AN EXAMINATION OF DEVELOPMENT OF WYOMING’S ALTERNATIVE ASSESSMENT SYSTEM, THE BODY OF EVIDENCE

by

Sharla Kay Dowding

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Education in Education

MONTANA STATE UNIVERSITY
Bozeman, Montana

July 2011
APPROVAL

of a dissertation submitted by

Sharla Kay Dowding

This dissertation has been read by each member of the dissertation committee and has been found to be satisfactory regarding content, English usage, format, citation, bibliographic style, and consistency and is ready for submission to The Graduate School.

Dr. Elisabeth Swanson

Approved for the Department of Education

Dr. Jayne Downey

Approved for The Graduate School

Dr. Carl A. Fox
STATEMENT OF PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this dissertation is allowable only for scholarly purposes, consistent with “fair use” as prescribed in the U.S. Copyright Law. Requests for extensive copying or reproduction of this dissertation should be referred to ProQuest Information and Learning, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted “the exclusive right to reproduce and distribute my dissertation in and from microform along with the non-exclusive right to reproduce and distribute my abstract in any format in whole or in part.”

Sharla Kay Dowding

July 2011
ACKNOWLEDGEMENTS

I would like to acknowledge those people who supported me through the process of completing my dissertation. First, I would like to thank my husband, Rod, for being patient with me over the last six years and encouraging me to “finish what you start.” Thank you for letting me embark on this challenge and see it to the completion. I would like to thank my committee – especially Dr. Elisabeth Swanson for her reviewing, editing, commenting and overall support throughout my time in the CLTW program. Dr. Art Bangert provided insights and recommendations in preparing the statistical analysis. Dr. Maurice Burke, Dr. Walt Woolbaugh, and Dr. Carrie Meyers provided continued methodology guidance, encouragement and editorial comments. Thank you to Dr. Geoff Poole for stepping in at the last minute as my grad rep. Thank You also to Walt and Rick Jones for introducing me to the CLTW program. I am grateful for the opportunity to have meet so many inspiring teachers! In addition, a huge thank you to Micki MacGregor for always keeping me informed about deadlines and due dates. I would like to thank Dottie Thorson for her help and for sharing the information she always has at her fingertips, Phyl Sundstrum for editing. Tracy Ragland, Doug Scribner and Melissa Cade were constant sounding boards through the process. I would also like to thank Dr. Alan Moore for reading and providing helpful suggestions. In addition, I am very appreciative to all of my interview participants and survey respondents!
TABLE OF CONTENTS

1. INTRODUCTION .................................................................................................................. 1

   Background ......................................................................................................................... 1
   Wyoming’s Framework for Assessing Skills ................................................................. 2
      PAWS Test Design ......................................................................................................... 4
      History of the PAWS ....................................................................................................... 6
      Alignment of PAWS to Standards ................................................................................. 8
   District Assessment Systems: Body of Evidence (BOE) ........................................... 8
   Statement of Problem ..................................................................................................... 9
      BOE Peer Review Teams ............................................................................................... 11
   Research Questions ......................................................................................................... 12
   Definition of Key Terms ................................................................................................. 14
   Significance of Study ...................................................................................................... 18
   Potential Limitations of Study ...................................................................................... 18

2. REVIEW OF LITERATURE ................................................................................................. 21

   Introduction ...................................................................................................................... 21
   Topic I. Perspectives from Cognitive Science ............................................................. 21
   Topic II: Principles & Components of Assessment .................................................... 31
      Principles of Quality Assessments ............................................................................ 32
      Design of Large-Scale Assessment – Prevailing Practices .................................. 33
      Performance-Based Assessment ............................................................................... 36
      Learner-Centered Assessment .................................................................................. 37
      Instructionally Supportive Assessments ................................................................ 38
   Topic III: Literature that relates IV (Assessment Design) to
   DV (Wyoming’s BOE Development) ........................................................................... 41
      Wyoming’s Framework for Assessing Skills .......................................................... 44
      History of the PAWS .................................................................................................... 45
      PAWS Test Design ...................................................................................................... 47
      Alignment of PAWS to Standards ............................................................................ 48
   District Assessment Systems: Body of Evidence (BOE) .......................................... 49
      Alignment .................................................................................................................... 51
      Consistency .................................................................................................................. 52
      Fairness ....................................................................................................................... 52
      Standard Setting ....................................................................................................... 52
      Comparability ............................................................................................................ 53
      Multiple Measures .................................................................................................... 54
      Credibility .................................................................................................................. 54
      Consequences ............................................................................................................ 54
      Rewards and Sanctions ............................................................................................. 55
TABLE OF CONTENTS - CONTINUED

Summary ............................................................................................................................ 56

3. RESEARCH METHODOLOGY.................................................................................. 58
   Introduction.................................................................................................................... 58
   Context of the Study .................................................................................................... 58
   Research Perspectives ............................................................................................... 60
   Research Purpose and Focus Questions ..................................................................... 61
   Research Design ......................................................................................................... 63
   Development, Implementation and Analysis of the Orienting Survey ................. 68
      Pilot Survey ............................................................................................................ 68
      Survey Implementation ......................................................................................... 70
      Instrument Reliability ............................................................................................ 72
      Survey Data Analysis ............................................................................................ 73
   Development, Implementation and Analysis of the Case Studies ..................... 74
      Case Participants .................................................................................................. 74
      Pilot Case Study Interviews .................................................................................. 77
      Case Study Implementation .................................................................................. 78
      Case Study Data Analysis ..................................................................................... 79
      Case Study BOE Analysis ..................................................................................... 80
   Study Fidelity and Quality ......................................................................................... 82
   Researcher Perspective .............................................................................................. 84
   Delimitations and Limitations ................................................................................... 85

4. RESULTS OF THIS STUDY .................................................................................... 88
   Introduction ............................................................................................................... 87
   BOE Survey ............................................................................................................... 87
      BOE Participation ................................................................................................. 89
      BOE Activity Structure ......................................................................................... 91
      Likert Items .......................................................................................................... 96
      Open-Ended Responses ...................................................................................... 100
      Summary of Survey Responses ........................................................................... 100
   District Case Studies ............................................................................................... 102
      Cottonwood .......................................................................................................... 106
         BOE Characteristics ......................................................................................... 107
         Alignment to State Standards ......................................................................... 109
         Additional Wyoming BOE Design Principles ............................................... 110
         Support for BOE Processes ............................................................................. 112
         Data Reporting and Usage ................................................................ .......... 112
         Reflection about Student Learning ................................................................. 113
      Cypress ............................................................................................................... 114
TABLE OF CONTENTS – CONTINUED

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOE Characteristics</td>
<td>115</td>
</tr>
<tr>
<td>Alignment to State Standards</td>
<td>116</td>
</tr>
<tr>
<td>Additional Wyoming BOE Design Principles</td>
<td>117</td>
</tr>
<tr>
<td>Support for BOE Processes</td>
<td>120</td>
</tr>
<tr>
<td>Data Reporting and Usage</td>
<td>121</td>
</tr>
<tr>
<td>Reflection about Student Learning</td>
<td>122</td>
</tr>
<tr>
<td>Larch</td>
<td>123</td>
</tr>
<tr>
<td>BOE Characteristics</td>
<td>124</td>
</tr>
<tr>
<td>Alignment to State Standards</td>
<td>125</td>
</tr>
<tr>
<td>Additional Wyoming BOE Design Principles</td>
<td>125</td>
</tr>
<tr>
<td>Support for BOE Processes</td>
<td>127</td>
</tr>
<tr>
<td>Data Reporting and Usage</td>
<td>128</td>
</tr>
<tr>
<td>Reflection about Student Learning</td>
<td>129</td>
</tr>
<tr>
<td>Linden</td>
<td>131</td>
</tr>
<tr>
<td>BOE Characteristics</td>
<td>132</td>
</tr>
<tr>
<td>Alignment to State Standards</td>
<td>134</td>
</tr>
<tr>
<td>Additional Wyoming BOE Design Principles</td>
<td>135</td>
</tr>
<tr>
<td>Support for BOE Processes</td>
<td>137</td>
</tr>
<tr>
<td>Data Reporting and Usage</td>
<td>138</td>
</tr>
<tr>
<td>Reflection about Student Learning</td>
<td>139</td>
</tr>
<tr>
<td>Willow</td>
<td>140</td>
</tr>
<tr>
<td>BOE Characteristics</td>
<td>141</td>
</tr>
<tr>
<td>Alignment to State Standards</td>
<td>142</td>
</tr>
<tr>
<td>Additional Wyoming BOE Design Principles</td>
<td>143</td>
</tr>
<tr>
<td>Support for BOE Processes</td>
<td>145</td>
</tr>
<tr>
<td>Data Reporting and Usage</td>
<td>146</td>
</tr>
<tr>
<td>Reflection about Student Learning</td>
<td>147</td>
</tr>
<tr>
<td>Tesota</td>
<td>148</td>
</tr>
<tr>
<td>BOE Characteristics</td>
<td>149</td>
</tr>
<tr>
<td>Alignment to State Standards</td>
<td>151</td>
</tr>
<tr>
<td>Additional Wyoming BOE Design Principles</td>
<td>152</td>
</tr>
<tr>
<td>Support for BOE Processes</td>
<td>154</td>
</tr>
<tr>
<td>Data Reporting and Usage</td>
<td>155</td>
</tr>
<tr>
<td>Reflection about Student Learning</td>
<td>156</td>
</tr>
<tr>
<td>Cross-Case Analysis</td>
<td>158</td>
</tr>
<tr>
<td>Domains of BOE Development and Implementation</td>
<td>159</td>
</tr>
<tr>
<td>BOE plans that utilize a combination of Wyoming BOE consortium activities along with district-developed assessments</td>
<td>160</td>
</tr>
<tr>
<td>BOE plans that utilize only district developed assessments</td>
<td>162</td>
</tr>
<tr>
<td>BOE plans that are in a state of revision</td>
<td>163</td>
</tr>
<tr>
<td>Nature of District Support</td>
<td>164</td>
</tr>
<tr>
<td>Comprehensive Support</td>
<td>166</td>
</tr>
</tbody>
</table>
TABLE OF CONTENTS – CONTINUED

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marginal Support</td>
<td>167</td>
</tr>
<tr>
<td>Severely Limited Support</td>
<td>168</td>
</tr>
<tr>
<td>Communication Factors</td>
<td>168</td>
</tr>
<tr>
<td>Comparison of Findings across Quantitative and Qualitative Methods</td>
<td>171</td>
</tr>
<tr>
<td>Summary of Findings</td>
<td>173</td>
</tr>
<tr>
<td>5. FINDINGS, CONCLUSIONS, AND IMPLICATIONS</td>
<td>175</td>
</tr>
<tr>
<td>Introduction</td>
<td>175</td>
</tr>
<tr>
<td>Review of the Study Context</td>
<td>175</td>
</tr>
<tr>
<td>Findings of the Study</td>
<td>177</td>
</tr>
<tr>
<td>Research Question 1</td>
<td>177</td>
</tr>
<tr>
<td>Research Question 2</td>
<td>180</td>
</tr>
<tr>
<td>Research Question 3</td>
<td>181</td>
</tr>
<tr>
<td>Implications</td>
<td>185</td>
</tr>
<tr>
<td>Recommendations for Policy Makers</td>
<td>186</td>
</tr>
<tr>
<td>Recommendations for State –Level Leadership</td>
<td>187</td>
</tr>
<tr>
<td>Recommendations for District Level Coordinators</td>
<td>190</td>
</tr>
<tr>
<td>Recommendations for Teacher Leaders</td>
<td>192</td>
</tr>
<tr>
<td>Suggestions for Further Research</td>
<td>193</td>
</tr>
<tr>
<td>Looking Ahead</td>
<td>195</td>
</tr>
<tr>
<td>Summary</td>
<td>196</td>
</tr>
<tr>
<td>REFERENCES CITED</td>
<td>197</td>
</tr>
<tr>
<td>APPENDICES</td>
<td>207</td>
</tr>
<tr>
<td>APPENDIX A: BOE Activity Alignment</td>
<td>208</td>
</tr>
<tr>
<td>APPENDIX B: BOE Survey</td>
<td>212</td>
</tr>
<tr>
<td>APPENDIX C: Interview Questions</td>
<td>219</td>
</tr>
<tr>
<td>APPENDIX D: Research Question Matrix</td>
<td>226</td>
</tr>
<tr>
<td>APPENDIX E: Body of Evidence Peer Review Rubrics</td>
<td>231</td>
</tr>
<tr>
<td>APPENDIX F: Excerpt from NCLB Law</td>
<td>238</td>
</tr>
<tr>
<td>APPENDIX G: Excerpt from Adequate Yearly Progress</td>
<td>240</td>
</tr>
<tr>
<td>APPENDIX H: Wyoming House Bill 0078</td>
<td>243</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table Table
1. Cognitive Perspectives Supporting Learner-Centered Assessment 26
2. Comparisons of Learning Perspective Skills 27
3. Learner-Centered Assessment Practices 29
4. Overview of Quantitative and Qualitative Data Sources to Address the Research Questions 65
5. School District Demographics 75
6. Number of Years and Current Teaching Assignment of Interview Participants 77
7. Roles of Survey Participants and Number of Years in Current Position 88
8. Results Regarding Participation in BOE Peer-Review at the State Level by Respondent Role 89
9. Cross-tabulation Results of Role x Participation in BOE-science Plan Development 90
10. Cross-tabulation of Role by Familiarity with District Science Matrix 91
11. Number of Consortium Activities Used in BOE Science Assessment Plans 95
12. Results of Likert-scale Questions about BOE Impact 97
13. Range of Responses to Experience Developing the BOE Plan 97
14. Range of Responses to Experience Implementing the BOE Plan 98
15. Range of Opinion about Impacts on Student Learning as a Result of BOE Plan 99
16. Means of Likert Items for Respondents whom had Peer Review Experience 100
17. Wyoming BOE Design Principle Terminology 105
<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>18. Domains of BOE Development and Case Study Classifications</td>
<td>159</td>
</tr>
<tr>
<td>19. Participant Engagement in BOE Development &amp; Implementation</td>
<td>164</td>
</tr>
<tr>
<td>Activities</td>
<td></td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The Standards-Based System in Wyoming</td>
<td>3</td>
</tr>
<tr>
<td>2. Four Spheres of Work in Educational Assessment Practice</td>
<td>30</td>
</tr>
<tr>
<td>3. Content-process Space of Assessment Tasks</td>
<td>81</td>
</tr>
<tr>
<td>4. Integration of Assessment Pieces</td>
<td>92</td>
</tr>
<tr>
<td>5. Multiple Opportunities for Assessment Pieces</td>
<td>93</td>
</tr>
<tr>
<td>6. Use of Consortium BOE Activities in a Science BOE Plan</td>
<td>94</td>
</tr>
</tbody>
</table>
ABSTRACT

The overarching purpose of this qualitative study is to explore the patterns of development and implementation of Body of Evidence (BOE) science systems throughout the state of Wyoming, using an emerging and relatively open mixed methods design. BOEs were first launched throughout Wyoming a decade ago, and are ongoing today. Through interviews with teachers and curriculum coordinators and analysis of BOE science plans, the following research questions were explored: 1) What design elements and implementation processes characterize BOE science assessment plans across Wyoming school districts? 2) What support do teachers receive to develop a BOE science plan, to align the plan to the state science education standards, and to implement the intended science plan? And 3) How do districts handle this opportunity to design and implement a locally useful assessment?

