagrees with the evidence found in mobility research literature in two ways. First, studies have found that a combination of mobility and other factors, such as poverty persistently lead to low achievement (Kerbow, 1996). Our study found that in most cases poverty had no impact on the performance of districts where mobility rates were low and average, yet found consistent results across high mobility districts. Second, the previous research claims that high levels of mobility are consistently associated with low achievement (Mao, Whitsett, & Mellor, 1998; Texas Education Agency, 1997). Our study has found that even districts with high rates of mobility are capable of

performing as well as districts with average or low rates of mobility, especially in schools with low to average poverty levels.

There are several limitations inherent in our study. First, due to the aggregate nature of the data, it was not possible to examine the interaction of poverty and mobility on achievement with mobile students only. The study can only be generalized to district mobility, poverty, and achievement. Perhaps the use of individual student data may have clarified some of the inconsistencies. Furthermore, it was difficult to eliminate the effects of potential confounding variables with the use of aggregate data. Achievement data may also be positively skewed based on the additive nature of rounding scores. Additionally, this does not provide insight into specific district factors that may impact performance; student's affected by mobility such as advantageous district policies or classroom practices.

### **IMPLICATIONS**

Strong implications for education policy can be proposed based on the results of the present study. First, mobility in schools and districts can not be perceived as a problem that can only be solved beyond school walls and district limits. The ability to cope with the effects of mobility on achievement can be managed based on the inconsistent impact of mobility found in this study. Second, the effects of poverty and mobility, though considerable factors of student achievement, are not definitive in districts' achievement. Their effect is likely to vary depending on districts themselves - their programs, policies and practices. Thirdly, districts with high poverty and high levels of mobility should evaluate their policies and practices in working with mobile students and should investigate the effects of mobility in classrooms. Perhaps individualized attention to this issue could serve to minimize the negative effects that high mobility and high poverty have on achievement in combination. Finally, in districts with average rates of mobility a slightly stronger focus should be given to the issue of mobility. This implication is drawn from the empirical evidence revealing high chances of these districts to significantly reduce the effect of mobility of achievement and start performing on the same level as do districts with very low mobility rates. Districts with high mobility rates may be able to diminish any negative effects provided that suitable interventions, policies and practices are introduced.

In future research specific attention should be given to the policies and interventions of schools and districts that have high levels of poverty and high mobility rates. Successful strategies and interventions to bridge the achievement gap might be located within isolated schools and districts and have not yet been made public for all schools to implement. Additionally research should be done to determine which strategies are more successful than others, because if poor strategies are not substituted, unsuccessful strategies will create persistent grounds for the mobility problem to linger. Considering the consequences in both cases, there is a dire necessity for revealing both effective and non effective strategies. Also, there is a need to provide teachers, school administrators and policy makers with this information to assist them in making research based and data driven decisions relative to ineffective methods, adoption of successful practices, and the improvement of existing policies.

Thus, future research should assist the districts and schools in finding interventions, policies, and practices that will facilitate success in addressing the mobility problem. In the future,

this study should be replicated using individual student data in order to focus on the achievement of mobile students. Additionally, qualitative studies should be conducted to give a voice to school principals and teachers. These voices will tell true stories of how schools and districts approach the problem of mobility and how they can facilitate achievement among students affected by mobility.

## REFERENCES

- Alexander, K. L., Entwisle, D. R., & Dauber, S. L. (1996). Children in motion: School transfers and elementary school performance. *The Journal of Educational Research, 90* (1), 3-12.
- Alspaugh, J. (1991). Out-of-school environmental factors and elementary-school achievement in mathematics and reading. *Journal of Research and Development in Education, 24*(3), 53-55.
- Audette, R., Algozzine, R., & Warden, M. (1993). Mobility and school achievement. *Psychological Reports, 72*(2), 701-702.
- Benson, G. P., Haycraft, J. L., Steyaert, J. P., & Weigel, D. L. (1979). Mobility in sixth graders as related to achievement, adjustment, and socioeconomic status. *Psychology in the Schools, 16*(3), 444-447.
- Benson, G. P., & Weigel, D. L. (1980). Ninth grade adjustment and achievement as related to mobility. *Educational Research Quarterly, 5*(4), 15-19.
- Berg-Cross, L., & Flanagan, R. (1988). Effects of orientation program on mobile transfer students. *Journal of Early Adolescence, 8*(3), 311-324.
- Blane, D. (1985). A longitudinal study of children's school mobility and attainment in mathematics. *Educational Studies in Mathematics, 16*(2), 127-142.
- Brent, G., & Diobilda, N. (1993). Effects of curriculum alignment versus direct instruction on urban children. *Journal of Educational Research, 86*(6), 333-338.
- Bruno, J. E., & Isken, J. A. (1996). Inter and intraschool site student transience: Practical and theoretical implications for instructional continuity at inner city schools. *Journal of Research and Development in Education, 29*(4), 139-252.
- Christensen, D. (2006). State assessments, national tests: How they are different, what they tell us. Retrieved on March 28, 2008 from [http://www.nde.state.ne.us/COMMISH/121506\\_editorial.pdf](http://www.nde.state.ne.us/COMMISH/121506_editorial.pdf).
- Demie, F. (2002). Pupil mobility and educational achievement in schools: An empirical analysis. *Educational Research, 44*(2), 197-215.
- Engec, N. (2006). Relationship between mobility and student performance and behavior. *Journal of Educational Research, 99*(3), 167-178.
- Fisher, T. A., Matthews, L., Stafford, M. E., Nakagawa, K., & Durante, K. (2002). School personnel's perceptions of effective programs for working with mobile student and families. *The Elementary School Journal, 102*(4), 317-333.
- Gallagher, C. W. (2007). *Reclaiming Assessment: A Better Alternative to the Accountability Agenda*. Portsmouth, NH; Heinemann.
- General Accounting Office. (1994). *Elementary school children: Many change schools frequently, harming their education*. (Report No. GAO/HEH-94-45). Washington, DC; U.S. General Accounting Office. (ERIC Document Reproduction Service No. ED369526.

- Heinlein, L. M., & Shinn, M. (2000). School mobility and student achievement in an urban setting. *Psychology in the Schools, 37*(4), 349-357.
- Ingersoll, G. M., Scamman, J. P., & Eckerling, W. D. (1989). Geographic mobility and student achievement in an urban setting. *Educatoinal Evaluation and Policy Analysis, 11*(2), 143-149.
- Kerbow, D. (1996). Patterns of urban student mobility and local school reform. *Journal of Education for Students Placed at Risk, 1*(2), 147-169.
- Lash, A. A., & Kirkpatrick, S. L. (1990). A classroom perspective on student mobility. *The Elementary School Journal, 91*(2), 176-191.
- Levine, M., Wesolowski, J. C., & Corbett, F. J. (1966). Pupil turnover and academic performance in an inner city elementary school. *Psychology in Schools, 2*(3), 153-158.
- Long, L. H. (1992). International perspectives on the residential mobility of America's children. *Journal of Marriage and the Family, 54*(4), 861-869.
- Mao, M. X., Whitsett, M. D., & Mellor, L. T. (1998). Student mobility, academic performance, and school accountability. *ERS Spectrum, 16*(1), 3-15.
- Morrise, J. L., Pestaner, M., & Nelson, A. (1967). Mobility and achievement. *The Journal of Experimental Education, 35*(4), 74-80.
- Nakagawa, K., Stafford, M. E., Fisher, T. A., & Matthews, L. (2002). The "city migrant" dilemma: Building community at high-mobility urban schools. *Urban Education, 37*(1), 96-125.
- Nebraska Department of Education (2008). 2006-2007 State of the Schools Report. A Report on Nebraska Public Schools. Retrieved from May 22, 2008: <http://reportcard.nde.state.ne.us/Page/DemoMobility.aspx?Level=st>
- Offenberg, R. M. (2004). Inferring adequate yearly progress of schools from student achievement in highly mobile communities. *Journal of Education for Students Placed at Risk, 9*(4), 337-355.
- Reynolds, A. J. (1991). Early schooling for children at risk. *American Educational Research Journal, 28*(2), 392-422.
- Rumberger, R. W. (2003). The causes and consequences of student mobility. *The Journal of Negro Education, 72*(1), 6-21.
- Rumberger, R. W., Larson, K. A., Ream, R. K., & Palardy, G. J. (1999). *The educational consequences for California students and schools. Research Series.* (Report No. PACE-99-2). Berkeley, CA; Stuart Foundation. (ERIC Document Reproduction Service No. ED441040).
- Schuler, D. B. (1990). Effect of family mobility on student achievement. *ERS Spectrum, 8*(4), 17-24.
- Temple, J. A., & Reynolds, A. J. (1999). Mobility and achievement: Longitudinal findings from an urban cohort. *Journal of School Psychology, 37*(4), 355-377.
- Texas Education Agency. (1997). *A study of student mobility in Texas public schools. Statewide Texas educational progress study report No 3.* (Report No. RE7-601-03). Austin, TX; Texas Education Agency. (ERIC Document Reproduction Service No. ED424664).

**CHARTING STARS:  
*Engaging Conversations*  
Section 3: Research Studies**

**Study VII: The Effect of Nebraska's Standards & Accountability System (STARS) on School Improvement Practices**



**Study VII: The Effect of Nebraska's Standards & Accountability System (STARS) on School Improvement Practices**

**Jerald Riibe, Assistant Superintendent for Curriculum and Instruction  
Ralston Public Schools**

**INTRODUCTION**

The initial emphasis for standards and accountability in Nebraska was to guide instruction and promote school improvement. The standards-wide accountability process in Nebraska is the School-based Teacher-led Assessment and Reporting System (STARS). The rationale for standards and the subsequent assessment of student achievement on those standards was to provide a catalyst for school improvement. One measure of the impact of standards and accountability in Nebraska is the role it plays in school-wide conversations regarding school improvement.

The effectiveness of standards for accountability depends greatly on the perceptions of teachers and principals. One perception can be that STARS is primarily a reporting and compliance activity. Another perception could be that STARS is a guide for instruction and school improvement. Schools are left with two contrasting messages regarding standards assessment: standards as a basis for school improvement and standards as a compliance activity. The interpretation and implementation of those messages will define a school's priorities and actions.

Since 2000, all Nebraska public school districts have reported annually on progress toward meeting state standards in math and reading in grades four, eight, and eleven. The passage of *The No Child Left Behind Act* in 2001 (NCLB) resulted in the addition of grades three, five, six, and seven in the reporting of at least one standard in math and reading (Isernhagen, 2007). The reporting grades for federal accountability include the third grade through eighth grade and at least one grade in the high school. Reporting grades for STARS include the fourth, eighth, and eleventh grades. Kindergarten, first, and second grades do not report student achievement to either the Nebraska Department of Education (NDE) for STARS or federal accountability (No Child Left Behind).

Are the intended results of STARS reflected in perceptions of teachers regardless of class assignment? If STARS is a part of a school-wide improvement process to improve instruction and achievement, there should be little difference in perceptions between teachers in reporting grades and non-reporting grades. If STARS is perceived as primarily a compliance activity, teachers may feel removed from the process, unaware and indifferent to the results. If STARS is perceived as a school improvement activity, teachers will relate STARS to school improvement practice in the classroom.

Nebraska teachers were divided into two groups: those that report student achievement results to the NDE and those that do not. Does teacher perception of STARS differ between these groups? A survey to gauge perceptions will give insight into the question of whether STARS is perceived to be primarily about compliance or about school improvement. Second-grade teachers would provide insight into teachers' perceptions of STARS as those who are not directly involved in reporting student achievement results to the NDE. Fourth-grade teachers represent the heart of the STARS process in the elementary grades. Fourth-grade teachers embody the group responsible for all state reading and math standards and subsequent reporting.

## **PURPOSE OF STUDY**

The purpose of this quantitative study was to compare the perceptions of Nebraska second-grade teachers (a non-reporting grade) with perceptions of Nebraska fourth-grade teachers (a reporting grade) regarding the effect of Nebraska's standards accountability system (STARS) on school improvement practices.