Methods of data collection used in this mixed methods study began with a survey, conducted in Spring 2010, to investigate characteristics of the BOE systems being implemented within and across Wyoming school districts, followed by district case studies, implemented in Fall 2010, that included an interview with the curriculum coordinator and a science teacher in each district and a review of the BOE plan for that district. A total of 110 survey responses were received and analyzed. The six schools that participated in the interview component of the study represent a range of Wyoming school districts.

The first set of themes reveals the ways BOE assessments and curriculums are aligned to state science teaching standards. The second set of themes depicts the nature of district support that is provided for the BOE development and implementation. Finally, because BOE is an ongoing process, a third theme was identified that discovered factors that affect the communication and support needed to sustain BOE implementation. The results presented from both the survey and case study interviews indicated that there is a great variation in how this state-mandated assessment requirement is implemented and sustained across school districts in Wyoming. It was obvious that the perceived levels and categories of support are different both from district to district and within districts.
CHAPTER 1

INTRODUCTION

Background

At no other time in the history of American education has large-scale assessment been in the spotlight as it is currently. Public Law 107-110, known as the No Child Left Behind Act (NCLB) of 2001, made major amendments to the Elementary and Secondary Education Act of 1965, which is where standardized testing found its roots (Popham, 2001). NCLB mandates each state to develop and implement a single, statewide accountability system. In this era of high-stakes testing, how can a classroom teacher prepare students to do well on large-scale assessments without narrowing the curriculum and teaching to the test?

The Proficiency Assessments for Wyoming Students (PAWS) summative assessments are designed to meet the current NCLB testing requirements. Furthermore, Wyoming’s graduation requirements include “evidence of proficient performance, at a minimum, on the uniform student content and performance standards for the common core of knowledge and skills” (Wyoming Department of Education, 2003a). Although many states nationwide require students to pass a single, state-wide graduation exit exam, in Wyoming state legislation requires a unique component to the Wyoming Assessment System in the use of multiple assessments by which students can demonstrate proficiency on the standards (Wyoming Department of Education, 2005). Submitted in a Body of Evidence Plan (BOE), individual districts develop data-driven plans which reflect more
ways for students to show what they know than can be measured on a single, high-stakes assessment. BOE plans are difficult to use as large-scale assessments for a few reasons. First, performance assessments are more expensive to grade than multiple-choice tests, due to each piece needing to be scored by two or more raters. Second, although common assessment rubrics have been developed for use statewide for several common activities in science, inter-rater reliability is lower than on multiple-choice questions which have obvious correct responses (Dr. Alann Moore, personal communications, July 27, 2009). Jones, Jones & Hargrove (2003) describe similar difficulties when Kentucky used a state math portfolio in 1996. Unfortunately, Kentucky’s math portfolio project was short-lived.

**Wyoming’s Framework for Assessing Skills**

A major difference between Wyoming BOE plans and other state portfolio assessments is that four-point/task rubrics were developed for 28 science activities to be used across the state. Through a Body of Evidence Consortium, activities were first piloted, followed by a gathering of student work samples that a team of teachers evaluated in order to write indicators of advanced, proficient, partially proficient and novice performance. Using these indicators, rubrics were designed for the assessments. The rubrics were field tested, and the teacher team gathered again to evaluate student work against the rubric for inter-rater reliability. Each consortium activity is accompanied by an alignment chart to the Wyoming State Science Standards. An example of this alignment piece can be seen in Appendix A. After this two-year process,
the assessments mentioned above were placed into the statewide BOE Consortium activity bank. This is a lengthy process that cannot be feasibly done for every classroom assessment, but it has been used to develop science activities commonly used around the state in BOE plans.

Figure 1 organizes the framework of the standards-based system in Wyoming. The right-hand assessment bubble could be further sub-divided. The state assessment is the PAWS test. The district assessment is the BOE plan, which can include selected BOE Consortium activities and even PAWS scores in certain districts. The classroom assessments are individual teacher grades.

Figure 1. The Standards-Based System in Wyoming. (Wyoming Department of Education, 2008b, p.3).
BOE plans (from every Wyoming school district) have been scrutinized by peer review processes in 2003 and in 2009. In contrast, the first data release from state-wide testing in science didn’t occur until 2008. Because Wyoming’s assessment system is different from other states in that it uses multiple assessments instead of a single, high-stakes test to determine graduation requirements, I am interested in knowing what characteristics emerge from an investigation into the ways school districts go about designing, implementing and using data from their BOE science plans.

**PAWS Test Design**

Wyoming uses a two-tiered assessment process, a state test designed to meet NCLB testing requirements and a BOE assessment system mandated by the state legislature. Readers will benefit from a history of this system. The Proficiency Assessments for Wyoming Students (PAWS) summative assessments are designed to meet the current NCLB testing requirements for language arts and mathematics in grades 3-8, and once in the grade span of 10-12. In March of 2007 the PAWS science exam was field-tested in Wyoming schools. In April 2008 the official PAWS-science assessment was administered and student results evaluated. Students are tested for science in grades 4, 8, and 11. The science test has three subtests: life-science, physical-science, and earth/space science. Wyoming’s framework for assessing science is based on the skills of science as inquiry, and therefore each subtest is designed to address a component of inquiry-based science.

In March 2010 multiple choice questions of the PAWS were completed online, and extended response questions were written in test booklets. The goal was to
eventually have all sections of the PAWS administered online, but due to technical difficulties at both Pearson and in Wyoming school districts, the exam was solely on paper in March 2011. It is expected to be administered on paper in 2012 as well. There are suggested time parameters for administering the PAWS that will help districts in planning test administration, but there are no time limits within which a student must complete each subset of the PAWS. Students on Individual Education Plans (IEPs) may take the PAWS in the special education room where the exam can be read to them.

The Wyoming Department of Education (WDE) website (www.k12.wy.us) hosts the most extensive information about the PAWS. Finding specific information can be a bit tricky in the web-of-information. However, by locating the PAWS homepage (http://edu.wyoming.gov/Programs/Statewide_Assessment_System.aspx), one can navigate numerous pdf-files and word-documents. Searching the PAWS website, under the links for “Grade Specific Resources” one will find in-depth “assessment descriptions” for math, reading, writing, and science. These are lengthy documents that describe Skills, Cognitive Processes, Context and Evaluative Criteria for Mathematical Skills, Science Skills, Reading Skills, and Writing Skills. The documents also provide a rationale for the PAWS-science assessment.

The Wyoming Science Content and Performance Standards ask students to demonstrate the knowledge, skills, and habits of mind necessary to safely perform scientific inquiry. To achieve this end, the PAWS assessment asks students to connect knowledge of scientific concepts with the problem at hand, design procedures to find an answer to the problem, investigate phenomena through data collection, and construct meaning through the use of evidence and logic. Therefore, Wyoming’s framework for assessing science is based on the skills of science as inquiry. (Wyoming Department of Education, 2005, p.1)
In the past, teachers have been faced with the perplexing task of trying to prepare their students to take the state test without knowing which standards/skills would or would not be assessed on a given form of the test. To ensure that PAWS is instructionally supportive, assessment descriptions have been developed to clearly and thoroughly describe the knowledge and skills that will be tested and the evaluative criteria that will be used to assess the skill.

History of the PAWS

The Commission on Instructionally Supportive Assessment (2001) presented nine requirements of statewide achievement tests to ensure exams both benefit students, by providing educators with information they can use to improve instruction, and also provides information for accountability purposes. The Wyoming Department of Education has taken these recommendations to heart in the creation of the state assessment system, and in particular in creation of the PAWS tests. “PAWS represents the nation’s first attempt to construct NCLB tests in complete accord with key recommendations of a national advisory group, the Commission on Instructionally Supportive Assessment” (Popham, 2005c, p. 3).

Reading the PAWS technical report (Wyoming Department of Education, 2010) one will find PAWS history. Here we see the first reference is the Wyoming Statewide Task Force on Student Assessment & Education Accountability Report & Recommendations from October 2003. “Pursuant to State law, and based on six months of deliberation and outreach, the Task Force recommends modifications to Wyoming’s
statewide assessment system and the establishment of a single statewide accountability system with consequences” (p. 2).

In this report, the Task Force recommended the establishment of a statewide assessment system that includes a summative assessment. The Task Force recommends modifications designed to improve teaching and learning, foster school improvement, and enhance Wyoming’s compliance with the NCLB law. As a result, PAWS replaced WyCAS as the statewide accountability assessment. The WyCAS (Wyoming Comprehensive Assessment System) was initially designed to comply with the provisions of the 1994 reauthorization of the Elementary and Secondary Education Act. The change in focus from WyCAS to PAWS reflects a shift in assessment philosophy toward instructionally supportive assessments. Rewriting the statewide summative assessment, the Wyoming Department of Education decided to follow guidance recommendations from the Commission on Instructionally Supportive Assessment. In the PAWS Technical Report one can also read about the PAWS test design and development, test administration, scoring procedures, and how scores are reported.

PAWS is produced by Pearson Assessment Inc. with input on item design, standard setting and scoring from Wyoming teachers, administrators and assessment expert, James Popham. The first science field-test was not able to be administered in 2006 because Wyoming teachers did not feel it was standards-based or that the questions were aligned correctly to cognitive levels and it lacked an inquiry focus. This was worked on for another year and field-testing occurred in April 2007. The first classes received PAWS-science scores for data interpretation in April 2008. Data interpretation
by Harcourt (now Pearson) started in May 2008. Re-alignment to Wyoming State
Science Standards occurred in July 2008. Unlike the process during initial assessment
development, data review at the state level now occurs annually in late July of each year.

Alignment of PAWS to Standards

Additional references for teachers are available under the PAWS link
http://edu.wyoming.gov/Programs/statewide_assessment_system/paws.aspx. One will
find documents showing alignment of PAWS to state standards.

These documents are designed to allow teachers to become informed about what skills
will be assessed on the PAWS. Sample items are posted for grades 3-8 and 11. The
sample items include the grade level, the content standard being addressed by each
sample question, the cognitive skill level being addressed (using 1=low to 4=high
cognitive skill levels), the context of the problem, and the item format, for example,
multiple choice or short response.

District Assessment Systems: Body of Evidence (BOE)

Wyoming’s assessment system is different from that in other states in that the
Wyoming system uses multiple assessments, instead of a single high-stakes test, to
determine graduation requirements. Information about other assessments by which
students can demonstrate proficiency, besides through the PAWS tests, is available from
the WDE homepage via the “Programs” tab. Under this tab is also located information
about “District Assessments”

http://edu.wyoming.gov/Programs/district_assessment/boe.aspx. District assessments are
a key component of the multiple assessment systems employed in Wyoming. Each
district develops a Body of Evidence (BOE) Plan for each of the nine content fields, in
which the PAWS can be a component. The Wyoming Department of Education provides
to districts submission guidelines, recommendations of documents to collect, and sample
BOE plans for use during BOE plan development.

Statement of Problem

Assessment, as defined in the National Science Education Standards (National
Research Council, 1996), is “a systematic, multi-step process involving the collection and
interpretation of educational data” (p.76). Large-scale assessments, such as those
statewide testing programs often employed to meet NCLB requirements, often turn into
high-stakes testing operations (Popham, 2001). Large scale state assessments become
high stakes when referenced to whether 1) they are used to determine significant
consequences linked to individual students’ test performance, or 2) the students’ test
scores are used to evaluate the instructional success of a school or district (Popham,
2001).

For more than two decades there has been a call to consider alternative assessment
practices that benefit students directly. Several studies in the 1980s and 1990s indicate
standardized testing distorts curricula and instruction and leads to cheating, test pollution
and useless data (Parris, 1998; Falk, 2000). “In a perfect world, standardized assessment,
whether paper/pencil or performance-based, would be designed to address standards-
based content and process in ways that reflect varied ways of knowing” (Hammerman,
2005, p. 26). Yet since NCLB was enacted the time spent taking tests in American
classrooms has increased, with no significant increase in student performance on
measures such as the National Assessment of Educational Progress (NAEP) (Nichols &
Berliner, 2008). It can be said that testing provides only limited views of students’
proficiencies, or of the thinking strategies they use in their learning (Falk, 2000).

In 1994 Parris and Ayres provided a list of principles for assessment reforms that
support learner-centered assessment. The authors noted that the “emphasis is on ongoing
assessments because they promote students’ analysis and direction of their own learning”
(Parris and Ayres, 1994, p. 51). One strategy for promoting learner-centered assessment,
including student involvement in their own learning if designed carefully, is the use of
portfolios. Portfolio assessment is defined in Classroom Assessment and the National
Science Education Standards (NSES) as “A purposeful and representative collection of
student work that conveys a story of progress, achievement and/or effort” (NRC, 2001, p.
31). Portfolios are one method of alternative assessment that aligns nicely with the 12
principles of assessment reform called for by Parris and Ayres in 1994.

A key component to the Wyoming Assessment System is the use of multiple
assessments, instead of a single, high-stakes assessment (WDE, 2005). The BOE system
determines, through multiple assessments, whether Wyoming’s high school students have
met state graduation standards. The Wyoming Body of Evidence system is intended to
produce a collection of data similar to portfolios of student work gathered throughout
their high school years. The plans are designed and implemented at the local level, with
each district responsible for creating its own plan. Districts must submit a BOE plan for
nine content areas: math, language arts, science, social studies, health, physical education, fine & performing arts, foreign language, and career & vocational. To assist districts, the Wyoming Department of Education has developed a series of recommendations and guidelines for districts to follow. In a district’s BOE assessment plan the PAWS-scores can be a component.

Just as teachers vary tremendously in the ways they teach students, districts vary greatly in the ways they write their BOE plans. There is no standard format. However, to ensure that the efforts of all districts are aligned with Wyoming standards, the BOE plans are subject to a peer review.

**BOE Peer Review.** The BOE peer review process is the evaluation of BOE plans based on the concept that a larger and more diverse group of people will be able to make a more impartial evaluation than will just the person or group responsible for creating the plan. All districts convene at one meeting location during the BOE peer review event. The review teams consist of three to five reviewers and a team leader. The peer review panel is made up of educators and administrators from across the state. Reviewers and team leaders are selected from a pool of experienced educators familiar with the BOE system. Reviewers have participated in a half-day training and calibration session prior to conducting reviews of district BOE systems. Each district submits two content area plans for the BOE peer review, as opposed to random selection of the plans to be reviewed from the nine BOE plans developed by each district. Each district’s plan is reviewed independently by two teams. For each BOE system, teams look for evidence that the design criteria of alignment, consistency, fairness, standard-setting, and
comparability have been satisfactorily met and that the system has been implemented. The rubrics by which each BOE is scored are located in Appendix B. Each district is provided with a report that includes suggestions for improving its system (Wyoming Department of Education, 2008b). If the BOE system is found to be satisfactory, no further review is required and no change in accreditation status is recommended to the State Board of Education. If a district’s BOE system is found to be unsatisfactory, the district will be required to participate in the next BOE Peer Review, a letter will be sent to the district Superintendent specifying which criteria have been satisfactorily met, which have not been satisfactorily met, and suggestions for what steps might be taken to address those criteria that have not been met. At the follow-up peer review meeting, only the district plans that were found to be unsatisfactory at the first meeting are required to resubmit for additional review. In the case that a district plan was found unsatisfactory at the second meeting, a recommendation to reduce the accreditation status to “Accreditation with follow-up” may be made to the State Board of Education.

In light of the BOE use as an alternative to high stakes testing, there is a need to explore the patterns of development and implementation of BOE science systems throughout the state of Wyoming.

Research Questions

The overarching purpose of this qualitative study is to explore the patterns of development and implementation of Body of Evidence (BOE) science systems throughout the state of Wyoming, using an emerging and relatively open mixed methods
design. BOEs were first launched throughout Wyoming a decade ago, and are ongoing today. Through interviews with teachers and curriculum coordinators and analysis of BOE science plans, the following research questions will be explored:

1. What design elements and implementation processes characterize BOE science assessment plans across Wyoming school districts?
   a. What are the characteristics of the BOE systems being implemented today, within and across districts?
   b. How are the design principles presented in the Wyoming Assessment Handbook integrated into district BOE science plans?
   c. What processes do different districts utilize to ensure that assessments are coordinated with state science standards and benchmarks?

2. What support do teachers receive to develop a BOE science plan, to align the plan to the state science education standards, and to implement the intended science plan?
   a. What support do teachers receive while developing, aligning and implementing district BOE science plans, and how does this vary among districts?
   b. Is there a relationship between the extent or nature of the activities within the BOE science plan and the level of support teachers receive to develop, align, and implement the intended plan?
c. Is the support teachers perceive they receive in a given district comparable to the support curriculum directors perceive they provide in developing and sustaining a district’s BOE plan?

3. The BOE approach was originally envisioned by the 2000 WY state legislature as a way to provide districts with a measure of decision-making autonomy, participation and ownership within the structure of state-mandated assessment. How do districts handle this opportunity to design and implement a locally useful assessment?

    a. What are the different ways districts utilize BOE data?
    b. How do districts disaggregate assessment results?
    c. In what ways do teachers, curriculum coordinators, and administrators see student learning taking place as a result of BOE assessment plans?

Definition of Key Terms

ALIGNMENT – provides assurances that 1) district assessments are aligned to the standards, 2) all standards are adequately assessed, 3) assessments are aligned to the level of cognitive complexity called for by the standards, and 4) students are evaluated against the standards. “The combination of assessments that comprise the system are aligned with district content and performance standards so that the full set of standards, both in terms of content and cognitive complexity, is assessed” (Wyoming Department of Education, Assessment Handbook, 2008, p. 38).
AYP - Adequate Yearly Progress - “Each State plan shall demonstrate, based on academic assessments described in paragraph (3), and in accordance with this paragraph, what constitutes adequate yearly progress of the State, and of all public elementary schools, secondary schools, and local educational agencies in the State, toward enabling all public elementary school and secondary school students to meet the State’s student academic achievement standards, while working toward the goal of narrowing the achievement gaps in the State, local educational agencies, and schools” (No Child Left Behind Act of 2001, 2002, 115 STAT. 1446).

BOE – Body of Evidence - The philosophy at the heart of the Wyoming Body of Evidence system is to provide multiple measures to assess student mastery of the content standards; in this way, no single assessment can disqualify a student from graduation (Wyoming Department of Education, 2008b). “The Body of Evidence is a collection of assessments in a district system used to determine student performance levels in each of the nine content areas in grades 9-12” (Wyoming Department of Education, 2008c, p. 1).

COMPARABILITY – provides assurances that 1) teachers use comparable assessments across schools and classrooms within the district, and 2) assessment results are comparable from year to year. “The assessments comprising the system should be comparable across schools and classrooms within the same school district both within a given year and across years. The comparability requirement is, in some sense, a specific instance of the fairness criterion. If students’ difficulty in meeting the graduation requirements is associated with the particular school attended or the year they graduated,
the system is not fair. The district should work to avoid changes in the difficulty of the assessments over time due to the ‘memorability’ of specific assessments or changes in the way that scorers rate students’ papers or projects” (Wyoming Department of Education, 2008b, p. 41).

**CONSISTENCY** – provides assurances that 1) assessments yield consistent decisions about student performance related to the standards, 2) teachers employ similar techniques when scoring assessments, 3) there is commonality of assessment procedures and scoring across the district, and 4) teachers will avoid the use of compliance issues when making decisions about student performance. “The decision regarding whether or not a student has met the graduation requirements for a given content area must demonstrate a high degree of consistency, such that the misclassification rates are minimal. The focus of this evaluation should be concentrated on the system and should examine, for example, how different judges would evaluate the same set of data about a group of potential graduates. In order to satisfy this criterion, the district should also document that the results of the assessments are not overly influenced by error due to raters or the specific tasks/items used comprising the assessments. Individual assessments within the system need to be evaluated for consistency, in terms of error due to raters, tasks, administration conditions, and occasions” (Wyoming Department of Education, 2008b, p. 39).

**FAIRNESS** – a “fair test is one that yields comparably valid inferences from person to person and group to group” (National Research Council, 2001, p. 214).
INSTRUCTIONALLY SUPPORTIVE ASSESSMENT – assessment intended to promote more effective classroom instruction (Popham, 2003).

NCLB – No Child Left Behind, Public Law 107-110, known as the No Child Left Behind Act of 2001.

PAWS – Proficiency Assessments for Wyoming Students (PAWS) is the assessment system developed by the Wyoming Department of Education in cooperation with Harcourt Assessment, Incorporated (Wyoming Department of Education, 2008a).

STANDARD SETTING – provides assurances that 1) the district has employed a deliberative process in setting passing scores, 2) there is an agreed upon description about what it means for a student to pass or to be proficient, and 3) teachers assign passing scores tied closely to definitions of proficiency. “The method for establishing cut scores between various performance levels on the Body of Evidence should be based on a defensible methodology, and the district should indicate a clear rationale for choosing one method over another. The method selected should incorporate clear descriptions of the performance levels and not be based on arbitrary performance distinctions” (e.g., traditional percentages) (Wyoming Department of Education, 2008b, p. 40).

VALIDITY – the “extent to which the test or assessment task is, in fact, credible for making the types of conclusions or inferences desired” (Vitale, Romance, & Dolan, 2006, p. 2).
Significance of Study

Calls for reforms in educational assessment have been ongoing for several years. Pellegrino (1999) argues that “assessment that is integral to the process of learning and teaching can impact achievement significantly” (p. 5). Additionally, he offers insight into how the educational assessment community can substantially improve assessment design and implementation to meet the challenges of integrating assessment into the learning environment. One recommendation is that assessments need to be more valid and complex, as well as embedded into the fabric of instruction.

Nebraska, like Wyoming, employs a non-traditional assessment program, School-based Teacher-led Assessment and Reporting System (STARS), whereby districts develop their own assessment system or in collaboration with other districts (Gallagher, 2007). This system has six quality criteria overarching the commitment to using multiple measures of student performance, making it very similar to the BOE criteria. The uniqueness of Wyoming’s two-tiered assessment system, one embedded in instruction (the BOE) and one external (the PAWS), makes analysis of the Wyoming BOE system valuable to the educational research community as an alternative to high-stakes testing.

Potential Limitations of Study

The Wyoming BOE-plans have been in place for eleven years, yet this is still a relatively young system, and data from this system must be considered preliminary. While the five components of each BOE plan are all technical features of assessments
emerging from traditional psychometrics, more sensitive tools are needed to examine and document the quality of classroom assessment.

Wyoming is small in population, supporting fewer students in the whole state than populations many large city districts serve. Thus the support in number of administrators and support staff is greater, and the student-to-teacher ratio is generally lower than in many areas of the United States. Opportunities for BOE training and collaboration have been a priority in the staff development structure originating within the Wyoming Department of Education. Without the same support system, similar results cannot be expected in other states.

Haertel (1999) points out that some research shows indications of ways in which tested proficiency fails to generalize. “If year-to-year score gains do not even generalize to similar tests, it is hard to imagine that such changes tell us much about proficiency across the rest of the school curriculum, let alone the set of academic knowledge in real-world contexts” (p.8).

Inferences from test scores to quality of schooling are problematic, and this study will not consider the great depth of contextual information that would be needed to make any such inferences about the quality of schooling among districts.

Generalizability of the study is a potential limitation of the study. Merriam (1998) and Bogdan and Biklen (2006) agree that the case study method limits generalizability. It is difficult to generalize beyond the specific research subjects and the time and place of the study. It is not possible to predict with absolute certainty from this study what educators in other states would do. Yet educators working in settings, or with assessment
systems, that have similarities to those described here may find these results useful and relevant.
CHAPTER 2

REVIEW OF LITERATURE

Introduction

The first section of this chapter provides an overview of learning theories from recent cognitive science and explains how these perspectives can inform our thinking about student assessment. The second part examines components of learner-centered assessments. Section three reviews literature relating learning theories to assessment practices. The final part concludes with dual discussions of the designs of the Proficiency Assessment of Wyoming Students (PAWS) in science and the Wyoming Body of Evidence (BOE) Assessment Plan.

Topic 1: Perspectives from Cognitive Science

The American way of testing pupils’ attainments has been around since the Committee of Ten published their famous Report on Secondary School Studies in 1892. Although the Committee promoted that the same curriculum should be offered to all students, regardless of their destinies in life, members probably also presumed that high schools would continue to serve only a very small proportion of the population. This fact has changed the landscape of schools over the last 100 years. Along with changing demographics of high schools, cognitive sciences developed at the same time. Cognitive science, the science of mind, explores the mechanisms by which people acquire, process, and use knowledge. “A science that focuses on knowledge acquisition and use, by
definition, has implications for education,” (McGilly, 1994, p. 3). The cognitive perspective emphasizes that knowledge is actively constructed by learners. Human constructivists believe we seek meaning in interactions with objects, events and other people (Mintzes, 2000). Constructivists believe science is best understood through extended periods of interaction with the natural world, including others who are making meaning of, or in, the same situations. How knowledge is structured and connected determines its availability for later use. In a review of lessons from cognitive science, vanGelder (2005) draws six key lessons for teachers: 1) critical thinking is a highly contrived activity, 2) mastering skills takes deliberate practice, 3) teachers must teach for transfer and students must practice this skill, 4) students have to learn new vocabulary to master a body of knowledge, 5) critical thinking skills improve faster when instruction is based on mapping concepts, and 6) belief preservation must be purposefully overcome by weighing evidence to guide opinions.

Applications of cognitive science to meaningful science understanding reveal four common themes (Vitale, Romance, Dolan, 2006, p. 3). First, “curriculum structure and student curriculum mastery of that content area should be consistent with the conceptual organization and proficiency of experts in that field.” Second, “classroom instruction should be organized according to curricular structure and core concepts and concept relationships.” Third, continually relate “what students are asked to do in instruction and assessment to the core concepts and relationships that such activities represent.” Fourth, teachers should “recognize the distinction between declarative knowledge and procedural knowledge.”
An important goal for the standards movement is to promote the learning of academic subjects with a deeper understanding of the underlying concepts and their applications than has been the case in the past. This requires a different type of teaching. In the past few years, more research has been done on constructivist learning (Brooks & Brooks, 1999), authentic performance assessments (Wiggins, 1993, 1998), standards-based education (Resnick & Resnick, 1992), and conceptual versus procedural learning (Driscoll, 2000). Assessment in support of standards must not only measure results, but must also contribute to the educational process itself. Assessment is a way of showing what a student knows and can do with that knowledge in an authentic setting. Wiggins (1998) advocates that authentic performances should be based on complex, problem-solving situations that students may encounter in their lives; that they should have clear targets and that those targets should be shared with students; that authentic performances should have an aspect of self-assessment to them to allow students to grow.

Evidence is summarized by McCombs and Marzano (1990), highlighting the importance of self-initiation, development, and continuation of self-regulated learning. “Although a skill component can enhance self-regulation, it is not sufficient. Students’ will or desire to engage in self-regulation is not only necessary, but primary,” (McCombs and Marzano, 1990, p. 51). The concept of metacognition plays a vital role in understanding oneself. Educative interventions must link self and metacognitive awareness to help students display the will and develop the skills for self-regulated learning. Utilizing a real-time information processing model, one can explain an individual’s variance in performance across homogeneous tasks. An example is the
observation that a student can perform quite well one moment and then poorly on the next opportunity (McCombs and Marzano, 1990, p. 59). This processing model can be related to a testing situation. When the task (the test) is judged as a high-risk situation then low task motivation results in low levels of efficiency instead of engagement in the task. If the environment threatens self-belief and evaluation systems, then metacognition and cognitive skills are not reinforced. “In effect, norm-referenced testing and normative grading practices negatively impact metacognitive and cognitive skill development for students who characteristically do poorly on tests and grades,” (McCombs and Marzano, 1990, p. 68). This serves as an argument against testing individuals repeatedly year after year. If students are armed with the necessary cognitive and metacognitive abilities they can experience success at tasks they once performed poorly. This is one argument for multiple opportunities in high-stakes testing situations. However, this style of self-regulation is usually seen to emerge in late adolescence and adulthood, which may be too late for accountability in the public education arena. “Research indicates that the most accurate way to evaluate students’ proficiency is through multiple assessments,” (Mid-Continent Research for Education and Learning, 2001, p. 1) but not more of the same; instead, choose assessments which engage students in self-evaluation to develop critical and creative thinking. Table 1 summarizes the cognitive perspectives and authors supporting learner-centered assessment practices.