## **METHODS**

The population considered relevant to this study consisted of second and fourth-grade teachers in Nebraska public schools. A sample of this population consisted of a minimum sample population of not less than 64 teachers per grade level. This sample size was calculated using *G\*Power 3* statistical software (Buchner, Erdfelder, & Faul, 2007) with the following inputs: an alpha of .05, a moderate effect size estimation of .5, with a power of .8. To insure adequate sample size, 125 surveys were mailed to each grade level in the study. A post hoc analysis using *G\*Power 3* (Buchner et al., 2007) to compute achieved power with an effect of .5, an alpha of .05, and sample sizes of 70 and 74 resulted in an achieved power of .84. The sample population did not use paired samples. The second-grade teachers were selected in isolation of the fourth grade teacher sample. This was done to limit the effects of school system practice on the sample population perceptions. Demographic data regarding

school district size was collected for the purpose of describing the population sample but was not used to compare sub-group responses.

A survey was used to generate quantitative data describing second and fourth-grade teachers' perceptions of Nebraska's state standards process. The survey data was analyzed to discover relationships between variables. A quantitative design is suggested when a problem or issue requires trends to be described and variable relationships be explained (Creswell, 2002). The goal of this research project was to evaluate the mean difference between two populations: second and fourth-grade teachers. The independent-measures *t* statistic was used to draw inferences between two populations (Creswell, 2002). Each research question used a *t* test to measure the statistical difference between the sample populations.

### **Research Questions**

1. Is there a difference in perceptions of second-grade teachers and the perceptions of fourth-grade teachers as to whether or not the implementation of the STARS process has led to improved student achievement?
2. Is there a difference in perceptions of second-grade teachers and the perceptions of fourth-grade teachers as to whether or not the implementation of the STARS process has led to improved school curriculum?
3. Is there a difference in perceptions of second-grade teachers and the perceptions of fourth-grade teachers as to whether or not the implementation of the STARS process has led to improved school climate?
4. Is there a difference in perceptions of second-grade teachers and the perceptions of fourth-grade teachers as to whether or not the implementation of the STARS process has led to improved classroom instruction?
5. Is there a difference in perceptions of second-grade teachers and the perceptions of fourth-grade teachers as to whether or not the implementation of the STARS process has led to improved assessment practice?
6. Is there a difference in perceptions of second-grade teachers and the perceptions of fourth-grade teachers as to whether or not the implementation of the STARS process has led to improved image of the teaching profession?

## **RESULTS**

Data were collected through a survey (FAppendix F) mailed to 125 second-grade teachers and 125 fourth-grade teachers. The survey netted 70 responses (56%) from second-grade teachers and 74 responses (59%) from fourth-grade teachers. The response rate fell within the parameters outlined in the study's methodology. The survey questions were constructed using 5-point Likert scale. The scale breakdown was: 1 - strongly disagree, 2 - disagree, 3 - no opinion, 4 - agree, and 5 - strongly agree. Means were computed for each research question and each subscale question. All subscale questions were without an omitted response. An alpha level of .05 was used for all statistical tests. The data will be presented by each individual research question.

**Research Question 1:** Is there a difference in perceptions of second-grade teachers and the perceptions of fourth-grade teachers as to whether or not the implementation of the STARS process has led to improved student achievement?

**Result:** Statistical analysis of the grouped subscale questions relating to this research question found no statistically significant difference between the second-grade teachers' responses ( $M = 16.86$ ,  $SD = 3.42$ ) and the fourth-grade teachers' responses ( $M = 17.32$ ,  $SD = 3.77$ ) for this research question:  $t(142) = -0.777$ ,  $p = .44$ .

**Research Question 2:** Is there a difference in perceptions of second-grade teachers and the perceptions of fourth-grade teachers as to whether or not the implementation of the STARS process has led to improved school curriculum?

**Result:** Statistical analysis of the grouped subscale questions relating to this research question found no statistically significant difference between the second-grade teachers' responses ( $M = 12.88$ ,  $SD = 3.09$ ) and the fourth-grade teachers' responses ( $M = 13.62$ ,  $SD = 2.94$ ) for this research question;  $t(142) = -1.461$ ,  $p = .15$ .

**Research Question 3:** Is there a difference in perceptions of second-grade teachers and the perceptions of fourth-grade teachers as to whether or not the implementation of the STARS process has led to improved school climate?

**Result:** Statistical analysis of the grouped subscale questions relating to this research question found no statistically significant difference between the second-grade teachers' responses ( $M = 12.08$ ,  $SD = 2.88$ ) and the fourth-grade teachers' responses ( $M = 11.78$ ,  $SD = 3.30$ ) for this research question:  $t(142) = 0.583$ ,  $p = .56$ .

**Research Question 4:** Is there a difference in perceptions of second-grade teachers and the perceptions of fourth-grade teachers as to whether or not the implementation of the STARS process has led to improved classroom instruction?

**Result:** Statistical analysis of the grouped subscale questions relating to this research question found no statistically significant difference between the second-grade teachers' responses ( $M = 10.42$ ,  $SD = 2.13$ ) and the fourth-grade teachers' responses ( $M = 10.65$ ,  $SD = 1.91$ ) for this research question:  $t(138.47) = -0.650$ ,  $p = .52$ .

**Research Question 5:** Is there a difference in perceptions of second-grade teachers and the perceptions of fourth-grade teachers as to whether or not the implementation of the STARS process has led to improved assessment practice?

**Result:** Statistical analysis of the grouped subscale questions relating to this research question found no statistically significant difference between the second-grade teachers' responses ( $M = 10.97$ ,  $SD = 2.11$ ) and the fourth-grade teachers' responses ( $M = 10.70$ ,  $SD = 2.55$ ) for this research question:  $t(139.691) = 0.689$ ,  $p = .49$ .

**Research Question 6:** Is there a difference in perceptions of second-grade teachers and the perceptions of fourth-grade teachers as to whether or not the implementation of the STARS process has led to improved image of the teaching profession?

**Result:** Statistical analysis of the grouped subscale questions relating to this research question found no statistically significant difference between the second-grade teachers'

responses ( $M = 9.1714$ ,  $SD = 2.38$ ) and the fourth-grade teachers' responses ( $M = 9.35$ ,  $SD = 2.45$ ) for this research question:  $t(142) = -0.446$ ,  $p = .66$ .

## DISCUSSION

The purpose of this study was to compare the perceptions of Nebraska second-grade teachers (a non-reporting grade) with the perceptions of Nebraska fourth-grade teachers (a reporting grade) regarding the effects of STARS on school improvement. The study was not intended to provide a judgment on the effectiveness of the STARS process. In fact, the perceptions studied would be relevant to any accountability system that included some grades and not others for reporting.

The research questions involved the impact of an accountability system on school improvement practices. The results of the study present two different perspectives. Teachers in reporting and non-reporting grades did not have a statistically significant difference in how STARS was perceived. This would indicate that the processes in place to meet STARS requirements are not isolated to reporting grades. However, the large standard deviation found within the responses provides another interpretation of the study. This may indicate that there was a lack of common understanding about the STARS process and its effect on school improvement practices among teachers.

### Summary

The data from this study will satisfy neither the detractors of the STARS process or its supporters. The finding of no statistical difference of perspectives in the research questions indicates that STARS is not limited to reporting grades. The similar perspectives of reporting and non-reporting grade teachers indicate a certain universality that is important for any accountability system. If the current system were to be replaced by a high stakes statewide test, the questions asked in this study would be just as relevant to that system.

The Nebraska standards accountability system is not without blemish and concerns. The positive finding is that teachers in this study rarely linked STARS to negative developments in achievement, curriculum, or assessment. STARS has moved teachers, buildings, and districts toward accountability for student achievement. That is important and provides solid footing for future school improvement efforts.

A question raised at the beginning of this study was to determine if STARS was perceived as a compliance activity or a school improvement activity. The answer to that question is not clear. Teachers in reporting and non-reporting grades perceive STARS as a component of school improvement. The response to the research questions also indicated a sense that a common perception of STARS may be isolated to certain classrooms, buildings, and districts.

Nebraska's STARS process has been legislated to resemble other state universal testing systems. The legacy that STARS will leave is not in the assessments created or the accountability reports. STARS may be labeled successful if the idea of blending accountability and school improvement transcends grade levels and content areas. The findings of this study would indicate STARS has had an effect in moving some teachers from isolated classrooms to an accountable and improving school culture.

## **Recommendations**

The single most critical concern found in this study is the large standard deviation found within the responses. This may indicate that there was a lack of common understanding about the STARS process and its effect on school improvement practices among teachers. After ten years of work with the STARS system, there is still a sense that some teachers may lack a global understanding of the process. The inconsistent perception of STARS may indicate a lack of understanding or commitment among second and fourth-grade teachers in Nebraska. This may be due to the lack of administrative leadership in connecting STARS to everyday practice, a lack of understanding among teachers of the STARS process, a lack of total school conversation about STARS, or the confusion that NCLB has created with the myriad of changes to STARS.

The following recommendations are derived from the results of this study.

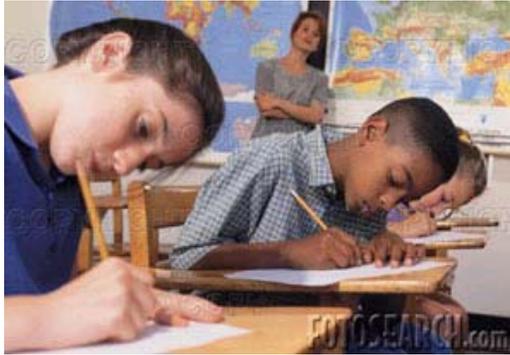
1. A study that investigated perceptions of reporting and non-reporting grade teachers by length of experience, gender, and professional preparation.
2. A study of the affective factors that influence teacher self-image is important to address retention problems, recruitment problems, and public image issues.
3. A thorough examination of systemic district practices that have evolved through STARS is critical. The effective and imbedded STARS processes for school improvement need to be clearly identified and included in the transition to a high stakes testing model.

## **REFERENCES**

- Creswell, J. W. (2002). *Educational research planning, conducting, and evaluating quantitative and qualitative research*. Upper Saddle River, NJ: Merrill.
- Isernhagen, J. C. & Mills, S. (2007). *Charting STARS: Pathways to Success. Year-Six Report. Comprehensive Evaluation of Nebraska's School-based Teacher-led Assessment and Reporting System*. Lincoln, NE: University of Nebraska-Lincoln.

**CHARTING STARS:  
ENGAGING CONVERSATIONS**  
**Section 3: Research Studies**

**Study VIII: Nebraska Educators Review the Local Math  
Assessment Process**



**Study VIII: Nebraska Educators Review the Local Math  
Assessment Process: Reliability of Peer Review of Assessment Portfolios  
2007-2008**

**John L. Moon, Assessment Coordinator  
Nebraska Department of Education**

**INTRODUCTION**

Since 2000, Nebraska School Districts have built a local assessment system to measure student achievement in math and reading that is different from any other state-wide system in the United States. Because the assessment process was developed at the local level and used for state and federal accountability, schools districts have been challenged by the Nebraska Department of Education (NDE) to assemble a district assessment portfolio to describe the development and documentation of their assessment systems. Beginning in 2006, the evaluation of portfolios included a new District Assessment Portfolio Rubric and the utilization of on-site peer reviewers. Therefore, there was a need to examine the validity and reliability of the revised system.

Because quality assessments are necessary for schools to report reliable and valid data on student achievement, this study will examine the impact of the revised rubric and the use of on-site reviewers for the examination of each portfolio. The revised system still uses assessment specialists (3-4), but the information is collected by on-site reviewers and brought back to the assessment specialists for rating.