Much of what we learn is acquired through interactions with others. In situated cognition theory (situated learning), declarative knowledge (knowing facts) and procedural knowledge (knowing how to do something) are integrated within a single
framework (Driscoll, 2000). The situative perspective, as developed by Driscoll, links knowledge from a person’s participation in communities of practice. “Proponents of situated learning argue that knowledge remains inert and unused if taught in contexts that separate knowing from doing,” (Driscoll, 2000, p. 155). The key components of the situated learning model, cognitive apprenticeship, collaboration, reflection, coaching, multiple practice, and articulation of learning skills change the culture of classrooms. Among critical aspects for the development of skills, practice and timely feedback are of utmost importance (National Research Council, 2001, p. 91).

With these changes in classroom culture, what products serve as valid evidence of students’ learning? McLellan (1993) recommended that a model be adopted as an approach to assessing situated learning, with three parts providing three different kinds of assessment measures: 1) diagnosis, 2) summary statistics, and 3) portfolios. Diagnosis is continually analyzing the progress of learners. Summary statistics keep records of student performance over time in order to show trends and patterns in performance, and portfolios consist of products that reflect the processes of learning over time. Paris (1998) recommends portfolios should contain three kinds of evidence about children’s learning and development, reflecting knowledge, skill and attitude. In 1994 McLellan added that technology is also a consideration in situated learning because “technology expands the power and flexibility of the resources that can be deployed to support the various components of situated learning” (p. 8). Studies of learning in social settings show learners take on proficiencies they see in others to improve their own performances (National Research Council, 2001, p. 89).
Haertel (1999) puts forth that from both the cognitive and situative perspectives, there still is not strong evidence that proficiency demonstrated in testing contexts will transfer to proficiency in other, perhaps more meaningful, contexts. “What research we have suggests surprisingly weak correlations between success on school achievement tests and success on tasks performed in nonacademic settings” (Haertel, 1999, p. 8). In a study of experimental design, Chen and Klahr (1999) also concluded that transfer tasks were difficult for almost all of the children in the study. Table 2 compares skills essential from the cognitive and situated learning perspectives.

Table 1. Cognitive Perspectives Supporting Learner-Centered Assessment.

<table>
<thead>
<tr>
<th>Cognitive Perspectives</th>
<th>Supporting Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Constructivism:</strong></td>
<td></td>
</tr>
<tr>
<td>Vocabulary acquisition</td>
<td>vanGelder–2005; Mintzes–2000</td>
</tr>
<tr>
<td>Curriculum alignment within field</td>
<td>Vitale–2006</td>
</tr>
<tr>
<td>Instruction &amp; assessment alignment</td>
<td>Vitale–2006; McCombs &amp; Marzano–1990</td>
</tr>
<tr>
<td>Declarative vs procedural knowledge</td>
<td>Vitale–2006; Champagne–2008</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Situated Learning:</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Declarative &amp; procedural knowledge integrated</td>
<td>Driscoll–2000</td>
</tr>
<tr>
<td>Communities of Practice</td>
<td>Driscoll–2000; Brown–1998; NRC–2000</td>
</tr>
</tbody>
</table>
Table 1 – continued

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Reflection</td>
<td>Driscoll–2000</td>
</tr>
<tr>
<td>Coaching</td>
<td>Driscoll–2000; NRC–2000</td>
</tr>
<tr>
<td>Multiple Practice</td>
<td>Driscoll–2000</td>
</tr>
</tbody>
</table>

Table 2. Comparison of Learning Perspective Skills.

<table>
<thead>
<tr>
<th>Active Construction</th>
<th>Shared</th>
<th>Situated Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transfer of skills</td>
<td>Practice</td>
<td>Collaboration</td>
</tr>
<tr>
<td>Vocabulary acquisition</td>
<td>Metacognition</td>
<td>Coaching</td>
</tr>
<tr>
<td>Concept strands</td>
<td></td>
<td>Technology</td>
</tr>
<tr>
<td>Misconceptions overcome</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curriculum alignment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instruction &amp; assessment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>alignment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Declarative vs procedural knowledge</td>
<td></td>
<td>Integrated declarative and procedural knowledge</td>
</tr>
</tbody>
</table>

In taking Baron’s advice (1998), a powerful way in which assessments can be learner-centered is by attending to what we know about how students learn. From cognitive psychology we know: 1) students learn best when they actively construct their own meaning (Driscoll, 2000, p. 187); 2) students learn best when they use metacognition and self-assessment (Bruner, 1994, p. 280); and 3) students learn best when they have experience in applying and transferring their knowledge (National Research Council, 2000). With respect to assessment, many important findings from detailed observations of children’s learning behavior have assessment implications. Children do not move
simply from erroneous to optimal solution strategies. Prior knowledge provides a foundation for future learning, acquiring skill takes time, individuals acquire skills more rapidly if they receive feedback, and research suggests that knowledge does not transfer very readily (National Research Council, 2001, chpt 3). “Assessments of academic achievement need to consider carefully the knowledge and skills required to understand and answer a question to solve a problem, including the context in which it is presented, and whether an assessment task or situation is functioning as a test of near, far, or zero transfer” (National Research Council, 2001, p. 92).

“Most current tests, and indeed many aspects of the science of educational measurement, have theoretical roots in the differential and behaviorist traditions. The more recent perspective – the cognitive and the situative – are not well reflected in traditional assessments but have influenced several recent innovations in the design and use of educational assessments” (National Research Council, 2001, p. 60). From motivational and educational psychology there are additional findings that can be incorporated into the design of effective assessments: 1) students learn best when problems are interesting, meaningful, challenging, and engaging (Resnick, 1987), 2) students learn best when they exercise choice, control, and personal responsibility (McCombs, 1986), 3) students learn best when they have a sense of efficacy (Driscoll, 2000, p. 311), and 4) students learn best when they work in groups (Tudge, 1990). Models of cognition and learning in science should be used to design and select assessment tasks that support the examination of a variety of kinds of student performances (National Research Council, 2001, p. 92). The Committee on Test Design
for K-12 Science Achievement recommends that states develop “a system of assessment that incorporates multiple measures and a range of assessment strategies” (National Research Council, 2006, p. 21). Table 3 summarizes practices of learner-centered assessment.

Table 3. Learner-Centered Assessment Practices.

<table>
<thead>
<tr>
<th>Practice</th>
<th>Supporting Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis</td>
<td>McLellan–1993</td>
</tr>
</tbody>
</table>

“A key component to the Wyoming Assessment System is the use of multiple assessments, instead of a single, high-stakes assessment” (Wyoming Department of Education, 2005). District assessments are a key component of the multiple assessment systems employed in Wyoming. Wyoming’s assessment system is different from other states in that it uses multiple assessments, instead of a single, high-stakes test, to determine graduation requirements. Each district compiles a Body of Evidence (BOE) Plan which represents a portfolio of student work throughout their high school years. Because research indicates learner-centered assessment is a key characteristic of a high-quality assessment system, this study will focus on the development of the Body of Evidence assessment system in six Wyoming school districts.

Pellegrino’s assessment of progress in educational assessment is that “we have made progress in connecting cognitive theory with assessments designed to support
learning and instruction at the classroom level. This effort, however, has involved little or no connection with formal psychometric criteria and practices or with large-scale assessment practices” (Pellegrino, 1999, p. 10). These connections can be seen depicted in Figure 2. Connections are made between cognition theory and research of classroom-based assessment practices in education, as evidenced in compilations by McMahon (2006) and Coffey, Douglas and Stearns (2008). Connections are also made between psychometrics theory and research to large-scale assessment in educational practices as evidenced by authors Wittrock (1991) and Popham (2005). However, the diagram depicts no cross-connections between cognition theory & research and large-scale assessment, or between psychometrics and classroom-based assessments. Mintzes (1998) also noted these areas need to be explored more by researchers. The National Research Council reinforced this message when the Committee on Test Design concluded “the design of assessment systems is not an exact science and has not been thoroughly researched” (National Research Council, 2006, p. 18).

Figure 2: Four Spheres of Work in Educational Assessment Practice (Pellegrino, 1999, p. 10)
The standards movement emerged as a means for improving public schools in the 1990s. By establishing higher standards for all students, tests are used to determine if students and schools are meeting the standards. “Assessment serves as an education tool with at least these two distinct functions: (a) improving learning and teaching and (b) providing measures of performance for accountability” (Paris, 1998, p. 189). Large-scale assessments are generally used to evaluate programs or to obtain information about student progress toward learning goals as defined in state standards. However, according to the Center on Education Policy, “19 states gave exit exams and withheld diplomas from students who did not pass” in 2005 and by “2012, 26 states – enrolling 72% of the nation’s public school students – will have mandatory exit exams” (p. 1). One can hope that states will embark upon designing effective assessment for such high stakes results.

Effective assessments should be based on three foundations, known as the assessment triangle: 1) grounded in a theory about how people learn, 2) elicits demonstrations of important knowledge and skills, 3) interprets evidence to draw meaningful inferences about what students know and can do (National Research Council, 2001). Effective assessment should provide feedback at many levels - to students, to teachers, to parents, to policymakers - about educational services. However, it is the conclusion of the Committee on the Foundations of Assessment, supported by the National Science Foundation, that one type of assessment cannot fit all of these needs (National Research Council, 2001). There are inevitable trade-offs in design of
classroom assessments for instruction guidance versus large-scale assessments for accountability purposes. It is the committee’s recommendation that an assessment system should be comprehensive, coherent and continuous. Comprehensiveness refers to a range of measurement approaches used to provide evidence to support decisions. Coherence refers to alignment of learning models across classrooms, schools, districts and state levels so that how students learn is useful to instruction and assessment. Continuous refers to how assessments should measure progress over time.

**Principles of Quality Assessments**

“Good assessment addresses the need for sound measurement through four principles: (1) developmental perspective, (2) a match between instruction and assessment, (3) the generating of quality evidence, and (4) management by instructors to allow appropriate feedback, feed forward and follow-up” (Wilson, 2009, p. 718). The *Atlas of Science Literacy* (2001) is designed to provide educators with a roadmap of developmentally appropriate benchmarks for particular science concepts. Principle two makes clear that the framework for assessment and the framework for curriculum must be aligned. Thus criterion-referenced tests are aligned to specified standards of what students should know and be able to do. Assessments must be quickly, readily, and reliably scored. This could be a combination of assessment blueprints and/or exemplars of student work at each performance level. In order to ensure changes in test scores are due to student achievement rather than to changes in what a test is measuring, procedures are needed to ensure comparability of results across time.
Design of Large-Scale Assessment – Prevailing Practices

Large-scale assessments are often used to evaluate programs and set expectations for student learning. They are often designed to sample a broad set of curriculum standards within a limited time frame. “A standardized achievement test is typically composed of many (usually dichotomous) items that are often all linked substantively in some way to a common construct variable” (National Research Council, 2001, p117). Criterion-referenced testing has become the norm, in order that student’s performance is reported in terms of what they can or cannot do in terms of performance against state standards in core content areas. Virtually every state has standards in place outlining what all students should know and be able to do in core subjects; yet as of 2009, nearly one-fifth of the states face penalties from the federal government for failure to meet standards that haven’t even been approved yet (NEA-Today, Aug/Sept 2009, p. 38). Item response modeling is one popular psychometric model, developed in the 1960’s, to enable comparisons among examinees who take different tests. This type of multiple-matrix design is used in the NAEP (National Assessment of Educational Progress) whereby not all students in the country are sampled (National Research Council, 2001). To avoid status models, which are single snapshots of student achievement, models of change and growth may appear to have a future in the arena of educational assessment.

Assessment design is influenced by 1) purpose of the assessment, 2) context in which it is administered, and 3) constraints such as time and money (National Research Council, 2001). Standardized tests can provide a look at how groups of students are performing within their local population, but do not effectively influence classroom
practice because they are limited at identifying underachieving students and do not provide timely information for classroom planning (Hulley & Dier, 2009). For the purpose of summative assessment of state standards, a model that focuses on the development of common constructs may suffice. Because large-scale assessments are typically only administered once a year, or every three years as is the case for science in Wyoming, the results seldom provide timely information for use by teachers in planning for individualized instruction. “Large-scale, standardized assessments can communicate across time and place, but by so constraining the content and timeliness of the message that they often have limited utility in the classroom,” (National Research Council, 2001, p. 8). There are problems that accompany almost all large-scale assessments. However, some “formal probability-based models for assessment were developed to overcome some of these limitations, especially for assessment purposes that (1) involve high stakes; (2) are not limited to a specific context, such as one classroom; or (3) do not require immediate information” (National Research Council, 2001, p. 112).

“The familiar practice of writing test items and handing them off to psychometricians to model the results cannot be sustained in complex assessments,” (National Research Council, 2001, p. 165). By studying results from two consecutive years (1999 & 2000) of administration of a Massachusetts 10th-grade science assessment, O’Neil, Sireci & Huff (2004) suggest cognitive skills were consistently measured; however, they were predominantly lower level skills being measured. Because examination is usually limited to answers on multiple-choice tests, the task of comparing
intended versus enacted constructs on large-scale assessments is demanding (Gorin, 2006).