With a quality assessment system generating quality student data, the impact of assessment information on student learning and school improvement can be realized and measured. Robert Stake identifies two choices for program evaluators either “to try to compare it to another program, a model program; or to try to compare it to a set criteria that represents a model program, with standards marking different levels for each of the criteria” (Stake, 2004, p. 8). In June 2000, the NDE and Buros Institute for Assessment Consultation and Outreach (BIACO) agreed upon a standards based evaluation using a “technical quality rubric” (Impara, Buckendahl, & Plake, 2001). A team of assessment specialists was recruited to “apply the final technical quality rubric (NDE, 2003) to district assessment portfolios” (Buckendahl, Impara, Plake, Ferdous, and Haack, 2003) submitted by districts in 2003 and 2004. Starting in 2006 the portfolio review process changed. The District Assessment Portfolio Rubric added a sufficiency requirement to the alignment criteria and adjusted the rating scale. The last review process to look at the reliability of the rubric was September 2003. “The results of this review were that 25 of 40 (63%) resulted in the same rating (exact agreement) between the original rater and the second rater” (Impara, et. al., 2003). The revised assessment rubric needed to be evaluated to ensure that the collection of information was applied in a consistence manner to the district assessment portfolios.

### **PURPOSE OF THE STUDY**

“Nebraska Educators Review the Local Math Assessment Process: Reliability of Peer Review of Assessment Portfolios” is a study undertaken by the Nebraska Department of Education. This study evaluates the Peer Review Process by establishing the reliability of reviewer decisions in collecting information from district assessment portfolios. The study will attempt to answer the following question:

- Is there agreement between the reviewers on rating indicators of assessment portfolios for each of the Six Quality Assessment Criteria?

### **LITERATURE REVIEW**

The STARS system is dependent upon the use of a portfolio as a demonstration that the assessment process meets assessment quality. Beginning in 2001 and continuing through 2007 districts have developed assessment portfolios to provide evidence of the quality assessment process at the local school level for the benchmark grades of four, eight, and eleven. Schools submitted the portfolios in 2001 (reading), 2002 (math), 2003 (reading), and 2004 (math) to the Department of Education for a group of assessment specialists to evaluate. Reports submitted by Buros Institute for Assessment Consultation and Outreach documented the reliability of reviewers in classifying district portfolios in final reports. “Reliability varied in consistency across the six quality criteria” (Buckendahl, Impara, & Plake, 2002; Impara, et. al., 2002).

In 2006-2007, the Department began a new process called the “Nebraska-led Peer Review of STARS” where two peer reviewers visited the district and collected information for evaluating the district’s portfolio (Guidelines for Assessment Quality for STARS, 2007-2008, p. 3). This decision was consistent with a recommendation made in the Year One Report on Charting the Stars (Gallagher, 2002) to “offer more, concrete, feedback to districts on their assessment systems” (p. 31). As noted later, the inclusion of educators in the Portfolio Review training process accomplished this recommendation. Just like in the

previous process, there was a need for the NDE to verify that reviewers were consistent in collecting data across districts since the information was used in the portfolio rating process. The NDE needed to complete a study to assure the consistency of data collection across raters.

In the first year of the change (2006) from the previous process of having districts submit assessment portfolios for review to having peer reviewers collect the information in on-site interviews, districts felt the change was positive. An evaluation of the Peer Review Process found that “many educators felt the Peer Review Process allowed the district to honor their teachers and the amount of work they have done to make the process strong and viable” (Isernhagen & Mills, 2007, p. 73). Additional findings in the study stated that “teachers and leaders were very positive about the reviewers themselves and their ability to lead the review process at the building and district level” (Isernhagen & Mills, 2007, p. 69). Reviewers indicated benefits in the new process with “many reviewers commenting about how much *they* personally learned” (Isernhagen & Mills, 2007, p. 69). These comments from the STARS evaluation indicated a positive perception of the process from those involved. But, the Department still needed information about the reliability of the process.

Further changes to the process added extra challenges to the training of peer reviewers. Adjustments made to the *District Assessment Portfolio Rubric* for meeting the six quality criteria (QC) became effective in 2006-2007. The *District Assessment Portfolio Rubric* still continues to measure the technical quality of district assessment portfolios according to the six quality assessment criteria (Plake, Impara, & Buckendahl, 2004).

- QC 1 - Alignment to standards and range of assessment items (sufficiency)
- QC 2 - Opportunity to learn
- QC 3 - Unbiased assessments
- QC 4 - Appropriate assessment level
- QC 5 - Reliability of scoring
- QC 6 - Setting of cut scores

As noted in the 2006 evaluation of STARS, “the Six Quality Assessment Criteria are reflective of good practices in educational testing and assessment” (Lane, 2006, p. 7). She also noted that the following changes for the rubric became effective for 2006-2007 school year:

- For Quality Criteria 1, sufficiency requirement became more explicit (Minimum of 12 items or equivalent on selected reading and math standard).
- For Quality Criteria 1 through 6, consistency needs to be clear and complete between the criteria.

In reviewing district assessments to meet the sufficiency requirement, most districts made changes to the assessment thus precipitating a review of all six QC’s on the revised assessments. The description for each QC had to be consistent with each of the other criteria (Guidelines for Assessment Quality for STARS, 2007-2008, p. 4). Year Two report on STARS recommended that the process should be simplified by allowing “for reporting on fewer standards” (Gallagher, 2003). This suggestion resulted in the identification of STAR (Standards That Are Reported) standards. This change impacted district’s assessment portfolios and became effective in 2006-2007 with many districts deciding to report on a subset of the State Standards instead of the full slate (Nebraska Department of Education,

August 2007, p. 2). Districts revised their portfolios based on this decision. Changes were made to the training process as a result of these new requirements.

In an *Evaluation Report on the NDE's District Assessment Portfolio Training Process*, Dr. Forte Fast recommended from her observations that the training “continues to involve local educators” to gain valuable knowledge about the process to share within their districts (Forte Fast, 2004). This recommendation was first implemented into the training during the summer and continued through trainings into 2007. A *Qualitative Case Study of the STARS Portfolio Review Process* looked at the communication between educators and assessment specialists during the STARS Portfolio Review Training sessions in the summer of 2005 to discover differences between the two groups (McEntarffer & Norman, 2005, p. 3). In this study the researchers found that the educators who participated “learned a great deal about the meaning behind the language in the six quality criteria rubric” (McEntarffer & Norman, 2005, p. 11). The findings included statements about expanding the training to other educators as “a valuable experience that needed to be shared with other educators” (McEntarffer & Norman, 2005, p. 15). During the training process the researchers observed that “teachers and assessment specialists both communicated their ideas effectively but differently” (McEntarffer & Norman, 2005, p. 18). In the end, the researchers identified the discussions during the training “as the most important way to create understanding about the realities and promise of the STARS process” (McEntarffer & Norman, 2005, p. 21).

## **METHODS AND PROCEDURES**

Teachers, administrators, staff developers, and college staff were selected from educators who completed a peer reviewer application sent to the Statewide Assessment Office in the Nebraska Department of Education. Applications to become a peer reviewer were sent to superintendents and curriculum/assessment directors of each district, directors and staff developers of each education service unit, and college representatives. From the applications, individuals were selected and invited to participate in the peer review training. The reviewers contract with NDE for three days of training and four days of on-site reviews with three school districts. As part of their training each reviewer completed the District Assessment Portfolio worksheet on one of three sample portfolios. One hundred and one (101) educators who participated in the peer review training and conducted district portfolio reviews participated in the study and completed the worksheet in two hours.

In this study, data was collected using the “assessment rubric” of the six quality criteria indicators to measure reviewer consistency in evaluating district assessment portfolios. On the third day of training the reviewers were given a sample portfolio to evaluate using the District Assessment Portfolio Rubric. Each reviewer completed a review of the six quality criteria and recorded the results on a “District Assessment Portfolio Worksheet” (Appendix G). From the data collected using three different portfolios in January, February, and April, an analysis was made to determine reliability of evaluations made by different reviewers on the same portfolio. The reviewer’s performance in assigning a rating of “Missing”, “Incomplete” or “Complete” to Quality Criterion indicators was used to obtain the percentage of agreement across the reviewers. Assessment specialists along with the reviewers used the indicators (five to nine per criteria) to rate whether the portfolio met the criterion. A count of ratings for each indicator was used to calculate percentage of the rating

agreement between the reviewers. The on-site peer review process's success was dependent on the reviewers correctly evaluating each indicator in the district's portfolio.

## RESULTS

Analysis of the peer reviewers' evaluations of the sample portfolios provided a method to determine the percent of agreement between reviewers for the six Quality Criteria indicators. The results from the Quality Criteria indicators present a consistent picture of what the reviewer observed in the sample portfolios and how they evaluated the indicators. The results will be presented for each of Quality Criteria.

**Quality Criterion 1 – The assessment match the standards.** For this criterion the reviewers showed an average agreement on indicators from a high of 1.00 to a low of 0.86 with most averages being 0.90 or above. See Table 14 for reviewer decisions by indicator for each sample portfolio. Overall agreement for each indicator is calculated with a weighted average of the rater agreement for each of the three sample portfolios. For Criterion One, all these averages were 0.87 or above.

QC 1.1: Qualifications of the independent reviewers are clear and complete.	QC 1.5: Sufficiency process is clear and complete.
QC 1.2: Evidence of an independent review for match to standards is clear and complete (reviewers did not write the assessments.)	QC 1.6: Sufficiency results are clear and complete (sufficiency required for both number of items/ performances and levels of difficulty. Minimum 12 items or equivalent on math standards 4.2.1, 8.2.2, and 12.2.1)
QC 1.3: The process for matching assessments to standards is clear and complete.	QC 1.7: Consistency between Criterion 1 and other criteria is clear.
QC 1.4: Results of the matching process are clear and complete.	

<b>Table 14. Quality Criterion 1 – The assessments match the standards.</b>								
District	Rating	QC 1.1	QC 1.2	QC 1.3	QC 1.4	QC 1.5	QC 1.6	QC 1.7
A - Jan	Missing	0	0	0	0	0	0	0
	Incomplete	2	5	5	4	7	5	7
	Complete	48	45	45	46	43	44	38
Rater Agreement		0.96	0.90	0.90	0.92	0.86	0.90	0.84
B - Feb	Missing	0	0	0	0	0	0	0
	Incomplete	0	0	0	1	1	0	3
	Complete	30	30	30	29	29	30	22
Rater Agreement		1.00	1.00	1.00	0.97	0.97	1.00	0.88
C - Apr	Missing	1	0	0	0	0	0	0
	Incomplete	1	1	0	2	0	2	3
	Complete	29	30	31	28	31	29	25
Rater Agreement		0.94	0.97	1.00	0.93	1.00	0.94	0.89
Overall Agreement		0.96	0.95	0.95	0.94	0.93	0.94	0.87

**Quality Criterion 2 – The students have the opportunity to learn.** For this criterion, the reviewers showed an average agreement on indicators from a high of 1.00 to a low of 0.46 with many averages being 0.90 or above. See Table 15 for reviewer decisions by indicator for each sample portfolio. Overall agreement for each indicator is calculated with a weighted average of the rater agreement for each of the three sample portfolios. For Criterion Two, all these averages were 0.71 or above.

<b>QC 2.1: Qualifications of the opportunity to learn reviewers are clear and complete.</b>	<b>QC 2.6: Dates are provided when standards are assessed and are clear and complete</b>
<b>QC 2.2: The process for alignment of standards with local curriculum is clear and complete.</b>	<b>QC 2.7: 80% of instruction should take place prior to assessment.</b>
<b>QC 2.3: The process for timing of assessment/instruction is clear and complete.</b>	<b>QC 2.8: Consistency between Criterion 2 and other criteria is clear and complete.</b>
<b>QC 2.4: The results of the process for alignment of standards with local curriculum are clear and complete.</b>	<b>QC 2.9: Opportunity to learn information provided for all standards.</b>

<b>Table 15. Quality Criterion 2 – Students must have the opportunity to learn.</b>										
District	Rating	QC 2.1	QC 2.2	QC 2.3	QC 2.4	QC 2.5	QC 2.6	QC 2.7	QC 2.8	QC 2.9
A - Jan	Missing	0	0	11	1	35	28	7	1	8
	Incomplete	3	4	14	11	15	15	16	13	19
	Complete	47	45	24	37	0	7	27	34	23
Rater Agreement		0.94	0.92	0.49	0.76	0.70	0.56	0.54	0.71	0.46
B - Feb	Missing	0	0	0	0	0	0	1	0	0
	Incomplete	0	0	4	2	8	1	3	0	4
	Complete	30	30	26	28	22	29	25	24	26
Rater Agreement		1.00	1.00	0.87	0.93	0.73	0.97	0.86	1.00	0.87
C - Apr	Missing	1	0	1	0	0	0	0	0	0
	Incomplete	0	1	1	1	8	9	1	5	1
	Complete	29	28	27	29	22	21	30	18	28
Rater Agreement		0.97	0.97	0.93	0.97	0.73	0.70	0.97	0.78	0.97
Overall Agreement		0.96	0.95	0.71	0.86	0.72	0.71	0.75	0.80	0.71

**Quality Criterion 3 – The assessments are free of bias and sensitive situations.** For this criterion, the reviewers showed an average agreement on indicators from a high of 1.00 to a low of 0.54 with many averages 0.90 or above. See Table 16 for reviewer decisions by indicator for each sample portfolio. Overall agreement for each indicator is calculated with a weighted average of the rater agreement for each of the three sample portfolios. For Criterion Three, all averages were 0.75 or above.

<b>QC 3.1: Qualifications of the bias reviewers are clear and complete.</b>	<b>QC 3.4: Results of a bias review are clear and complete.</b>
<b>QC 3.2: The description of the bias orientation/training process is clear and complete.</b>	<b>QC 3.5: Bias information provided for all standards (used for reporting.)</b>
<b>QC 3.3: The process for bias review of assessment items is clear and complete.</b>	<b>QC 3.6: Consistency between Criterion 3 and other criteria is clear and complete</b>

<b>Table 16. Quality Criterion 3 – The assessments are free of bias and sensitive situations.</b>							
District	Rating	QC 3.1	QC 3.2	QC 3.3	QC 3.4	QC 3.5	QC 3.6
A - Jan	Missing	2	1	4	3	6	0
	Incomplete	8	17	15	14	27	8
	Complete	40	32	31	33	17	39
Rater Agreement		0.80	0.64	0.62	0.66	0.54	0.83
B - Feb	Missing	0	0	0	0	0	0
	Incomplete	1	0	3	1	1	1
	Complete	29	29	27	29	28	27
Rater Agreement		0.97	1.00	0.90	0.97	0.97	0.96
C - Apr	Missing	0	0	1	0	0	0
	Incomplete	0	2	4	4	2	3
	Complete	31	29	25	25	27	25
Rater Agreement		1.00	0.94	0.83	0.86	0.93	0.89
Weighted average		0.90	0.82	0.75	0.80	0.76	0.88

**Quality Criterion 4 – The assessments are at the appropriate level.** For this criterion, the reviewers showed an average agreement on indicators from a high of 1.00 to a low of 0.79 with many averages 0.90 or above. See Table 17 for reviewer decisions by indicator for each sample portfolio. Overall agreement for each indicator is calculated with a weighted average of the rater agreement for each of the three sample portfolios. For Criterion Four, all averages were 0.86 or above.

QC 4.1: Qualifications of the reviewers for appropriate level are clear and complete.	QC 4.4: Appropriate level information is provided for all standards (used for reporting.)
QC 4.2: Process for appropriate level review is clear and complete.	QC 4.5: Consistency between Criterion 4 and other criteria is clear and complete.
QC 4.3: Results of the appropriate level review are clear and complete.	

<b>Table 17. Quality Criterion 4 – The assessments are at the appropriate level.</b>						
District	Rating	QC 4.1	QC 4.2	QC4.3	QC 4.4	QC 4.5
A - Jan	Missing	0	1	2	1	0
	Incomplete	2	8	4	1	4
	Complete	48	41	44	48	44
Rater Agreement		0.96	0.82	0.88	0.96	0.92
B - Feb	Missing	0	0	0	0	0
	Incomplete	0	0	2	2	2
	Complete	30	30	28	28	28
Rater Agreement		1.00	1.00	0.93	0.93	0.93
C - Apr	Missing	0	1	1	0	0
	Incomplete	0	5	5	2	2
	Complete	31	25	23	29	26
Rater Agreement		1.00	0.81	0.79	0.94	0.93
Overall Agreement		0.98	0.86	0.87	0.95	0.92

**Quality Criterion 5 – There is consistency in scoring.** For this criterion, the reviewers showed an average agreement on indicators from a high of 1.00 to a low of 0.48 with many averages being 0.90 or above. See Table 18 for reviewer decisions by indicator for each sample portfolio. Overall agreement for each indicator is calculated with a weighted average of the rater agreement for each of the three sample portfolios. For Criterion Five, all averages were 0.77 or above.

<b>QC 5.1: Qualifications of the reliability process participants are clear and complete.</b>	<b>QC 5.4: Procedure for improving reliability is clear and complete.</b>
<b>QC 5.2: Appropriate process for reliability is clear and complete.</b>	<b>QC 5.5: Reliability is reported for all standards (used for reporting).</b>
<b>QC 5.3: Reliability value provided and calculations are at or above the minimum acceptable level. (Minimum level of acceptable reliability is 0.70, mean or median, averaged across all standards.)</b>	<b>QC 5.6: Consistency between Criterion 5 and other criteria is clear and complete.</b>

<b>Table 18. Quality Criterion 5 – There is consistency in scoring.</b>							
District	Rating	QC 5.1	QC 5.2	QC 5.3	QC 5.4	QC 5.5	QC 5.6
A - Jan	Missing	2	1	0	1	0	0
	Incomplete	21	11	1	3	3	4
	Complete	26	38	48	46	47	44
Rater Agreement		0.53	0.76	0.98	0.92	0.94	0.92
B - Feb	Missing	0	0	0	1	0	0
	Incomplete	0	4	26	14	3	3
	Complete	28	25	4	14	26	22
Rater Agreement		1.00	0.86	0.87	0.48	0.90	0.88
C - Apr	Missing	0	1	0	0	0	0
	Incomplete	2	4	1	4	1	2
	Complete	28	25	27	25	28	24
Rater Agreement		0.93	0.83	0.96	0.86	0.97	0.92
Weighted average		0.77	0.81	0.94	0.79	0.94	0.91

**Quality Criterion 6 – Mastery levels are appropriate.** For this criterion, the reviewers showed an average agreement on indicators from a high of .97 to a low of 0.42 with many averages 0.90 or above. See Table 19 for reviewer decisions by indicator for each sample portfolio. Overall agreement for each indicator is calculated with a weighted average of the rater agreement for each of the three sample portfolios. For Criterion Six, all averages were 0.66 or above.

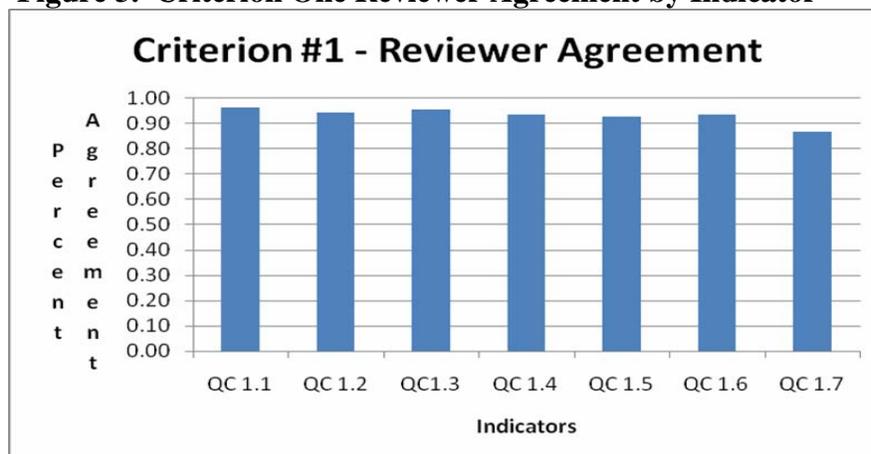
<b>QC 6.1: Qualifications for mastery level participants are clear or complete.</b>	<b>QC 6.4: Mastery level information is provided for all standards (used for reporting.)</b>
<b>QC 6.2: Evidence of mastery level process is clear or complete.</b>	<b>QC 6.5: Consistency between criterion #6 and other criteria is clear and complete.</b>
<b>QC 6.3: Results of the mastery level process are clear and complete.</b>	

<b>Table 19. Quality Criterion 6 – Mastery levels are appropriate.</b>						
District	Rating	QC 6.1	QC 6.2	QC6.3	QC 6.4	QC 6.5
A - Jan	Missing	0	2	1	8	2
	Incomplete	7	16	15	21	11
	Complete	42	32	34	21	34
Rater Agreement		0.86	0.64	0.68	0.42	0.72
B - Feb	Missing	0	0	1	0	0
	Incomplete	1	2	12	2	8
	Complete	28	27	17	28	18
Rater Agreement		0.97	0.93	0.57	0.93	0.69
C - Apr	Missing	0	0	0	1	0
	Incomplete	2	6	8	5	7
	Complete	29	23	22	22	20
Rater Agreement		0.94	0.79	0.73	0.79	0.74
Weighted average		0.91	0.76	0.66	0.66	0.72

### SUMMARY

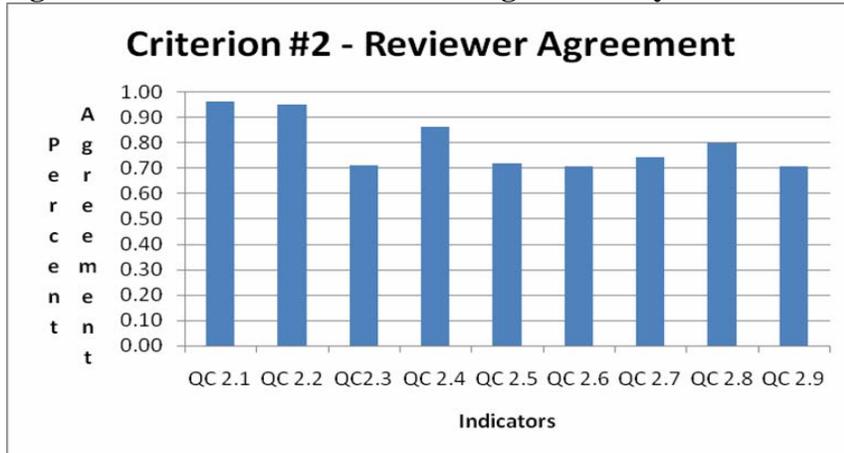
This study was designed to determine whether the new peer review process provided a reliable way to collect information about a district’s peer review portfolio assessment process. Part of the process was the evaluation of the district’s portfolio in meeting the indicators for each of the six criteria. Peer reviewers met with district representatives to discuss the district assessment process as they reviewed the district portfolio and collected information about the indicators for each quality criterion. In examining the results of the study, it was evident that the reviewers showed some variation in their evaluations of some criteria, but the overall agreement between raters was 0.66 or above. Each Quality Criterion presents a different picture in reviewer agreement on the relevant indicators. The results pictured in the Figures below will be discussed for each Quality Criterion.

**Figure 3. Criterion One Reviewer Agreement by Indicator**



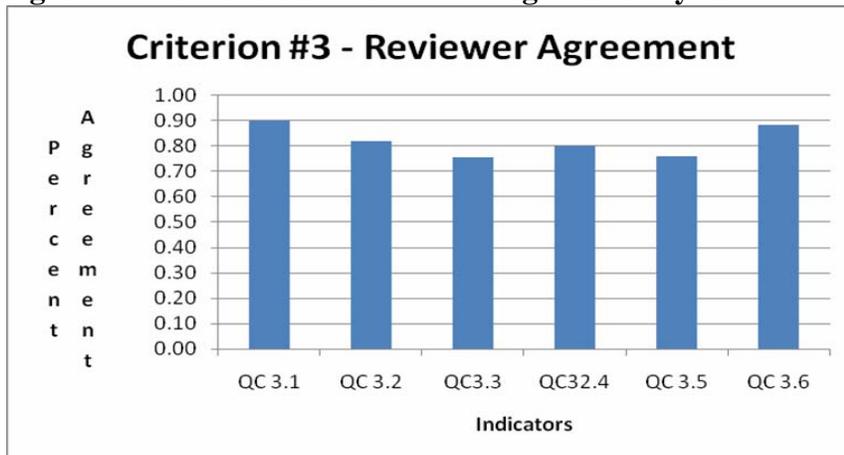
For Criterion One, the reviewers attained a high level of agreement with the highest percentage being 0.96 for indicator QC 1.1, “identifying whether the reviewers were qualified and independent of the assessment writers.” The indicator with the lowest agreement was QC 1.7 “establishing that this criterion was consistent with the other criteria.” Six out of the seven indicators showed an agreement rating greater than 0.90.