In a 2003 study done by Boston College for the National Board on Educational Testing and Public Policy, teachers - with various levels if institutional and student sanctions regarding their state tests - were surveyed. Two areas addressed were alignment of classroom practices with the state test and impact on the content and mode of instruction. Results from teachers in low-stakes states found that teaching the state standards resulted in better test performance. More teachers in states with high-stakes tests felt their own tests reflected the format of the state test than did teachers from low-stakes test states. Across all types of testing programs, teachers reported that testing had influenced the time spent using a variety of instructional methods such as whole-group instruction, individual seatwork, cooperative learning, and using problems similar to those on the test. The survey contained seven items that focused on teaching practice. Across all seven items, teachers agreed that the state test influenced their use of specific pedagogical practices and instructional emphasis. (Pedulla et al., 2003)

Results of a meta-analysis by Cimbricz (2002) contend that state-mandated testing influenced teachers and their work negatively. Consequences reported included: (a) narrowing of the curriculum and instruction; (b) fostering anxiety, confusion, fear, shame, anger, and/or mistrust; (c) deskilling of teachers an/or perceptions of powerlessness; (d) loss of instruction time due to test preparation and testing (Cimbricz, 2002). The results of these tests have little to say to an individual teacher about how to improve classroom instruction.
Because state standards are generally written with too many curricular aims to meet in any given school year, this typically translates into sampling a collection of content standards. Educational leaders often foster the image that state assessments measure the complete array of content standards, a case that is generally not true. “Students will have studied content that isn’t assessed, and students will not have studied content that is assessed” (Popham, 2005a, p. 79).

**Performance-Based Assessment**

In a survey of 2000 students, in grades 2 through 11 in four states, Paris & colleagues found three developmental trends: 1) growing suspicion about the validity and usefulness of the tests, 2) decreasing motivation and effort, and 3) increasing incidence of counterproductive test-taking strategies. “Even criterion-referenced tests produce a large number of participants who score below others because the cutoff scores are established by political motives to show the relatively good or poor performance of the test takers,” (Paris & Ayres, 1994, p. 38). They recommend ongoing assessments to promote students’ analysis and direction of their own learning and to provide a list of twelve principles, corollaries to the APA Presidential Task Force’s Learner-Centered Psychological Principles, for assessment reform. In an effort to find alternatives to traditional assessment tests, states such as Connecticut and Wyoming are looking toward approaches that include more open-ended tasks. Performance assessments, an enduring concept which dates back to the 1950’s, represent one such effort. Performance assessments require students to perform tasks that involve the application of combined knowledge and skills in the context of a project. “If multiple assessments grow out of a
shared knowledge base about cognition and learning in the domain, they can provide valuable multiple perspectives on student achievement while supporting a core set of learning goals” (National Research Council, 2001, p. 225). The main purpose of performance assessments is to more closely align teaching with assessment practices in a manner that embeds learning and instruction in a reflective cycle (Mintzes, Wandersee, & Novak, 2000). The term *embedded* means that opportunities to assess student progress and performance are integrated into the instructional materials, and assessments are virtually indistinguishable from the day-to-day classroom activities (Wilson & Sloane, 2000). However, analysis of science tasks by Baxter and Glaser (1998) found mismatches between the intentions and the tasks and the associated scoring rubrics.

**Learner-Centered Assessments**

Are learner-centered and large-scale assessments compatible? Research by Davies (2007) and best practices tell us what some teachers have known for a while. “When students are involved in the classroom assessment process, they are more engaged and motivated, and they learn more. This deep student involvement is essential if students are to learn and to achieve at high levels,” (Davies, 2007, p. 31). Usage of learner-centered assessments, often referred to as performance assessments or authentic assessments, can be traced back as far as a 20-year study initiated in 1973 at Alverno College (Mentkowski, 1998). As performance assessments became increasingly popular in the 1990’s, questions about implementation and validation arose. Messick (1994) points out there is a continuum in response structures, ranging between multiple-choice items and open-ended performance tasks, with respect to what is captured by the score on the task
and the interpretation of the score. Gardner (1993) believes that a good assessment instrument can be a learning experience. Also in 1994, Paris and Ayres published 12 principles of learner-centered assessment which provided guidelines for assessment reforms that support use of learner-centered, large-scale assessment systems.

Learner-centered assessments share three characteristics. 1) They represent content, skills and dispositions that society currently values. 2) They function as learning events. 3) They encourage students to self-assess their work and progress using published performance criteria. (Baron, 1998, p. 212) Sewell et al (1998) purport that “differences in achievement attributable to factors related to learning styles, socio-cultural experiences, racial and ethnic distinctiveness, and levels of ability necessarily require an assessment system diverse in conceptualization, ideology, methodology, and purpose” (p. 323).

Rarely used for large-scale assessment purposes, these assessment approaches embedded in the classroom curriculum offer additional benefits over on-demand testing situations. They can provide opportunity for students to manipulate information to derive conclusions about science phenomena and can promote self-regulated learning, a skill that testing cannot foster.

**Instructionally Supportive Assessments**

A major part of the initial motivation for alternative assessment was to create a better match between desired instructional goals and actual assessment practices (Wilson & Sloane, 2000). There is a call for assessment to be part of the teaching and learning process. “Knowledge of children’s learning and the development of expertise clearly
indicates that assessment practices should focus on making students’ thinking visible to themselves and others by drawing out their current understanding so that instructional strategies can be selected to support an appropriate course for future learning” (National Research Council, 2001, p. 90). Analysis by Popham (2003, 2005, 2008) describes how to foster and create instructionally supportive assessments, that is, “assessments intended to promote more effective classroom instruction” (Popham, 2003, p. 1). From Popham’s viewpoint, an appropriate accountability test is one that can be used to secure accurate evidence about the instructional success of schools, districts, and states. In contrast, a test that spurs improved classroom instruction is a test that lets teachers know what skills and knowledge they need to promote in their classrooms and helps teachers determine whether their teaching was successful. Can a state test meet guidelines called for in No Child Left Behind legislation (NCLB) and assess adequate yearly progress (AYP)? In 2001, five national education organizations (American Association of School Administrators, the National Association of Elementary School Principals, the National Association of Secondary School Principals, the National Education Association, and the National Middle School Association) felt the need to commission an advisory group, the Commission on Instructionally Supportive Assessment, to provide key recommendations and to identify a series of requirements that large-scale accountability tests must satisfy if they are to supply accurate accountability evidence yet also support improved instruction. Three of the most important requirements are: 1) the task assesses important, and teachable, skills and/or knowledge, 2) the task provides assessment descriptions, and 3) the task provides standard-by-standard reporting of test results (Popham, 2005b).
Instructionally supportive accountability tests narrow the focus to a modest number of standards – and the skills and knowledge being assessed must be clearly important (Popham, 2003). Most states have a sprawling array of content standards which could never be assessed during a two-hour testing window with anything but an instructionally insensitive test. Instructionally supportive tests must therefore, focus on an important body of knowledge that must be effectively teachable. Proponents of standardized testing point out that one utility in assessments designed to measure state curriculum frameworks is that they ensure that what is taught is important (Sireci, 2008).

In order for a test to be instructionally supportive, teachers must be able to understand what is going to be measured. Assessment descriptions for each skill or body of knowledge must be provided so that teachers do not have to try to figure out what is being tested. Teachers need to know the curricular outcomes they are held accountable for addressing. In 1990 the California state legislature created the California Assessment Collaborative to study and support the development and implementation of performance assessments (Jamentz, 1994). One key practice uncovered by that project was that improvement comes from having teachers debate the meaningfulness of standards and tasks, and articulate the standards and assessment design process.

Finally, if teachers are to be in a position to provide effective instruction on tested skills and knowledge, they must be provided with their students’ test results on a standard-by-standard basis, and in a timely manner, in order to provide instructional support to students whose mastery of certain standards is unsatisfactory. Realistically, even if curricular-aim reports are not delivered in time for individual student remediation,
teachers can still ascertain the effectiveness of their current year’s teaching per standard and then can make changes in their next year’s lesson plans for certain segments of instruction (Popham, 2008, p. 134).

Topic III: Literature that Relates IV (Assessment Design) to DV (Wyoming’s BOE Development)

Wittrock (1991) summarizes that testing performs several useful, but different roles in education. Tests designed for diagnostic purposes would not be useful for comparing achievements of different students, schools, systems, states, or nations. The variety of purposes should not be confused. Assessments play a pivotal role in standards-led reform, by communicating the goals that school systems, teachers, and students are expected to achieve; providing targets for teaching and learning, and shaping the performance of educators and students (Linn & Herman, 1997).

The National Research Council’s Committee on test design for K-12 science achievement agreed there are four primary purposes for assessment: 1) application of 2001 recommendations of the Commission on Instructionally Supportive Assessments, 2) instructionally useful, teacher-led classroom-based program to meet the NCLB requirements, 3) multilevel state collaboration using technology, and 4) assessment to meet NCLB requirements in the context of psychometric considerations. The committee also agreed that none of these strategies could provide results that are valid for the purposes of the other (National Research Council, 2006). “The committee therefore recommended that states develop a system of science assessment that collectively would meet the various purposes of NCLB and provide education decision makers with
assessment-based information that is appropriate for each specific purpose for which it will be used” (Bertenthal, Wilson, Beatty, and Keller, 2008, p. 309). The committee recommends a developmental approach to science assessment, which is the use of multiple assessments providing evidence of student learning over time.

Subject-matter competence can be assessed by developing indices for proficient performance for various domains and skills. Because of cognitive science, we know that at specific stages of learning there exist differences in integration of knowledge, in forms of skills, and in the efficiency of performance. Glaser promotes that “test items can be designed to capture such evidence” (Glaser, 1991, p. 26). One roadblock to such design is the fact that much of what is known about the way knowledge is represented, organized, and processed in the mind has yet to be applied to the design of assessments for classroom or external evaluation purposes (National Research Council, 2001, chapter 3). Studies of ASVAB, TOEFL and GRE multiple choice items reveal sensitivity of score, meaning to relatively small changes in item structure, and suggest careful watch be given to unintended impact of item design (Gorin, 2006).

In 1999 educational testing specialist E.H. Haertel suggested possible studies that might address some testing assumptions that are often ignored. Haertel suggested the practical matter of wide-range inquiry into test scores often runs afoul due to short timelines and tight budgets, but infers that the real reason that such studies are not carried out is because results might be used to attack planned and ongoing testing programs. Thus he refers to his ideas as “an impractical quest for validity evidence” (p. 6). In contrast, Gallagher (2007) contends that “teachers and their classroom assessments are
causing the psychometric community to rethink its assumptions and practices” (p. 62). Gorin (2006) examines three aspects of test development - construct definition, validation, and item writing - that could be improved with the use of cognitive psychology principles. New ways to measure reliability of assessment driven by instruction and embedded in curriculum must be designed.

Starting in 2002, Reeves examined schools in Norfolk, VA, that experienced gains of 20 percent or more in academic achievement. The schools with the greatest gains were not similar demographically yet shared similar external variables of funding and labor agreements. What explained the extraordinary differences between the schools in Norfolk? “The keys to improved academic achievement were the professional practices of teachers and leaders, not the economic, ethnic, or linguistic characteristics of the students,” (Reeves, 2004, p. 66). From his work, Reeves identified nine characteristics that distinguished the schools with great academic gains: 1) time for teacher collaboration, 2) frequent feedback to students, 3) use of innovative schedules, 4) use of action research and midcourse corrections, 5) alignment of teacher assignments with teacher preparation, 6) analysis of student data from multiple sources, 7) analysis of common assessments, 8) respect for and utilization of every adult in the system, and 9) use and analysis of cross-disciplinary integration. Among those findings, six (1, 2, 4, 5, 6, 7) can be categorized as having influence on assessment. “The plain fact is that accountability for learning happens in the classroom,” (Reeves, 2004, p. 1).

NAEP reading committee member Peter Afflerbach (2008) purports that quality teaching and learning is not reflected in today’s assessment practices used throughout the
United States. There is an imbalance between attention given to high-stakes testing and classroom-based assessment. Afflerbach suggests that science assessment designers might benefit from challenges and accomplishments faced by other disciplines and avoid giving supreme attention to product assessments that “provide little information about the learning process, thus restricting the opportunity to influence instruction to meet students’ current needs” (p. 327). “Accountability is not achieved through testing – it is achieved through the hard work that surrounds successful classroom assessment and instruction” (p. 334).

Wyoming’s Framework for Assessing Skills

The Proficiency Assessments for Wyoming Students (PAWS) summative assessments are designed to meet the current NCLB testing requirements for language arts and mathematics in grades 3-8 and once in the grade span of 10-12. In March of 2007 the PAWS-science was field-tested in Wyoming schools. Students are tested for science in grades 4, 8, and 11. In April 2008 the first official PAWS-science assessment was administered and student results evaluated.

The Wyoming Department of Education (WDE) website (www.k12.wy.us) hosts the most extensive information about the PAWS. Finding specific information can be a bit tricky in the web-of-information. However, by locating the PAWS homepage (http://edu.wyoming.gov/Programs/statewide_assessment_system/paws.aspx), one can navigate numerous pdf files and word documents. These documents describe Skills, Cognitive Processes, Context and Evaluative Criteria for Mathematical Skills, Science Skills, Reading Skills, and Writing Skills. The documents also provide a rationale for the
PAWS-science assessment, “The Wyoming Science Content and Performance Standards ask students to demonstrate the knowledge, skills, and habits of mind necessary to safely perform scientific inquiry. To achieve this end, the PAWS assessment asks students to connect knowledge of scientific concepts with the problem at hand, design procedures to find an answer to the problem, investigate phenomena through data collection, and construct meaning through the use of evidence and logic. Therefore, Wyoming’s framework for assessing science is based on the skills of science as inquiry” (Wyoming Department of Education, 2005).

In the past, teachers have been faced with the perplexing task of trying to prepare their students to take the state test without knowing which standards/skills will or will not be assessed on a given form of the test. To ensure that PAWS is instructionally supportive, assessment descriptions have been developed to clearly and thoroughly describe the knowledge and skills that will be tested and the evaluative criteria that will be used to assess the skill.

History of the PAWS

In 2006, PAWS became the official statewide assessment used to measure individual student achievement against the Wyoming Content and Performance Standards, replacing the Wyoming Comprehensive Assessment System (WyCAS). This was as a result of recommendations by the Wyoming Statewide Task Force on Student Assessment & Education Accountability Report & Recommendations from October 2003. “Pursuant to State law, and based on six months of deliberation and outreach, the Task Force recommends modifications to Wyoming’s statewide assessment system and
the establishment of a single statewide accountability system with consequences” (Wyoming Department of Education, 2003b, p. 2). In this report, the Task Force recommends the establishment of a statewide assessment system that will include a summative assessment, which will maintain some, but not all, of the features of the previous state assessment, the WyCAS. The Task Force recommends modifications designed to improve teaching and learning, foster school improvement, and enhance Wyoming’s compliance with the NCLB law. The change in focus from WyCAS to PAWS reflects a shift in assessment philosophy. Attempting to re-write the statewide summative assessment, the WDE decided to follow guidance recommendations from the Commission on Instructionally Supportive Assessment.