**Figure 4. Criterion Two Reviewer Agreement by Indicator**



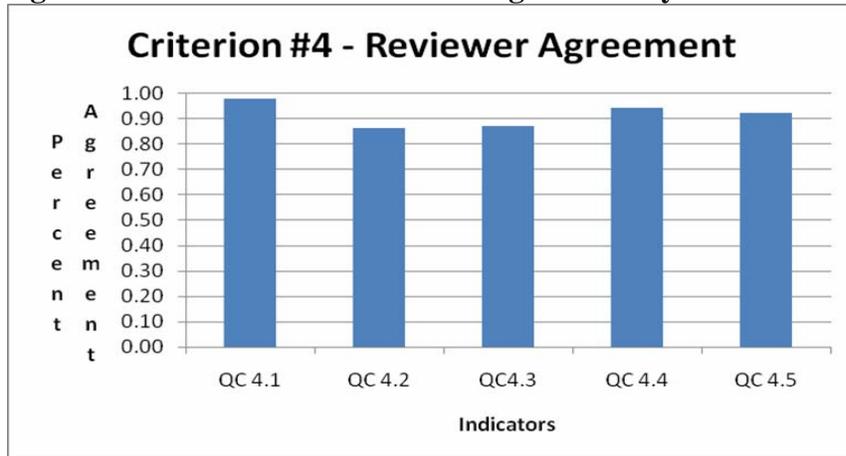
For Criterion Two, the reviewers showed more variability in agreement with the highest percentage being 0.96 for indicator QC 2.1, “identifying whether the reviewers were qualified.” The indicators with the lowest agreement were QC 2.3, QC 2.6, and QC 2.9, the process for timing, clarity of assessment dates, and information provided for all standards. Only two out of the nine indicators were above 0.90, but all ratings were greater than 0.70.

**Figure 5. Criterion Three Reviewer Agreement by Indicator**



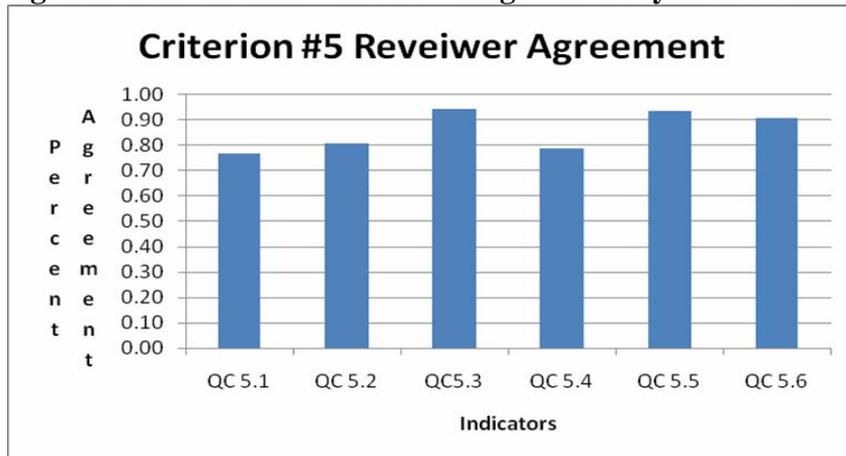
For Criterion Three, the reviewers again showed more variability across the indicators with the highest agreement being 0.90 for indicator QC 3.1, “identifying whether the bias reviewers were qualified.” The indicators with the lowest agreement were QC 3.3, “examining the process for bias review.” The highest indicator was 0.90, with all other ratings equal to or greater than 0.75.

**Figure 6. Criterion Four Reviewer Agreement by Indicator**



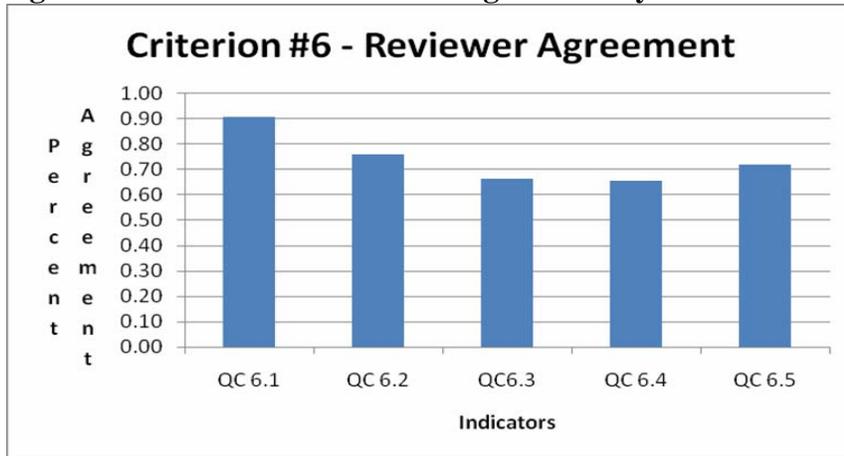
For Criterion Four, the reviewers showed less variability across the indicators with the highest agreement being 0.98 for indicator QC 4.1, “identifying whether the reviewers for appropriateness were qualified.” The indicator with the lowest agreement was QC 4.2, “examining the process for appropriate level review.” Three of the five indicators were greater than 0.90, and the remaining two quality criterion were greater than 0.85.

**Figure 7. Criterion Five Reviewer Agreement by Indicator**



For Criterion Five, the reviewers showed less variability across the indicators with the highest agreement being 0.94 for indicator QC 5.3, “reliability value provided” and QC 5.5, “reliability reported for all standards.” The indicator with the lowest agreement was QC 5.1, “identifying the qualifications of reliability participants.” All six indicators rated greater than 0.75 with three ratings greater than 0.90.

**Figure 8. Criterion Six Reviewer Agreement by Indicator**



For Criterion Six, the reviewers showed less variability across the indicators with the highest agreement being 0.94 for indicator QC 6.3, “results of mastery level process” and QC 6.5, “consistency between criterion six and other criteria.” The indicator with the lowest agreement was QC 6.1, “identifying the qualifications of mastery level participants.” All six indicators rated greater than 0.75 with two indicators rating greater than 0.90.

### **Discussion and Recommendations**

Tables 14 through 19 along with Figures 3 through 6 show the high level of agreement between reviewers on evaluating the indicators of the three sample portfolios. This level of agreement provides support that the initial step in the 2007-2008 evaluation process, namely the collection of data by the peer reviewers, does provide an accurate picture of the district assessment process. Additional steps in the process add more confidence in the ratings assigned for each of the Quality Criteria leading to an overall district rating to be published in the 2007-2008 State of the Schools Report. To reduce variability in the evaluations, the peer review process was supplemented with the following steps:

1. The peer review training was evaluated in 2006-2007 and adjustments to the training process were made based on this evaluation.
2. District portfolios were reviewed by two peer reviewers with the findings being synthesized between the two.
3. The reviewers’ findings were discussed with the district staff during the review. Clarification to the portfolio along with additional information was accepted during the on-site visit by the review team.
4. The Quality Criterion Rating (Not Met, Needs Improvement, Met w/comment, and Met) was assigned in conjunction with an assessment specialist. Discussion of the findings with the Assessment Specialists provided another consistency factor in the evaluation process.
5. Each district had the right to appeal the ratings by collecting missing or incomplete information and submitting it for further review at a later time. Assessment specialists reviewed the information presented during the appeal process.

Analysis of the data collected from “Evaluation Survey of Peer Review Training” (Appendix H) should be pursued with a focus on the different components of the peer review training and the effectiveness of the training in establishing accurate guidelines for the peer

reviewers. Other uses of this process may be appropriate for educational evaluations by the state in the future. Additional study in how the process was perceived by reviewers and district staff as well as assessment specialists would help to identify how the process benefits the educators in Nebraska and eventually the students. Using evaluation rubrics similar to the District Assessment Portfolio Worksheet when evaluating local school improvement processes should be encouraged to continue the benefits of the current state level assessment process.

## REFERENCES

- Buckendahl, C. W., Impara, J. C., & Plake, B. S. (2002). Final report evaluating Nebraska's district assessment portfolios and recommending model assessments for mathematics: 2001-2002. Retrieved on September 11, 2008 from <http://www.nde.state.ne.us/assessment/techreports/finalreportmatheval0102.pdf>
- Buckendahl, C. W., Impara, J. C., Plake, B. S., Ferdous, A., & Haack, M. K. (2003). Final report evaluating Nebraska's district assessment portfolios and recommending model assessments for reading, speaking, & listening: 2002-03 Retrieved on September 11, 2008 from <http://www.nde.state.ne.us/assessment/techreports/finalreport-rtsl03.pdf>
- Forte Fast, E. (2004). *Evaluation report on the Nebraska State Department of Education's district assessment portfolio training process*. Retrieved on September 11, 2008 from <http://www.nde.state.ne.us/assessment/techreports/documents/EFFreportonDAP10-12-04.pdf>
- Gallagher, C. W. (2002). *Charting STARS: The state of assessment in the State of Nebraska – research study and comprehensive evaluation of Nebraska's School-based, Teacher-led Assessment and Reporting System (STARS)*. Retrieved on September 11, 2008 from [http://www.nde.state.ne.us/assessment/documents/YEARONEREPORT.ALL\\_000.pdf](http://www.nde.state.ne.us/assessment/documents/YEARONEREPORT.ALL_000.pdf)
- Gallagher, C. W. (2003). *Charting STARS – Sustainability as challenge and opportunity – research study and comprehensive evaluation of Nebraska's School-based, Teacher-led Assessment and Reporting System (STARS)*. Retrieved on September 11, 2008 from <http://www.nde.state.ne.us/assessment/documents/yeartworeport--onefile.pdf>
- Impara, J. C., Buckendahl, C. W., & Plake, B.S. (2001). Final report evaluating Nebraska's district assessment portfolios and recommending model assessments for reading: 2000-2001. Retrieved on September 11, 2008 from <http://www.nde.state.ne.us/assessment/techreports/finalreportreadingeval0001.pdf>
- Isernhagen, J. C., & Mills, S. J. (2007). *Charting STARS: Pathways to success – comprehensive evaluation of Nebraska's School-based Teacher-led Assessment and Reporting System (STARS)*. Retrieved on September 11, 2008 from <http://www.nde.state.ne.us/assessment/documents/STARSTFinalReportFinalRevision092007.pdf>
- Lane, S. (2006). *Evaluation of the Six Quality Assessment Criteria used in the Nebraska School-based Teacher-led Assessment and Reporting System (STARS)*. Retrieved on September 11, 2008 from <http://www.nde.state.ne.us/assessment/documents/NebraskaReportLane06.pdf>
- McEntarffer, R., & Norman, R. (2005). *Teachers and assessment specialists in conversation: a qualitative case study of the STARS portfolio review process*. Paper presented by Rebecca Norman at University of Nebraska – Student Research Conference,

Nov. 17, 2007: Lincoln, NE. Available from Rebecca Norman, UNL, Buros Center for Testing.

Nebraska Department of Education (2007a). *Guidelines & requirements for documenting assessment quality for STARS*. Retrieved on September 11, 2008 from [http://www.nde.state.ne.us/assessment/documents/GuidelinestoDocumentAQMASTER\\_001.pd](http://www.nde.state.ne.us/assessment/documents/GuidelinestoDocumentAQMASTER_001.pd)

Nebraska Department of Education (2007b). *STARS Update #24*. Retrieved on September 11, 2008 from [http://www.nde.state.ne.us/assessment/documents/Update24\\_001.pdf](http://www.nde.state.ne.us/assessment/documents/Update24_001.pdf)

Stake, R. E. (2004). *Standards-based & responsive evaluation*. Thousand Oaks, CA: Sage.

Plake, B. S., Impara, J. C., & Buckendahl, C. W. (2004). Technical quality criteria for evaluating district assessment portfolios used in the Nebraska STARS. *Educational Measurement: Issues and Practice*, 23(2), 9-11.

**CHARTING STARS:  
PORTRAITS OF EXCELLENCE**

**Section 4: Appendices**



**Appendix A**  
**IRB Approval Letter**



January 16, 2008

Dr. Jody Isernhagen  
Dr. Shirley Mills  
132 TEAC  
(0360)

IRB# 2007-11-8184 EX

TITLE OF PROJECT: **Comprehensive Evaluation of School-based Teacher-led Assessment and Reporting System**

Dear Dr. Isernhagen:

The Institutional Review Board for the Protection of Human Subjects has completed its review of the Request for Change in Protocol submitted to the IRB.

1. It has been approved to accept the approval letters submitted with the 1-11-08 change request. Approval letters were accepted from: North Platte Public Schools, Fremont Public Schools, Columbus Public Schools, Garden Couth Public Schools, Crete Public Schools, Aurora Public Schools, and Omaha Nation.
2. It has been approved to accept the approval letter from Grand Island Public Schools submitted with the 1-16-08 change request.

We wish to remind you that the principal investigator is responsible for reporting to this Board any of the following events within 48 hours of the event:

- Any serious event (including on-site and off-site adverse events, injuries, side effects, deaths, or other problems) which in the opinion of the local investigator was unanticipated, involved risk to subjects or others, and was possibly related to the research procedures;
- Any serious accidental or unintentional change to the IRB-approved protocol that involves risk or has the potential to recur;
- Any publication in the literature, safety monitoring report, interim result or other finding that indicates an unexpected change to the risk/benefit ratio of the research;
- Any breach in confidentiality or compromise in data privacy related to the subject or others; or
- Any complaint of a subject that indicates an unanticipated risk or that cannot be resolved by the research staff.

This letter constitutes official notification of the approval of the protocol change. You are therefore authorized to implement this change accordingly.

If you have any questions, please contact Shirley Horstman, IRB Administrator, at 472-9417 or email [shorstman1@unl.edu](mailto:shorstman1@unl.edu).

Sincerely,

  
Dan R. Hoyt, Chair  
for the IRB



**Appendix B**  
**Researchers for the Project**



## **Researchers for the Primary Seventh Year Project 2007-2008**

### ***Principal Investigator***

**Dr. Jody C. Isernhagen** is an Associate Professor in Educational Administration at the University of Nebraska-Lincoln. She received her doctoral degree from Virginia Tech and has been a teacher, assistant principal, principal, supervisor of elementary education, and superintendent in pre-K through 12 schools. Dr. Isernhagen serves as the primary investigator for the STARS Process and is the primary instructor for the School Improvement Specialist Program, a joint program between the North Central Association on Accreditation and School Improvement (NCA CASI). She serves as the State Accreditation and North Central Accreditation External Leader for four school districts in Nebraska. Dr. Isernhagen was awarded the College of Education and Human Sciences Distinguished Teaching Award.

### ***Secondary Investigators***

**Jackie Florendo, M.Ed.**, received her Master's degree in Curriculum and Instruction from Doane College and is currently a doctoral student in Educational Administration at the University of Nebraska-Lincoln.

**Dr. Shirley J. Mills** is an Assistant Professor at the University of Texas-Pan American and served as a secondary investigator for the STARS Comprehensive Evaluation. She taught in Nebraska for 38 years prior to receiving her Doctorate in Leadership and Higher Education from the University of Nebraska-Lincoln in 2005.

**John Moon, M.A.**, is the Assessment Coordinator for the Nebraska Department of Education. He received his Masters in Educational Psychology from the University of Nebraska-Lincoln and is currently a doctoral student majoring in Quantitative and Qualitative Psychometric Measurement at the University of Nebraska-Lincoln.

**Dr. Jerald Riibe** is the Assistant Superintendent for Curriculum and Instruction for Ralston Public Schools. He received his Doctorate in Educational Administration from the University of Nebraska in 2008.

**Casey Tallent, M.A.**, received her Master's degree in Counseling and Guidance from the University of Missouri-Kansas City and is a doctoral student in Educational Psychology at the University of Nebraska-Lincoln.

**Nino Zhvania, M.A.**, a Muskie Scholar, completed her Masters degree in Educational Administration at the University of Nebraska-Lincoln. She is currently working for the Ministry of Education in the Republic of Georgia and has been accepted to the Doctoral Program in Educational Administration at the University of Nebraska-Lincoln.

### ***Project Administrative Coordinator***

**Susan Wilson** is on the staff at the University of Nebraska-Lincoln and holds an Associate of Science degree in Business Administration from the College of St. Mary, Omaha.



**Appendix C**  
**2007-2008 Study I: Nebraska-led Math Portfolio**  
**Peer Review Survey**



**Nebraska-Led Portfolio Peer Review Survey  
STARS Comprehensive Evaluation Project**

**INSTRUCTIONS:** Please provide the following demographic information by responding to the questions or marking the appropriate category for each area.

**1. Mark all the position(s) below that you hold in your school district:**

\_\_\_ Superintendent \_\_\_ Curriculum Coordinator \_\_\_ Assessment Coordinator Other: \_\_\_\_\_

Principal: \_\_\_ ELEM \_\_\_ MS \_\_\_ HS

Teacher: \_\_\_ ELEM \_\_\_ MS: Subject \_\_\_\_\_ HS: Subject \_\_\_\_\_

**2. My Gender:** \_\_\_ Male \_\_\_ Female

**3. Years of experience with assessment portfolio:** \_\_\_\_\_

**4. Total Years of Experience in Education:** \_\_\_\_\_

	None of the time	Very little of time	Some of the time	Most of the time	All of the time
<b>Directions:</b> Please circle the number that best describes your response to each statement.					
<b>ALIGNMENT</b>					
1. Our district assessment items/tasks reflect a match to the appropriate standards.	1	2	3	4	5
2. Our district assessment items/tasks reflect the content and skills found within the standards.	1	2	3	4	5
3. Our district had assessment items reviewed by district personnel.	1	2	3	4	5
4. Our district had assessment items reviewed by external personnel.	1	2	3	4	5
5. Our district has a list of specifications mapping the assessment items to the standards in order to show which items assess which standards.	1	2	3	4	5
6. There is a documentation process for alignment of assessments to standards.	1	2	3	4	5
7. Our district involved staff in the alignment of the assessments to standards.	1	2	3	4	5
8. Our district supports teachers working collaboratively to ensure assessments measure the standards.	1	2	3	4	5
<b>SUFFICIENCY</b>					
9. Our district measures all academic content standards in the assessment items/tasks.	1	2	3	4	5
10. Our assessment items/tasks are distributed across all performance levels.	1	2	3	4	5
11. Our assessment items/tasks use a variety of appropriate formats.	1	2	3	4	5
12. Our assessment items/tasks include higher order thinking skills.	1	2	3	4	5
13. Our district reviewed assessment items/tasks for sufficiency results.	1	2	3	4	5
<b>CLARITY</b>					
14. Our assessment directions for students are clear.	1	2	3	4	5
15. Our assessment directions for teachers are clear.	1	2	3	4	5
16. Our assessment directions for students are standardized across the district.	1	2	3	4	5
17. Our assessment directions for teachers are standardized across the district.	1	2	3	4	5
18. Our district sends individual reports to parents each school year.	1	2	3	4	5
19. Our district provides parents with reports that give an appropriate explanation of assessments results.	1	2	3	4	5
20. Our district provides individual student reports in appropriate language(s).	1	2	3	4	5
21. All district/school reports are appropriately disaggregated.	1	2	3	4	5
<b>APPROPRIATENESS</b>					
22. Our assessments are appropriate for the assessed grade level.	1	2	3	4	5
23. Our assessments demonstrate an increase of expectation from one grade level to the next.	1	2	3	4	5
24. Our assessments were screened for fairness, bias, and sensitivity.	1	2	3	4	5
25. Our assessments indicate our expectations for the students.	1	2	3	4	5
26. Our assessment plan provides for appropriate accommodations where necessary.	1	2	3	4	5
27. Our assessments have been administered with appropriate accommodations.	1	2	3	4	5
28. Our assessments were reviewed by internal or external groups.	1	2	3	4	5

**Nebraska-Led Portfolio Peer Review Survey  
STARS Comprehensive Evaluation Project**

	None of the time	Very little of the time	Some of the time	Most of the time	All of the time
<b>Directions:</b> Please circle the number that best describes your response to each statement.					
<b>SCORING PROCEDURES</b>					
29. Our performance level descriptors are clear and specific for each assessment.	1	2	3	4	5
30. Our performance level descriptors clearly differentiate for each proficiency level.	1	2	3	4	5
31. Our district consistently applies performance level descriptors to the cut scores for each assessment or standard.	1	2	3	4	5
32. Our performance level descriptors indicate increased expectations from one grade level to the next.	1	2	3	4	5
33. Our assessments have established scoring guidelines and directions.	1	2	3	4	5
34. Our subjectively scored assessments have clearly defined rubrics.	1	2	3	4	5
35. Our subjectively scored assessments have inter-rater reliability and decision consistency methods that are within acceptable ranges.	1	2	3	4	5
36. Our students are given instruction about behavioral objectives during the assessments.	1	2	3	4	5
37. Our district has taken test security measures to ensure results are not compromised.	1	2	3	4	5
38. Our district has monitoring procedures in place for inclusion, standardization, and security.	1	2	3	4	5
39. Our district provides training for those administering the assessments.	1	2	3	4	5
40. Our participation rates are documented.	1	2	3	4	5
41. Our district has local procedures in place for assuring appropriate accommodations for ELL students, students with disabilities, and students on 504 plans.	1	2	3	4	5
42. Our district has local assessment policies in place to assure comparability and consistency across the district.	1	2	3	4	5
<b>SUMMARIZING THE REVIEW PROCESS</b>					
43. I feel prepared to present my district portfolio to my peer reviewers.	1	2	3	4	5
44. I was provided time within the teaching day to prepare the district assessment portfolio.	1	2	3	4	5
45. Compensation is provided to prepare the district assessment portfolio when completed outside of the regular school day.	1	2	3	4	5
46. I have had adequate help in preparing the district assessment portfolio.	1	2	3	4	5
47. I have the necessary information to prepare the district assessment portfolio.	1	2	3	4	5

**PLEASE COMPLETE AND RETURN BEFORE YOU COMPLETE YOUR NEBRASKA LED PORTFOLIO PEER REVIEW.** Use the enclosed self-addressed postage-paid envelope or mail to:

Jody Isernhagen, Principal Investigator  
STARS Comprehensive Evaluation  
141 Teachers College Hall  
PO Box 880360, Lincoln, NE 68588-0360

**Appendix D**  
**2007-2008 Study I: Nebraska-led Math Portfolio**  
**Peer Review Interview Protocol**



**STARS PORTFOLIO PEER REVIEW INTERVIEW PROTOCOL  
2007-2008**

**Qualitative Research Purpose:** *Explore and understand the perceptions of educators about the STARS Portfolio Peer Review.*

Date of interview: \_\_\_\_\_ Time of interview: \_\_\_\_\_

Location of interview: \_\_\_\_\_

Interviewer: \_\_\_\_\_

**Participant Profile**

Participant: \_\_\_\_\_

District and School: \_\_\_\_\_

Position:   \_\_\_ Superintendent   \_\_\_ Assessment Coordinator   \_\_\_ Principal @   HS   MS   ELEM

          \_\_\_ Teacher   ELEM   MS   HS           Subject: \_\_\_\_\_

OTHER: \_\_\_\_\_

Years at present position and site: \_\_\_\_\_ Total Years in Education: \_\_\_\_\_

**Introduction:**

1. Thank you for taking the time to visit with me today.
2. I am serving as an interviewer for the STARS Comprehensive Evaluation conducted by the University of Nebraska-Lincoln. This research is being conducted so that the Nebraska Department of Education has a better understanding of how the STARS portfolio process is being implemented in school districts and schools across the state. Information gained from this research is used to improve the process and to provide insight into next steps.
3. First, I want to assure you that this interview is strictly confidential. Information provided by school and district staff is reported or released in aggregated form only. Districts, schools and individuals are not identified.
4. I have an Informed Consent form outlining your rights as a research participant. You are free to decide not to participate in this study or to withdraw from the study at any time without adversely affecting your relationship with the investigators, the University of Nebraska-Lincoln, or the Nebraska Department of Education. Contact persons for the project and the Institutional Review Board are provided on the Informed Consent Form in case you have questions or concerns. I have a copy for you to sign and one for you to keep for your use.
5. It is important that educators participating in this research be willing participants. You are free to decide not to participate or to withdraw from the interview at any time without harming your relationship with your district, this project, the University of Nebraska-Lincoln, or the Nebraska Department of Education. Should you decide not to participate you may either return to your normal activities or sit with me for the interview period. Are you willing to participate in this interview?
6. I am going to record this interview so that the interview can be transcribed (a typed copy of the interview will be made) and we have an accurate rendering of your responses.
7. It is important that I maintain the integrity of your words and intentions; therefore, I may ask you to review the transcription if I have any difficulties with the interpretation.
8. We are interested in finding out about the perceptions that you hold regarding the STARS assessment portfolio peer review process and its implementation in your school or district.