PAWS was first produced by Harcourt Assessment, Inc. with input on item design, standard setting and scoring from Wyoming teachers, administrators and assessment expert, James Popham. “PAWS represents the nation’s first attempt to construct NCLB tests in complete accord with key recommendations of a national advisory group, the Commission on Instructionally Supportive Assessment” (Popham, 2005c, p. 1). Harcourt was later obtained by Pearson.

The science test was not able to be field-tested in 2006 because Wyoming teachers did not feel it was standards-based or that the questions were aligned correctly to cognitive levels nor with an inquiry focus. This was worked on for another year and field testing occurred in April 2007. The first classes to receive the PAWS-science for data interpretation occurred in April 2008. Data interpretation started in May 2008 and in August 2008, results were released to schools. An alignment study to Wyoming State
Science Standards occurred in July 2008, and the first data review occurred in July 2008. Teams of Wyoming teachers meet annually in July to review items for future field-testing. Working to clarify passages and prompts helps resolve biases and ambiguities in item design. Also used are statistical differential item functioning (DIF) analysis, which identifies items that produce differing results for members of a particular group after the groups have been matched in ability; for example, after being matched in ability girls may outperform boys on an individual item. This aids in determining if biases are inherent in an item and decisions made about whether items should be kept in the test or discarded. Progress variables (known in Wyoming as red, yellow and green lights) were derived in part from professional opinion about what constitutes higher and lower levels of performance and were informed by data about how students performed on individual test items. Although “it would be educationally useful to analyze the difficulty of an assessment task in terms of which students get it wrong, and why it is so problematic for those students” (National Research Council, 2001, p. 196), practical constraints (time and money) prevent this from occurring.

PAWS Test Design

The Commission on Instructionally Supportive Assessment (2001) presented nine requirements of statewide achievement tests for testing that both benefits students by providing educators with information they can use to improve instructions and to provide information for accountability purposes. The WDE has taken these recommendations to heart in the creation of the state assessment system, and in particular in creation of the PAWS tests.
Reading, writing, and math tests are administered to students in grades 3-8 and in grades 10 and 11. Science is administered once in grades 4, 8, and 11. The science test has three subtests: life science, physical science, and earth/space science. Wyoming’s framework for assessing science is based on the skills of science as inquiry; therefore, each subtest is designed to address a component of inquiry-based science. There are suggested time parameters for scheduling the PAWS into the school day, but there are no time requirements for completing each subset of the PAWS test. Students on Individual Education Plans (IEPs) take the PAWS in the special education room where the exam can be read to them.

Alignment of PAWS to Standards

Additional references are located at the PAWS homepage. One will find documents showing alignment of PAWS to state standards. Here is an example from science:

Grade 11 - Content Standard: Science as Inquiry
Benchmark: Pose problems and identify questions and concepts to design and conduct an investigation.
Skill: Use observation to pose questions that can be addressed through a scientific investigation.
Evaluative Criteria: The student recognizes or poses appropriate scientific questions or hypotheses for a scientific investigation.
- The student identifies relevant evidence in scientific contexts.
- The student makes comparisons between objects or events observed.
- The student interprets evidence of natural phenomena.
- The student demonstrates an understanding of scientific concepts.
- The student poses and evaluates arguments based on evidence. (Wyoming Department of Education, 2005)

These documents are designed to allow teachers a way to become informed about what skills will be assessed on the PAWS. Sample items are posted for grades 3-8 and 11. The sample items include the grade level, the content standard being addressed by
each sample question, the cognitive skill level being addressed (using 1=low to 4=high cognitive skill levels), the context of the problem, and the item type (mc = multiple choice, sr = short response).

**District Assessment Systems: Body of Evidence (BOE)**

In 1998, the Wyoming legislature established graduation requirements for students that included mastery of the common core of knowledge and common core of skills. The State Board of Education and the Wyoming Department of Education decided to require that districts collect a convincing array of evidence that could be used to determine whether or not a student had met the graduation requirements. School districts were to create an assessment system, called the Body of Evidence, which would yield the most valid inferences possible about a student’s performance on Wyoming Content and Performance Standards (Annette Bohling, WDE Standards Coordinator, letter to superintendents, October 28, 1998).

With the guidance of Scott Marion, assessment director in the Wyoming Department of Education, Wyoming avoided a single graduation test, a practice common to many other states. In line with recommendation #8 from the Committee on the Foundations of Assessment, large-scale assessments should sample the broad range of competencies and forms of student understanding that research shows are important aspects of student learning. District assessments are a key component of the multiple assessments system employed in Wyoming. Wyoming’s assessment system is different from other states in that it uses multiple, curriculum-embedded assessments, instead of a
single, high-stakes test, to determine graduation requirements. Each district compiles a Body of Evidence (BOE) Plan, in which the PAWS can be a component.

What is a body of evidence? In Wyoming, the Body of Evidence is an assessment system that is designed to determine whether or not students have met graduation standards and to provide a collection of evidence that can be used to support this decision. An assessment system is a well-articulated set of assessments, each of which contributes toward support inferences related to the identified purposes of the system. A well-designed system will support inferences where the whole is greater than the sum of the parts.

Recording of teacher judgments about students’ work is aided by scoring rubrics, unique to each progress variable that is used for assessing levels of student performance and interpreting student work. Four-point/task rubrics were developed for 23 science activities to be used as assessments through the Body of Evidence Consortium. These consortium activities were written by high school teachers across the state (with the assistance of nationally-recognized specialists from the Center for Assessment in New Hampshire) and selected by teachers in the school districts based on their “fit” with the objectives taught in each course. Each assessment is an authentic performance assessment that was to be seamlessly embedded into normal instruction. The goal was an assessment system so transparent to students that the test questions themselves would be reflective of quality instruction. In such a system, it wouldn’t be clear to students whether they were taking a test or receiving instruction. As defined by Messick (1994) a construct-centered approach was used to design the common science activities.
Knowledge and skills to be assessed were identified, which guided the construction of tasks and scoring procedures. These activities were first piloted, then student work samples were gathered, and a team of teachers evaluated the student work in order to write indicators of advanced, proficient, partially proficient and novice performance. Using these indicators, rubrics were designed for the assessments. The rubrics were field tested, and the teacher team gathered again to evaluate student work against the rubric for reliability. The rubrics are augmented with exemplary papers, samples of actual student work illustrating performance at each score level for all assessment tasks. After this two-year process the assessments mentioned above were placed into the statewide BOE Consortium Activities Bank. This is a lengthy process that cannot be feasibly done for every classroom assessment. Here follows a summary of the eight design principles, as described in the Wyoming Assessment Handbook, spring 2008 update. These eight design principles will be used as the basis for interviewing district personnel about BOE design and implementation experiences.

Alignment. “In terms of the relationship between standards and assessment, alignment refers to the match between the items on the assessments and the knowledge and skills represented by the curriculum and the standards,” (Wyoming Department of Education, 2008b, p. 14). In the handbook, it is recommended that districts align not only assessment content to state content standards, but also align cognitive depth-of-knowledge in terms of performance levels as established in the standards. This is referred to as a two-way alignment process. It is also stated that “alignment needs to occur at a finer grain level than simply matching test items to broad standards. This
alignment should occur at the benchmark level to ensure that the standards are appropriately sampled,” (Wyoming Department of Education, 2008b, p. 17).

**Consistency.** “The set of assessments should yield consistent decisions about students’ performance related to the standards. One of the intentions behind using the Body of Evidence approach was to reduce the number of misclassified students compared to using a single test” (Wyoming Department of Education, 2008b, p. 18). The authors of the handbook suggest districts can evaluate the consistency of their systems by several methods; by having different panels rate the same evidence, by using statistical analysis of single exam results, or by using one of the reliability procedures described in the texts recommended by authors of the handbook.

**Fairness.** In an effort to assure that assessments are not designed to be unfair to certain subgroups of students, authors of the handbook raise awareness to two issues. First, the set of assessments should lead to fair inferences about students’ performance on the standards. Fairness is subsumed by construct validity (Messick, 1994) however, there is also responsibility for rechecking it after data have been collected. Second, the results of assessments should be disaggregated for identified subgroups of students. “Examining patterns in the scores on the different assessments in the system might allow you to flag particularly problematic assessments,” (Wyoming Department of Education, 2008b, p. 21).

**Standard Setting.** “Standard setting is the process of determining the scores that divide various performance levels (i.e., cutscores) on an assessment” (Wyoming
Department of Education, 2008b, p.22). There are many methods for setting standards and establishing cut-scores. “It is beyond the scope of this handbook to provide a comprehensive description of standard setting methods; rather our intent is to simply introduce the topic” (Wyoming Department of Education, 2008b, p. 22). However, the authors then go on to require “the evidence of any process used should be documented.” The Wyoming Department of Education did offer standard setting workshops in 2000, 2002, 2003, and 2008 for districts which chose to receive help on compliance with this design principle, as well as made available a list of references for such processes.

Comparability. Graduation requirements mandated by Wyoming Statute 21-2-304 were written to ensure that all Wyoming students reached a certain threshold of competency prior to graduation. Although it is hard to measure comparability without a set of common assessments, the BOE design principles were written in such a way to allow comparisons of proficiency on the standards to be determined. Districts should ensure efforts, through common assessments, to compare outcomes across classrooms and schools, within and across years. “Within a given year, it is crucial that a student’s chance of meeting the standards is not dependent on which school within a district a student attends or which set of teachers within a given school a student draws” (Wyoming Department of Education, 2008b, p. 23). Authors of the Wyoming Assessment Handbook did not intend for sophisticated equating designs to become employed as a means for ensuring year-to-year comparability, however, this has become one task of assessment managers.
Multiple Measures. “Multiple measures should be used to increase the validity, alignment, consistency, and fairness of the assessment system,” (Wyoming Department of Education, 2008b, p. 24). Allowing students multiple opportunities to take the same test many times is discouraged. “The multiple measures should, whenever possible, represent different formats and strategies to enhance the fairness and alignment of the system,” (Wyoming Department of Education, 2008b, p. 24). Determining how many measures are enough to determine proficiency is a troublesome issue. Authors of the Assessment Handbook have left this issue up to individual districts to determine for their BOE system. However, data must be presented at the statewide BOE peer review to support judgments of student proficiency ratings.

Credibility. Confidence in an assessment system is built through public discussions and forums. This may be accomplished by inviting stakeholders (e.g., parents, businesspeople, district employees) to participate in assessment discussions and have input on assessment design and review. Showcasing student work to the public is another way of building credibility in a system. This can be accomplished through a venue of events, such as science fairs, art shows, history days, speaking events, student performances, and other public invitations to school. “If the public can see the type of work expected of students, it will likely build credibility in the system,” (Wyoming Department of Education, 2008b, p. 25).

Consequences. Readers may find the title of this design principle deceiving. As described in the Assessment Handbook, it is meant as an aspect of a data-collection tool,
or a district data collection system, not a reflection of punishments. “Each district should
collect and analyze data to evaluate the consequences of the system, both intended and
unintended” (Wyoming Department of Education, 2008b, p. 25). In order to investigate
the consequences of the implemented assessment system, districts should use data
collected to examine whether the systems needs to be modified and system revisions and
improvements made periodically. In order to show that students are not being denied
opportunities to learn the standards, data collection and analysis should show student
progress toward the graduation requirements.

Rewards and Sanctions. Wyoming state law (HB126) requires that the State
Superintendent and the State Task Force on Assessment and Accountability “establish a
transitional plan to provide a temporary system of rewards and sanctions for all public
schools and school districts” (Wyoming Department of Education, 2003b). The rewards
for schools and districts that meet AYP are minimal. Schools and districts that meet AYP
will be notified of this accomplishment, and may elect to receive additional recognition
from the Wyoming Department of Education such as a presentation at the local Board of
Trustees meeting. Furthermore, schools and districts that meet AYP are encouraged to
seek awards under Wyoming’s Innovative Trust Fund, which provides grants to promote
innovative educational initiatives that can improve student achievement and may receive
additional consideration for those awards. Most importantly, members of the Wyoming
Statewide Task Force on Student Assessment and Education Accountability believed that
implementing locally designed assessment systems had the potential of leading powerful
changes in instruction and learning in classrooms.
There are different levels of consequences for schools and districts that do not meet AYP, depending on whether it is for the first year, or successive years. The schools and districts that do not meet AYP for the first time will be expected to examine why they did not meet AYP and address those issues (at the school level) as part of their school improvement plans. Schools that do not meet AYP for the second consecutive year, based on the same academic indicator as in the previous year (math, reading/language arts, or attendance/graduation rate), must revise their school improvement plans to address the specific issues on which the given school missed AYP. As required under NCLB, Title I schools must also spend not less than 10 percent of their Title I funds on high-quality professional development that directly addresses the issues on which the given school missed AYP, and must offer students the choice to transfer to another public school in the district that is not in “improvement status.” Non-Title I schools must focus their professional development efforts to address the particular issues on which the school did not meet AYP (Wyoming Department of Education, 2010).

Summary

Issues such as equitable assessments and designs to capture the complexity of science reasoning and understanding are still grappled with by researchers. Many researchers have recommended guidelines toward development of instructionally supportive tests, yet it is not an exact science and has not been successfully implemented in large-scale assessment systems. In addition, there are challenges associated with wide-scale professional development related to assessment implementation, review and
analysis. A body of research knowledge implies that “assessment practices need to move beyond a focus on component skills and discrete bits of knowledge to encompass the more complex aspects of student achievement” (National Research Council, 2001, p. 3). The twelve recommendations for research, from the Committee on the Foundations of Assessment, need to be implemented. Because it is unlikely that high-stakes testing will end anytime soon, advocates for instructionally supportive assessments must advance efforts to reduce the unintended consequences and move teaching and assessment in directions that will improve student learning.
CHAPTER 3

RESEARCH METHODOLOGY

Introduction

This chapter outlines the design for the study and the settings in which it took place. The first section of this chapter provides an explanation of the research design that was used. The second part describes participating districts and personnel in the study and sampling techniques. Section three explains the initial design and piloting, followed by full-scale implementation of the survey instrument which served as an orientation study to discover how views differ among high school science teachers and other district administrators with regard to BOE development and implementation. Section four explains methods associated with the district case studies. The final section addresses the scope and potential limitations of the study, and measures to ensure the quality and trustworthiness of the research.