Questions about alignment, sufficiency, clarity, appropriateness, scoring procedures, and any other topics of interest that would help improve the process.

9. Please feel free to discuss your views openly. From time to time, I may have additional questions to further understand a concept that you have shared.
10. Let's begin. Please state your name, school, district and indicate permission to record this interview by repeating this statement, "I (your name) at (school/district name) willingly give my permission to record this interview."

**Interview Questions**

*DIRECTIONS: Place a check when the participant mentions each probe so that you do not repeat the probe.*

1. How have you been supported with the preparation of the assessment portfolio?

**Probes**

- \_\_\_\_\_ a. What type of training did you receive to prepare your assessment portfolio for the review and who provided the training?
- \_\_\_\_\_ b. Did the training you received prepare you for the actual peer review process?

<b>Descriptive Notes:</b>	<b>Reflective Notes</b>

2. When the peer review process began, what were your initial thoughts?

**Probes**

- \_\_\_\_\_ a. How did you know that your expertise was valued?
- \_\_\_\_\_ b. Share some of your thoughts regarding the first part of the review?

<b>Descriptive Notes:</b>	<b>Reflective Notes</b>

3. Tell us about how your assessments matched and measured your standards?

**Probes**

- \_\_\_\_\_ a. Tell us about the changes you will make to your alignment process due to the interaction with the peer reviewers.
- \_\_\_\_\_ b. What recommendations could you offer to other districts about matching and measuring the assessments to standards to improve the process (alignment)?

Descriptive Notes:	Reflective Notes

4. Share about the process of distributing assessment items across all performance levels (sufficiency)?

**Probes**

- \_\_\_\_\_ a. How did you assure that assessment performance descriptors were clear, differentiated, and increased expectations for each grade?
- \_\_\_\_\_ b. Tell us about any changes you may make in how you will distribute assessment items across all performance levels due to your interaction with the peer reviewers (sufficiency).
- \_\_\_\_\_ c. What recommendations could you offer to make the sufficiency process easier for other districts?

Descriptive Notes:	Reflective Notes

5. Talk about how your district provided directions that were clear and appropriate for all teachers and students (clarity)?

**Probes**

- \_\_\_\_\_ a. How do you report assessment results to students and parents?
- \_\_\_\_\_ b. How do you report to students and parents about special populations?
- \_\_\_\_\_ c. What recommendations could you offer to make assessment directions clearer for other districts administering assessments?

Descriptive Notes:	Reflective Notes

6. Share how your district ensured that the assessments were free of bias?

**Probes**

- \_\_\_\_\_ a. Tell us about any changes you may make in how you will ensure that assessments are from bias.
- \_\_\_\_\_ b. What recommendations could you offer to make the bias review process easier for other districts?

Descriptive Notes:	Reflective Notes

7. How did your district ensure that your assessments were appropriate for grade levels across the district?

**Probes**

- \_\_\_\_\_ a. How did you ensure that assessment expectations increased from one grade level to the next?
- \_\_\_\_\_ b. How did you plan for and administer the needed accommodations for students?
- \_\_\_\_\_ c. Tell us about the changes you will make to your assessments due to the interaction with the peer reviewers.
- \_\_\_\_\_ d. What recommendations could you offer to make assessments meet the standards of appropriateness easier for other districts?

Descriptive Notes:	Reflective Notes

8. Share how your district ensured that assessments were reliable and consistent.

**Probes**

- \_\_\_\_\_ a. Tell us about the methods you used for meeting reliability and consistency?
- \_\_\_\_\_ b. How will the review process help you improve your assessments?
- \_\_\_\_\_ c. Tell us about the changes you will make to your scoring procedures due to the interaction with the peer reviewers.
- \_\_\_\_\_ d. What recommendations could you offer to make the scoring procedures easier for other districts?

Descriptive Notes:	Reflective Notes

9. What new learnings have you had due to your involvement in the Nebraska-led Assessment Portfolio Peer Review?

**Probes**

- \_\_\_\_\_ a. You have stated (one, two or what ever has been stated) new learnings due to your involvement in the peer review process. Are there others?
- \_\_\_\_\_ b. What was the value of your new learnings to you as a professional and to your school or district?
- \_\_\_\_\_ c. Did the process meet your expectations?

<b>Descriptive Notes:</b>	<b>Reflective Notes</b>

10. Please share anything that you believe will strengthen the Nebraska-led Assessment Portfolio Peer Review.

<b>Descriptive Notes:</b>	<b>Reflective Notes</b>



**Appendix E**  
**2007-2008 Study II: STARS Enhancement**  
**Research Survey**



## ***STARS ENHANCEMENT SURVEY***

### ***STARS COMPREHENSIVE EVALUATION PROJECT***

---

Please complete this survey and return it to Jody Isernhagen, Primary Investigator, for the STARS Comprehensive Evaluation Project.

**Demographic Information**

*Please answer each question as it pertains to you and your school.*

- Gender  Male  
 Female
- Employment  NDE  
 Policy Partner  
 School District
- Primary Job Role
- Total years of experience working in education
- Approximate district student population (if applicable)
- Approximate school student population (if applicable)
- What is your experience level with STARS?  Very Experienced  
 Experienced  
 Little Experience  
 No Experience

<b>Statements</b>	<b>Strongly Agree</b>	<b>Agree</b>	<b>Neutral</b>	<b>Disagree</b>	<b>Strongly Disagree</b>
-------------------	-----------------------	--------------	----------------	-----------------	--------------------------

*These questions are in reference to the work of the four advisory teams designing changes for STARS based only on the current status LB 653. Read each statement and select the level to which you agree or disagree with that statement.*

- The purpose of the standards, assessment and reporting system is to support accountability for continuous improvement.
- The changes to STARS are in the best interest of ALL students.
- The changes to STARS promote best practices for teaching and learning.
- The changes to STARS meet federal regulations.
- The changes to STARS meet state requirements.
- The changes to STARS include a balance of classroom-based and large-scale assessment.
- The changes to STARS are still manageable for schools/districts/.
- After the changes to STARS, the system remains fair, equitable, and accurate.
- STARS is enhanced by the changes made by my team.

**Overall**

*Read each statement and select the level to which you agree or disagree with that statement.*

I discussed the STARS Enhancement Design with

others.

I understand the new requirements that will be implemented by the Enhancement of STARS.

I understand how the Enhancement of STARS will impact my school/district.

I feel that I have a responsibility to help improve STARS with the new enhancements.

I feel free to express my opinions about the enhancement of STARS with my team.

**Short Answer**

*These questions are in reference to the work of the four advisory teams designing changes for STARS based only on the current status LB 653. Read each statement and type answers in the space provided.*

How will implementing the STARS enhancements allow you to use the skills that you've learned in STARS?	
In your opinion, what are the strengths of the new enhancements of STARS?	
In your opinion, what are the limitations of the new enhancements to STARS?	
What impact will the STARS enhancements have on you as a professional?	
What impact will the STARS enhancements have on your school/district/organization?	
What impact will the STARS enhancements have on standards and assessments?	
<b>Additional Comments</b>	

*Please select the enhancement team to which you belong.*

Select the enhancement team of which you are a member.

- Enhancement Design Team
- Standards Advisory Team
- Assessment Advisory Team
- Reporting Advisory Team

***Only complete the following section for the enhancement team to which you belonged.***

Enhancement Design Team (members only)	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
<p><i>These questions are in reference only to the work of the Enhancement Design Team in designing changes for STARS based only on the current status LB 653. Read each statement and select the level to which you agree or disagree with that statement.</i></p>					
I agree with the STARS Enhancement Criteria that our team came up with.					
My expertise was put to good use in developing the STARS Enhancement Criteria.					
I feel our team understands the impact that LB 653 will have on Nebraska schools.					
Our team meetings were productive.					
Our team meetings were convenient for me to attend.					
The time that I spent working with my team was worthwhile.					
What impact do you believe that your team will have on STARS.					
Are there any changes that you believe need to occur to allow the Enhancement Design team to better function?					
<b>Additional Comments</b>					

Standards Advisory Team (members only)	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
<p><i>These questions are in reference only to the work of the Standards Advisory Team in designing changes and revising standards for STARS based only on the current status LB 653. Read each statement and select the level to which you agree or disagree with that statement.</i></p>					
I feel that the design criteria provided guidance to my team.					
I agree with the Standards Revision Criteria that our team came up with.					
My expertise was put to good use in developing the Standards Revision Criteria.					
I feel that the revision of the standards will enhance STARS.					
Our team meetings were productive.					
Our team meetings were convenient for me to attend.					

The time that I spent working with my team was worthwhile.

What impact do you believe that your team will have on STARS.

Are there any changes that you believe need to occur to allow the Standards Advisory Group to better function?

**Additional Comments**

**Assessment Advisory Team** (members only)

**Strongly Agree**

**Agree**

**Neutral**

**Disagree**

**Strongly Disagree**

*These questions are in reference only to the work of the Assessment Advisory Team in designing changes for STARS based only on the current status LB 653. Read each statement and select the level to which you agree or disagree with that statement.*

I feel that the design criteria provided guidance to my team.

I feel that the Standards Revision criteria provided guidance to my team.

My expertise was put to good use in developing the Assessment Advisory Group Recommendations.

Our team understands the impact that LB 653 will have on assessment.

Our team meetings were productive.

Our team meetings were convenient for me to attend.

The time that I spent working with my team was worthwhile.

What impact do you believe that your team will have on STARS.

Are there any changes that you believe need to occur to allow the Assessment Advisory Group to better function?

<b>Additional Comments</b>					
<b>Reporting Advisory Team</b> (members only)	<b>Strongly Agree</b>	<b>Agree</b>	<b>Neutral</b>	<b>Disagree</b>	<b>Strongly Disagree</b>
<i>These questions are in reference only to the work of the Reporting Advisory Team in designing changes for STARS based only on the current status LB 653. Read each statement and select the level to which you agree or disagree with that statement.</i>					
I feel that the design criteria provided guidance to my team.					
I feel that the Standards Revision Criteria provided guidance to my team.					
I feel that the Assessment Advisory Group Recommendations provided guidance to my team.					
My expertise was put to good use in determining how data are to be reported.					
Our team understands the impact that LB 653 will have on reporting.					
Our team meetings were productive.					
Our team meetings were convenient for me to attend.					
The time that I spent working with my team was worthwhile.					
What impact do you believe that your team will have on STARS.					
Are there any changes that you believe need to occur to allow the Reporting Advisory Group to better function?					
<b>Additional Comments</b>					

**You have completed the survey. Please accept our sincere thanks for you willingness to share your experiences and opinions with us.**



**Appendix F**  
**2007-2008 Study VII: The Effect of STARS on**  
**School Improvement Practices Survey**



**A Survey of Nebraska Second and Fourth-grade teachers' Perceptions  
of Nebraska's STARS System**

**Teaching Assignment:**

2<sup>nd</sup> Grade \_\_\_\_\_

4<sup>th</sup> Grade \_\_\_\_\_

**Directions:**

Please use the scale below to indicate the extent of your agreement or disagreement with each of the following statements about Nebraska's standards system (STARS).

1 = Strongly Disagree

2 = Disagree

3 = No Opinion

4 = Agree

5 = Strongly Agree

1. In my building, the implementation of STARS has led to improved teacher morale.  
1            2            3            4            5
2. In my building, the implementation of STARS has led to improved student achievement.  
1            2            3            4            5
3. In my building, the implementation of STARS has led to improved curriculum alignment.  
1            2            3            4            5
4. In my building, the implementation of STARS has led to a broadening of the curriculum.  
1            2            3            4            5
5. In my building, the implementation of STARS has led to increased use of achievement data to guide classroom instruction.  
1            2            3            4            5
6. In my building, the implementation of STARS has led to improved classroom assessment practices.  
1            2            3            4            5
7. In my building, the implementation of STARS has led to meaningful professional development.  
1            2            3            4            5
8. In my building, the implementation of STARS has led to an increase in positive student participation in the classroom.  
1            2            3            4            5
9. In my building, the implementation of STARS has led to an increase in time teachers spend collaborating with other teachers designing strategies for teaching state content standards.  
1            2            3            4            5
10. In my building, the implementation of STARS has led to improved communication between teachers and students regarding learning.  
1            2            3            4            5

11. In my building, the implementation of STARS has led to expanded opportunities to assist struggling learners.  
1            2            3            4            5
12. In my building, the implementation of STARS has led to an increase in teachers' willingness to try different instructional strategies.  
1            2            3            4            5
13. In my building, the implementation of STARS has led to assessments that measure content or skills that is meaningful to the school community.  
1            2            3            4            5
14. In my building, the implementation of STARS has led to the development of a more rigorous curriculum.  
1            2            3            4            5
15. In my building, the implementation of STARS has led to the building meeting school improvement goals.  
1            2            3            4            5
16. In my building, the implementation of STARS has led to an increase in time spent in test preparation activities.  
1            2            3            4            5
17. In my building, the implementation of STARS has led to increased communication with the building principal regarding student achievement.  
1            2            3            4            5
18. In my building, the implementation of STARS has led to increased opportunities for teachers to be effective instructional leaders.  
1            2            3            4            5
19. In my building, the implementation of STARS has led to an improved public image of teachers.  
1            2            3            4            5
20. In my building, the implementation of STARS has led increased effectiveness in meeting individual student needs.  
1            2            3            4            5
21. In my building, the implementation of STARS has led to an increase of non-traditional learning opportunities for students.  
1            2            3            4            5
22. In my building, the implementation of STARS has led to raising achievement expectations for all students.  
1            2            3            4            5

**Appendix G**  
**2007-2008 Study VIII: Review of the Local**  
**Math Assessment Process District**  
**Assessment Portfolio Worksheets**



-----District Assessment Portfolio Worksheet -----

**Reliability of Peer Review Process**

Portfolio Number \_\_\_\_\_ Rater ID \_\_\_\_\_ Grade Level \_\_\_\_\_

*The purpose of this performance exercise is to determine reliability/consistency in reviewing district assessment portfolios using the Assessment Quality Rubric.*

Criterion One			<b>The assessments match the standards.</b>
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	<ul style="list-style-type: none"> <li>• Qualifications of the independent reviewers are clear and complete.</li> </ul>
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	<ul style="list-style-type: none"> <li>• Evidence of an independent review for match to standards is clear and complete (reviewers did not write the assessments.)</li> </ul>
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	<ul style="list-style-type: none"> <li>• The process for matching assessments to standards is clear and complete.</li> </ul>
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	<ul style="list-style-type: none"> <li>• Results of the matching process are clear and complete.</li> </ul>
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	<ul style="list-style-type: none"> <li>• Sufficiency process is clear and complete.</li> </ul>
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	<ul style="list-style-type: none"> <li>• Sufficiency results are clear and complete (sufficiency required for both number of items/ performances and levels of difficulty. Minimum 12 items or equivalent on reading standards 4.1.3, 8.1.1 and 12.1.1 and math standards 4.2.1, 8.2.2, and 12.2.1)</li> <li>• <i>*Districts with local standards must designate a reading and a math standard.</i></li> </ul>
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	<ul style="list-style-type: none"> <li>• Consistency between Criterion #1 and other criteria is clear.</li> </ul>
<b>Suggested RATING</b>			<input type="checkbox"/> Not Met <input type="checkbox"/> Needs Improvement <input type="checkbox"/> Met w/Comment <input type="checkbox"/> Met
Evidence supporting Criterion one:			
Evidence needed to meet Criterion One:			

-----District Assessment Portfolio Worksheet -----

--

**Reliability of Peer Review Process**

Portfolio Number \_\_\_\_\_ Rater ID \_\_\_\_\_ Grade Level \_\_\_\_\_

Criterion Two			<b>Students have an opportunity to learn.</b>
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• Qualifications of the opportunity to learn reviewers are clear and complete.
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• The process for alignment of standards with local curriculum is clear and complete.
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• The process for timing of assessment/instruction is clear and complete.
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• The results of the process for alignment of standards with local curriculum are clear and complete.
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• Dates are provided when standards are taught and they are clear and complete.
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• Dates are provided when standards are assessed and are clear and complete
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• 80% of instruction should take place prior to assessment.
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• Consistency between Criterion #2 and other criteria is clear and complete.
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• Opportunity to learn information provided for all standards.
<b>Suggested RATING</b>			<input type="checkbox"/> Not Met <input type="checkbox"/> Needs Improvement <input type="checkbox"/> Met w/Comment <input type="checkbox"/> Met
Evidence supporting Criterion two:			
Evidence needed to meet Criterion Two:			

-----District Assessment Portfolio Worksheet -----

---

**Reliability of Peer Review Process**

Portfolio Number \_\_\_\_\_ Rater ID \_\_\_\_\_ Grade Level \_\_\_\_\_

Criterion Three			<b>The assessments are free of bias and sensitive situations.</b>
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• Qualifications of the bias reviewers are clear and complete.
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• The description of the bias orientation/training process is clear and complete.
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• The process for bias review of assessment items is clear and complete.
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• Results of a bias review are clear and complete.
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• Bias information provided for all standards (used for reporting.)
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• Consistency between criterion #3 and other criteria is clear and complete.
<b>Suggested RATING</b>			<input type="checkbox"/> Not Met <input type="checkbox"/> Needs Improvement <input type="checkbox"/> Met w/Comment <input type="checkbox"/> Met
Evidence supporting Criterion Three:			
Evidence needed to meet Criterion Three:			

-----District Assessment Portfolio Worksheet -----

-----

**Reliability of Peer Review Process**

**Portfolio Number** \_\_\_\_\_ **Rater ID** \_\_\_\_\_ **Grade Level** \_\_\_\_\_

Criterion Four			<b>The assessments are at the appropriate level.</b>
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• Qualifications of the reviewers for appropriate level are clear and complete.
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• Process for appropriate level review is clear and complete.
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• Results of the appropriate level review are clear and complete.
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• Appropriate level information is provided for all standards (used for reporting.)
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• Consistency between Criterion #4 and other criteria is clear and complete.
<b>Suggested RATING</b>			<input type="checkbox"/> Not Met <input type="checkbox"/> Needs Improvement <input type="checkbox"/> Met w/Comment <input type="checkbox"/> Met
Evidence supporting Criterion Four:			
Evidence needed to meet Criterion Four:			

-----District Assessment Portfolio Worksheet -----

--

**Reliability of Peer Review Process**

Portfolio Number \_\_\_\_\_ Rater ID \_\_\_\_\_ Grade Level \_\_\_\_\_

Criterion Five			<b>There is consistency in scoring.</b>
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• Qualifications of the reliability process participants are clear and complete.
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• Appropriate process for reliability is clear and complete.
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• Reliability value provided and calculations are at or above the minimum acceptable level. (Minimum level of acceptable reliability is .70, mean or median, averaged across all standards.)
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• Procedure for improving reliability is clear and complete.
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• Reliability is reported for all standards (used for reporting.)
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• Consistency between Criterion #5 and other criteria is clear and complete.
<b>Suggested RATING</b>			<input type="checkbox"/> Not Met <input type="checkbox"/> Needs Improvement <input type="checkbox"/> Met w/Comment <input type="checkbox"/> Met
Evidence supporting Criterion Five:			
Evidence needed to meet Criterion Five:			

-----District Assessment Portfolio Worksheet -----

---

**Reliability of Peer Review Process**

Portfolio Number \_\_\_\_\_ Rater ID \_\_\_\_\_ Grade Level \_\_\_\_\_

Criterion Six			Mastery levels are appropriate
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• Qualifications for mastery level participants are clear or complete.
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• Evidence of mastery level process is clear or complete.
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• Results of the mastery level process are clear and complete.
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• Mastery level information is provided for all standards (used for reporting.)
Missing <input type="checkbox"/>	Incomplete <input type="checkbox"/>	Complete <input type="checkbox"/>	• Consistency between criterion #6 and other criteria is clear and complete.
Suggested RATING			<input type="checkbox"/> Not Met <input type="checkbox"/> Needs Improvement <input type="checkbox"/> Met w/Comment <input type="checkbox"/> Met
Evidence supporting Criterion Six:			
Evidence needed to meet Criterion Six:			

**Appendix H**  
**2007-2008 Study VIII: 2207-2008 Peer Reviewer**  
**Training Evaluation Survey**



## 2007-2008 Peer Review Training Evaluation Survey

**Reviewer ID** \_\_\_\_\_

### Reviewer Evaluation

Thank you for taking a few moments to complete this survey. Your responses will assist the Nebraska Department of Education in the continued implementation of STARS.

### Reviewer Information

1. Gender (circle) M F
2. Years of classroom teaching experience \_\_\_\_\_ Years of educational experience \_\_\_\_\_
3. Highest Level of education \_\_\_\_\_ Bachelors \_\_\_\_\_ Masters \_\_\_\_\_ Ed Specialist \_\_\_\_\_ Doctorate
4. Did you attend the 2006-07 training in the Peer Review Process? (circle) Yes No
5. Have you previously participated as a rater? (circle) Yes No
6. How many schools have you reviewed in 2006-07 and 2007-08? \_\_\_\_\_
7. What position do you currently hold in (circle one) school district, ESU, NDE, or college? Check all that apply.  
\_\_\_\_\_ teacher \_\_\_\_\_ principal \_\_\_\_\_ superintendent  
\_\_\_\_\_ staff developer \_\_\_\_\_ curriculum/assessment \_\_\_\_\_ retired
8. How big is your district – student enrollment? (circle) Less than 500, 500 to 2500, More than 2500

### Reviewer Training

Please indicate your level of agreement or disagreement with the following items by circling your choice (SD = Strongly Disagree: D = Disagree: N = Neither Agree nor Disagree: A = Agree: SA = Strongly Agree)

1. SD D N A SA The general content of the training for reviewers was appropriate (i.e. overview of the review process, understanding of rubric, 6 quality criteria)
2. SD D N A SA Strategies implemented during training to ensure reviewer accuracy were effective (i.e. strategies for rubric use, exemplars of portfolios, practice portfolio reviews)
3. SD D N A SA The training materials (i.e. rubric, worksheets, exemplars, guidebook) used during the training and review sessions were clear and understandable.
4. SD D N A SA The training leaders demonstrated expert knowledge of the review process.
5. SD D N A SA The training leaders demonstrated skill in teaching raters how to fairly and accurately review district assessment portfolios.

Comments: Please share any additional comments/thoughts you would like to share about your training experience on the back of this sheet.