Context of the Study

In 2000, the Wyoming State Board of Education developed new graduation requirements. These rules, known as Chapter 31 Rules, establish the requirements for earning a high school diploma. It is these rules that present and define Body of Evidence (BOE). The BOE is intended to be a collection of evidence that demonstrates a student’s mastery of the state content and performance standards, instead of a single high-stakes test. Chapter 31 calls for the BOE to be designed at the district level and provides
guidance for districts to answer two questions: 1) whether the student knows enough to graduate, and 2) whether the evidence supports the answer to question one. Chapter 31 Rules require multiple assessment opportunities for students to demonstrate competency in the standards. Planning a BOE requires considerable commitment of time and other resources, as does the implementation of the BOE. In The Wyoming Assessment Handbook the State Department of Education provides guidelines for districts to develop BOE plans via several approaches. The goal of the BOE system is to support inferences about students’ competency and to ensure equality of educational opportunities. Regardless of the type of approach and assessments selected by a district, the BOE plan must be submitted and approved by a BOE peer-review committee in order to meet state accreditation requirements.

The total number of students enrolled in Wyoming’s 48 public school districts is approximately 87,000 with 83% of those students being white, 10% Hispanic, 3.5% Native American, and 2.7% from other ethnic, cultural or language groups (U.S. Department of Education, 2010). According to 2009 data from The Rural School and Community Trust, 24% of Wyoming’s public school students are enrolled in a rural school. However, rural schools account for 57% of the schools in the state, almost double the national average of rural schools. Instructional expenditures per rural pupil are $2500 more than the national rural average of $5100. With 14.6% of rural students qualifying for special education services, the rate in rural Wyoming is high. There is also high rural household mobility in Wyoming, perhaps attributed to the shifting energy markets. Although Wyoming students rank relatively high in rural proficiency in
reading, per NCLB, and are ranked 16th nationally, rural graduation rates are 34th at 82%.
However, the picture is bleak for the rural minority students who graduate at a rate of
only 28% (The Rural School and Community Trust, 2009).

Research Perspective

The framework for this study developed from a postpositive knowledge claim.

The problems studied by postpositivists reflect a need to examine causes
that influence outcomes. The knowledge that develops through a
postpositivist lens is based on careful observation and measurement of the
objective reality that exists “out there” in the world. Thus, developing
numeric measures of observations and studying the behavior of individuals
become paramount for a postpositivist. (Creswell, 2003, p. 7)

My observations included requesting BOE plans, surveying participants, and
interviewing participants.

A mixed-methods approach utilizing both qualitative and quantitative strategies was
employed. Sequential procedures seemed to make sense. “This may involve beginning
with a qualitative method for exploratory purposes and following up with a quantitative
method with a large sample so that the researcher can generalize results to a population”
(Creswell, 2003, p. 16). The advantages of the quantitative survey approach is that it is
possible to measure reactions of a great number of people to a limited number of
questions, thus facilitating comparison and aggregation of data (Patton, 2002). By
contrast, qualitative methods produce a wealth of detailed information about smaller
numbers of people, increasing the depth of understanding the situation (Patton, 2002).
The qualitative approach I began with is grounded theory work, whereby, in a written
letter to superintendents I requested access to their BOE plans for this study. I collected BOE plans from 17 districts across the state and categorized them based on their similarities and differences. These characteristics included being composed of all tests, being composed of all projects, or a variety of approaches between the two extremes. After identifying categories of BOE plans, and surveying school personnel, Wyoming school districts were categorized by size, minority enrollment, free and reduced lunch enrollment and PAWS test results in order to select case study participants to interview.

“Triangulation strengthens a study by combining methods” (Patton, 2002, p. 247). The basic type of triangulation applied in this study was methodological triangulation, the use of multiple methods to study a single problem. As described by Clark, Creswell, Green & Shope (2008), mixed methods research has three key features: collection of both quantitative and qualitative data, analysis of both datasets, and development of an overall interpretation from the data sets. There is an enhanced understanding that arises when more than one kind of information is used to address a research problem. Looking at a phenomenon from multiple perspectives represents a more complete understanding of the issue.

**Research Purpose and Focus Questions**

The overarching purpose of this mixed-methods study is to explore the patterns of development and implementation of BOE science systems throughout the state of Wyoming, using an emerging and relatively open design. Wyoming’s assessment system is different from other states in that it uses multiple assessments, instead of a single high-stakes test, to determine graduation requirements. BOEs were first launched throughout
Wyoming a decade ago, and are ongoing today. Through interviews with teachers and curriculum coordinators, as well as analysis of BOE science plans, the following research questions will be explored:

Research Question 1: What design elements and implementation processes characterize BOE science assessment plans across Wyoming school districts?

a. What are the characteristics of the BOE systems being implemented today within and across districts?

b. How are the design principles presented in the Wyoming Assessment Handbook integrated into district BOE science plans?

c. What processes do different districts utilize to ensure that assessments are coordinated with state science standards and benchmarks?

Research Question 2: What support do teachers receive to develop a BOE science plan, align the plan to the state science education standards, and implement the intended science plan?

a. What support do teachers receive while developing, aligning and implementing district BOE science plans, and how does this vary across districts?

b. Is there a relationship between the extent or nature of the activities within the BOE science plan and the level of support teachers receive to develop, align, and implement the intended plan?
Research Question 3: The BOE approach was originally envisioned by the 2000 WY state legislature as a way to provide districts with a measure of decision-making autonomy, as well as participation and ownership within the structure of state-mandated assessment. How do districts handle this opportunity to design and implement a locally useful assessment?

a. What are the different ways districts utilize BOE data?
b. How do districts disaggregate assessment results?
c. In what ways do teachers, curriculum coordinators, and administrators see student learning taking place as a result of BOE assessment plans?

Research Design

The overarching purpose of this mixed-methods study is to explore the patterns of development and implementation of BOE science systems throughout the state of Wyoming using an emerging and relatively open design. Wyoming’s assessment system is different from other states in that it uses multiple assessments, instead of a single high-stakes test, to determine graduation requirements and, adding to the uniqueness of the Wyoming system, these assessments are permitted to be locally selected or designed. Many questions arise regarding this system of interest and importance to the broader education community. For example, are the assessments being seamlessly incorporated
into instructional plans, or are they stand-alone assessment practices, and what does this
tell us about people’s perception of the system? To what extent are people’s
understandings of what the BOE science system is, and what it has to offer, comparable
from one district to the next? How do time and other resources spent on various aspects
of the BOE – from professional development and planning to implementation and teasing
out lessons learned – vary across the state? These and other questions led to my interest
in conducting this study.

This study used a mixed-method design as described in Creswell (2003) in order
to capture the perspectives of the participants in rich, descriptive detail. The use of mixed
methods of data collection enhanced the strength of this study and helped the researcher
to better understand the factors that drive BOE science assessment practices for the
participants, and according to Denzin and Lincoln (2005), capture as much of the reality
as possible. The mixed-method design chosen here is the triangulation design presented
by Creswell and Plano-Clark (2007). With this mixed-method design the quantitative and
qualitative data are collected, analyzed, and merged into one interpretation and are
usually given equal emphasis.

Additional methods of data collection used in this mixed methods study began
with an orienting survey, conducted in Spring 2010, to investigate characteristics of the
BOE systems being implemented within and across Wyoming school districts (see
Appendix A), followed by district case studies, implemented in Fall 2010, that included
an interview (see Appendix B) with the curriculum coordinator and a science teacher in
each district and a review of the BOE plan for that district. Both the orienting survey and
interviews that occurred later were utilized, in part, to understand the influence the State of Wyoming’s BOE mandate has had on participants’ approach toward assessment and on their students’ science learning. The survey, completed by 119 school personnel, provided initial information about these issues, and the case studies provided a means to probe more deeply. The case studies, including in-depth interviews with a curriculum coordinator and a high school science teacher from six different school districts, also included a review of each district’s BOE plan to gain additional insights.

Together, the survey and case studies provided a rich pool of quantitative and qualitative data. Table 4 provides examples of questions examined through the surveys, the source of quantitative data for this study, as well as through the case studies, the primary source of qualitative data, to address specific research questions.

Table 4: Overview of Quantitative and Qualitative Data Sources to Address the Research Questions.

<table>
<thead>
<tr>
<th>Research Question 1: What design elements and implementation processes characterize BOE science assessment plans across Wyoming school districts?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1a: What are the characteristics of the BOE systems being implemented today within and across districts?</strong></td>
</tr>
<tr>
<td><strong>Survey Data</strong></td>
</tr>
<tr>
<td><strong>Case Study Data</strong></td>
</tr>
</tbody>
</table>
Table 4 - continued

<table>
<thead>
<tr>
<th>1b: How are the design principles presented in the Wyoming Assessment Handbook integrated into district BOE science plans?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Case Study Data</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1c. What processes do different districts utilize to ensure that assessments are coordinated with state science standards and benchmarks?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Case Study Data</strong></td>
</tr>
</tbody>
</table>

**Research Question 2: What support do teachers receive to develop a BOE science plan, to align the plan to the state science education standards, and to implement the intended science plan?**

<table>
<thead>
<tr>
<th>2a: What support do teachers receive while developing, aligning and implementing district BOE science plans, and how does this vary across districts?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Case Study Data</strong></td>
</tr>
</tbody>
</table>
Table 4 - continued

<table>
<thead>
<tr>
<th>Research Question 3: The BOE approach was originally envisioned by the 2000 WY state legislature as a way to provide districts with a measure of decision-making autonomy, as well as participation and ownership within the structure of state-mandated assessment. How do districts handle this opportunity to design and implement a locally useful assessment?</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>2b: Is there a relationship between the extent or nature of the activities within the BOE science plan and the level of support teachers receive to develop, align, and implement the intended plan?</th>
</tr>
</thead>
</table>

| Case Study Data | Interview Question: Would you describe your BOE science assessment plan as district-developed, consortium activities, or a blend of both? Would you describe the nature of your BOE assessments as projects or tests? |

<table>
<thead>
<tr>
<th>2c: Is the support teachers perceive they receive in a given district connected to the support curriculum directors perceive they provide in developing and sustaining a district’s BOE plan?</th>
</tr>
</thead>
</table>

| Survey Data | Survey item #1.3: Have you participated in a BOE peer-review at the state level?  
Survey item #1.4: Have you participated in developing a BOE science assessment plan in your current district position?  
Survey item #3.1: In the last three years, approximately how many hours have you spent working on your BOE science assessment plan?  
Survey item #3.2a: As a professional development experience, how do you rate your experience developing the BOE plan in your district?  
Survey item #3.2.b: How do you rate your experience with the implementation of the BOE plan in your district? |

| Case Study Data | Interview Question: Describe the support you receive from your district administration to develop and/or to sustain your BOE assessment plan in science. Whom do you see this support coming from? In what areas do you feel most supported, developing or implementing assessments? In what ways would you like to receive more support?  
Interview Question (for Curriculum Coordinators): Describe the support you provide for teachers to develop and/or to sustain the BOE science assessment plan in your district. In which area do you feel you provide the most support to teachers, in developing or in sustaining the BOE plan? In what ways would you like to provide more support? |
Table 4 continued

<table>
<thead>
<tr>
<th>3a. What are the different ways districts utilize BOE data?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Case Study Data</strong></td>
</tr>
<tr>
<td>Teacher Interview Question: What are the different ways your district utilizes BOE data? To whom do you report your BOE data?</td>
</tr>
<tr>
<td>Curriculum Coordinator Interview Question: What are the different ways your district utilizes BOE data? To whom do teachers report their BOE data?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3b. How do districts disaggregate assessment results?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Case Study Data</strong></td>
</tr>
<tr>
<td>Interview Question: How does your district disaggregate BOE science assessment results? Are the results reported to students, to parents, or to both?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3c. In what ways do teachers, curriculum coordinators, and administrators see student learning taking place as a result of BOE assessment plans?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Survey Data</strong></td>
</tr>
<tr>
<td>Survey item #3.2c: What is your opinion of the impact on student learning taking place in your classroom as a result of the BOE assessment plan your district has adopted?</td>
</tr>
<tr>
<td>Survey item #3.3: In what ways do you see student learning taking place in your classroom, as a result of the assessment plan adopted by your district?</td>
</tr>
<tr>
<td><strong>Case Study Data</strong></td>
</tr>
<tr>
<td>Interview Question: In what ways do you see student learning taking place in your classroom, as a result of the assessment plan adopted by your district?</td>
</tr>
</tbody>
</table>

Detailed descriptions of the survey and case studies follow in the next sections.

Development, Implementation and Analysis of the Orienting Survey

Pilot Survey

The BOE survey was designed with input from Dr. Art Bangert, a faculty member specializing in educational statistics and quantitative research methods, Department of Education, Montana State University, and Dr. Alan Moore, formerly Director of the Standards and Assessment Unit at the State of Wyoming Department of Education. Both advisors, familiar with the goals of this study and the BOE program, contributed to the development of the survey. This expert feedback contributed to the content validity of the
instrument. The survey was intended to gather preliminary information about the BOE from secondary science teachers throughout Wyoming, as well as school and district level administrators prior to development of the interview protocols. In order to evaluate the design of BOE science plans, research about requirements for plan development was used to guide the question development. To facilitate the evaluation, survey items included questions about whether or not survey participants had participated in a BOE peer-review at the state level, whether or not survey participants had participated in developing a BOE science assessment plan in their current district, and whether or not survey participants were familiar with the science assessment plan matrix in their current district.

This survey consisted of three sections. Section I elicited demographic information, for example, the participant’s primary role (high school science teacher, district curriculum coordinator, high school administrator, district administrator, other), and years in his or her current position. Section II elicited information about knowledge of the BOE science assessment plan, and Section III elicited information about personal experiences and opinions developing and implementing the BOE plan. An open-ended response allowed participants to share additional comments about ways they see student learning taking place in classrooms as a result of the BOE assessment plan adopted by their district.

Once the 16 questions for the survey were selected, a pilot was undertaken with two science teachers, one principal, one curriculum coordinator and one district data coordinator. These preselected individuals were all familiar with Wyoming BOE
requirements. The survey took participants about ten minutes to complete. Pilot survey respondents were asked for comments regarding clarity of the survey questions. Modifications were made as needed for the final survey. Modification included minor wording changes for clarity and understandability. The final version of the survey instrument is provided in Appendix A.

**Survey Implementation**

All surveys were administered online using Survey Monkey. In mid-April, 2010, a request to participate in the survey was sent electronically to potential participants through Dan Stephan, Executive Director of the Wyoming Association of School Administrators and through the Wyoming Science Teachers Association e-list serve. The requests included a brief statement describing the research study and providing contact information for the researcher. Each invitation was accompanied with a link to participate in the survey. A follow-up email was sent from the researcher two weeks later, and once again in early June 2010. All responses were anonymous, with the exception of the participants who left an e-mail for follow-up participation in retaking the survey in order to gauge reliability. These participants were assigned a number to maintain confidentiality and only the researcher had access to names and assigned numbers.

How large was the pool of survey respondents contacted, and how many from this pool actually completed the survey? There are 48 school districts in Wyoming employing about 232 full-time equivalent high school science teachers (Susan Williams, Data Analyst, Wyoming Department of Education, personal communication, June 16, 2011). In addition, there are 48 public school district superintendents, 76 high school
principals and 48 district curriculum coordinators invited to participate in the online survey. Of the population of high school science teachers, it is unknown exactly how many were contacted due to the composition of the Wyoming Science Teachers Association e-blast, which was the only comprehensive list of contact information for Wyoming science teachers available. This list contains 597 points of contact, or 2.6 times the number of secondary science teachers in the state. Those included in the e-blast list range from K-16 teachers of science to business and industry contacts. One hundred and seventy-two school administrators were invited to take the survey. Of the invited population, 79 participants completed the survey during the first two weeks in April, and 30 more eventually completed the survey during the next two months before the survey closed in late June. Including the ten participants that participated in retaking the survey, 119 surveys were collected from 67 high school science teachers, ten curriculum coordinators, 11 high school administrators, 13 district administrators, and nine other district personnel. Those responding represent about 28% of the state’s high school science teachers, 21% of the curriculum coordinators, and 14% of the high school administrators. Although these numbers are relatively low, they provide a context for understanding the survey responses and degree to which the survey findings represent the views of other teachers or school administrators.

With 404 potential survey participants, 300 responses would be needed (ideally distributed in proportions to their overall numbers to include 142 teachers, 42 superintendents, 42 curriculum coordinators, and 63 principles) for results to produce a
95% confidence level with a 5-point confidence interval. Indeed the 119 responses received fell short of the 300 responses needed to attain a 95% confidence level.

Instrument Reliability

Survey items were developed by consulting the Wyoming Assessment Handbook, the Wyoming BOE science consortium activities, The Survey Research Handbook (Alreck & Settle, 2004), Interpreting Assessment Data (Christmann & Badgett, 2009) and Designing Professional Development for Teachers of Science and Mathematics (Loucks-Horsley, et.al., 2003). The survey instrument was edited by the director of the Standards and Assessment Unit at the Wyoming Department of Education. The survey was piloted by a curriculum coordinator, a principal, a BOE coordinator and two science teachers.

Individual survey responses were available to the survey designer with a survey monkey log-in and results were coded into SPSS statistical software for analysis. Evidence for reliability of the survey was obtained using test-retest and internal consistency reliability analysis. Ten participants agreed to complete the survey twice in order to calculate test-retest reliability of the survey instrument. Test-retest reliability over a 16 day time period was obtained for a subsample of ten self-selected participants, yielding a stability coefficient of 0.817 for the total scale scores (p<.01). Analysis of the test-retest survey results revealed a Test 1 mean of 35.9 and a Test 2 mean of 35.6, Test 2 responses were not used in the analysis of the survey data.

Cronbach’s Alpha is a measure of internal consistency used to help determine which survey items become operational and which items are rejected. When Cronbach’s Alpha > 0.85, it is assumed that only one theoretical construct is being measured
(Schmitt, 1996). On the survey developed, Cronbach’s Alpha for the Likert scale questions ranged from 0.773 to 0.792, indicating the questions have a moderate to high degree of internal consistency reliability (Huck, 2008).

Survey Data Analysis

Survey responses were printed out, categorical variables were hand-coded for incorporation into SPSS software, and descriptive statistics were generated for all survey items. Questions were cross-tabbed and results evaluated using Chi-Square, Phi and Cramer’s V to determine significance of relationships between demographic variables. Cross-tabbing investigates the research questions about what is the extent of understanding of BOE science plans among participants. This phase of the analysis looked at the degree to which responses for Likert-type questions varied according to the categories of respondent - teachers, principals, superintendents, or curriculum coordinators.

Finally, responses to the open-ended survey question were analyzed qualitatively to obtain information about the ways participants see science learning affected by the BOE plan adopted in their school district. Inductive analysis (Patton, 2002) was utilized to discover patterns and themes from the open-ended responses. Specifically, the open-ended responses were read and reflective notes were made alongside each response, then these data were coded within and across demographic groups participating in the survey including teachers, administrators, ranging from superintendents and principals to curriculum coordinators, and other district personnel.
Development, Implementation and Analysis of the Case Studies

The case studies provide qualitative data regarding the processes involved in planning, implementing and using the results of the BOE. A focus on process involves looking at how something happens rather than examining outcomes. “Qualitative inquiry is highly appropriate for studying process because the experience of process typically varies for different people so their experiences need to be captured in their own words” (Patton, 2002, p. 159). The quality of investigation depends on collecting data from participants who can inform the researcher about the design and implementation of BOE science plans, thus first year teachers and novice curriculum coordinators were eliminated from consideration for participation in the case studies. Johnson and Christensen (2008) broadly “define case study research simply as research that provides a detailed account and analysis of one or more cases” (p. 406). For the purpose of this study, interviews of curriculum coordinators and science teachers from the same district, and review of the BOE plan from each district, provided an appropriate methodology. Thus, this is classified as an instrumental case study, as the cases are used to learn about something more general; to understand how science BOE plans were developed and implemented across six Wyoming school districts.

Case Participants

For the interview portion of the study, school districts were sorted by district enrollment size, minority population, free and reduced lunch status, and 2-year PAWS test score trends. Purposeful sampling was used to select districts for the interviews.
Patton (2002) recommends purposeful sampling as a means to select information-rich cases depending on the study purpose and resources available. After purposefully selecting six schools based on the range of indicators mentioned, curriculum coordinators were e-mailed and asked to participate in a phone interview. At the end of the interview, lasting about an hour, curriculum coordinators were then asked to refer high school science teachers who implement the BOE assessment plans for the teacher interview component.

The six schools that participated in the interview component of the study represent a range of Wyoming school districts. These will be referred to as Cottonwood SD1, Cypress SD2, Larch SD3, Linden SD4, Willow SD5, and Tesota SD6. Two-year PAWS performance trends were identified by the ranking the 48 high schools in Wyoming based on the percent of students scoring proficient or advanced in science. If a school ranked in the “Top 10” for two years in a row they are identified as such. School’s that were not consistent in scores are indicated as “No trend.” It is not indicated whether their scores are improving or declining. These six case-study schools are very different in size, demographics, and 2-year PAWS performance trends, in science, as outlined in Table 5.

Table 5. School District Demographics.

<table>
<thead>
<tr>
<th>School</th>
<th>Approximate K-12 student enrollment</th>
<th>% minority</th>
<th>% free-reduced lunch</th>
<th>BOE Plan*</th>
<th>2-year PAWS science trend</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cottonwood SD1</td>
<td>13,000</td>
<td>27</td>
<td>37</td>
<td>Blended, Content rich – process open</td>
<td>No trend</td>
</tr>
<tr>
<td>School</td>
<td>Population</td>
<td>Year</td>
<td>Grade</td>
<td>Description of BOE Plan</td>
<td>Trend</td>
</tr>
<tr>
<td>------------</td>
<td>------------</td>
<td>------</td>
<td>-------</td>
<td>-------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Cypress SD2</td>
<td>12,000</td>
<td>12</td>
<td>34</td>
<td>Blended, Content rich – process open</td>
<td>No trend</td>
</tr>
<tr>
<td>Larch SD3</td>
<td>3,600</td>
<td>24</td>
<td>31</td>
<td>Blended, Content rich – process open</td>
<td>Top 10</td>
</tr>
<tr>
<td>Linden SD4</td>
<td>2,300</td>
<td>30</td>
<td>16</td>
<td>Original, Content rich – process constrained</td>
<td>Top 10</td>
</tr>
<tr>
<td>Willow SD5</td>
<td>1,300</td>
<td>27</td>
<td>47</td>
<td>Blended, Content rich – process open</td>
<td>No trend</td>
</tr>
<tr>
<td>Tesota SD6</td>
<td>1,700</td>
<td>28</td>
<td>33</td>
<td>Blended, Content rich – process constrained</td>
<td>Middle</td>
</tr>
</tbody>
</table>

*The short descriptions of the BOE plans in this column use labels from the work of Baxter and Glaser (1998) on characterizing the cognitive complexity of science assessments in quadrants defined by two axes: content rich to content lean, and process open to process constrained. See more details below, in the Case Study BOE Analysis section.

All requirements were met for the protection of human subjects. No risks or inconveniences were anticipated, as I did not have access to individual teacher or student data. Nor did I disclose which school the data was from, and did not use the data in such a way that it identifies individuals or teachers. This research study was approved by the Institutional Review Board of Montana State University. Prior to the start of interviews, all participants agreed to be in the study, were informed that confidentiality would be
maintained throughout the study and that they could withdraw from the interview at any
time. Additionally, all names and identifying information of the participants were
changed to protect their identities in the case narratives. All district demographics and
the PAWS test results are available via the Wyoming Department of Education public
state website (www.k12.wy.us), but care was taken in this study to report these in a
manner designed to protect the anonymity of the case study districts and participants.
Thus, precise district enrollments and PAWS scores are not reported here.

A curriculum coordinator and a teacher were interviewed from each school district.
The number of years in the current position and teaching assignments are summarized in
Table 6.

<table>
<thead>
<tr>
<th>District</th>
<th>Number of Years in Current Position</th>
<th>Courses Teaching</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD1 CC</td>
<td>3</td>
<td>NA</td>
</tr>
<tr>
<td>SD1 Teacher</td>
<td>4</td>
<td>Physical Science, Biology</td>
</tr>
<tr>
<td>SD2 CC</td>
<td>4</td>
<td>NA</td>
</tr>
<tr>
<td>SD2 Teacher</td>
<td>7</td>
<td>Physical Science, Physics</td>
</tr>
<tr>
<td>SD3 CC</td>
<td>6</td>
<td>NA</td>
</tr>
<tr>
<td>SD3 Teacher</td>
<td>9</td>
<td>Chemistry</td>
</tr>
<tr>
<td>SD4 CC</td>
<td>2</td>
<td>NA</td>
</tr>
<tr>
<td>SD4 Teacher</td>
<td>4</td>
<td>Earth Science</td>
</tr>
<tr>
<td>SD5 CC</td>
<td>6</td>
<td>NA</td>
</tr>
<tr>
<td>SD5 Teacher</td>
<td>5</td>
<td>Physical Science, Earth Science, Biology</td>
</tr>
<tr>
<td>SD6 CC</td>
<td>11</td>
<td>NA</td>
</tr>
<tr>
<td>SD6 Teacher</td>
<td>14</td>
<td>Biology</td>
</tr>
</tbody>
</table>

Pilot Case Study Interviews

After analyzing the BOE survey data, an open-ended, semi-structured interview
guide was developed around the research questions. “An interview guide is prepared to
ensure that the same basic lines of inquiry are pursued with each person interviewed,”
(Patton, 2002, p. 343). See Appendix B for the interview prompts and questions. The
interview protocol was reviewed Dr. Tracy Ragland, Ed.D. Dr. Ragland is familiar with
the BOE system in Wyoming. The protocol was also reviewed by an assessment director,
Dottie Thorson, M.S. Expert feedback was received from three members of the MSU
education faculty (Dr. Elisabeth Swanson, Dr. Carrie Myers, and Dr. Walter Woolbaugh),
and one MSU mathematics/mathematics education faculty member (Dr. Maurice Burke)
who reviewed and commented on the interview questions. Interview questions were
revised based on their suggestions. The first interview was conducted as a conference
call involving the researcher, the participant, and an MSU science education faculty, Dr.
Elisabeth Swanson, who has extensive experience in designing and conducting semi-
structures interviews. Interviews were recorded, transcribed and reviewed with Dr.
Swanson as advisor. This expert feedback contributed to the content validity of the
instrument as well as my proficiency as a facilitator of relatively open-ended interviews
guided by a protocol.

Case Study Implementation

Each interview was conducted over the phone and was digitally recorded and
transcribed by me. The interviews occurred in the fall of 2010 after completion and
analysis of the online BOE Survey. Twelve interviews, including six curriculum
coordinators and a science teacher from the corresponding districts, were selected for
inclusion in the final data analysis. Only one science teacher declined to participate;
therefore, the next recommended name was used.
Case Study Data Analysis

Patton’s (2002) guideline for constructing case studies was used to guide the data analysis process. This process occurred in three steps: 1) assembly of the raw case data, 2) construction of a case record, and 3) writing of the final case narrative. The raw data consisted of transcripts of each interview, as well as the BOE plan for each district. After the individual case studies were created, a cross-case analysis was conducted to highlight similarities and differences in the experiences and perceptions about BOE development, implementation, and use by the participants across the six districts. Interview questions and responses were coded to indicate which research question they addressed. Thematic categories emerged within and across cases.

For this study, individual cases were analyzed separately and the case data was sorted categorically by different roles and experiences of curriculum coordinators versus classroom teachers. Categorical aggregation was used in creating codes from “a collection of instances in the data” and the codes were then reduced into a smaller number of themes (Creswell, 2007). Insightful differences from dialogue and/or BOE plans were marked for use in the case narratives.

Inter-rater reliability is the extent to which two or more individuals agree on the coding, rating or other interpretations of the data. Checking inter-rater reliability is one way to monitor the consistency of the data interpretation approaches in a study. Therefore, other researchers were asked to code raw data from one curriculum coordinator interview and one science teacher interview. The co-raters were an Ed.D. in educational leadership, an M.A. in science curriculum, an M.S. in mathematics, and a
Ph.D. in education from Weston County School District #1 and from Montana State University. Co-raters read through interview transcripts and were provided a list of numbered excerpts from interviews. Co-raters then coded these statements based on the categories and subcategories that emerged from the data. Co-raters were also asked to look for other categories. If discrepancies existed, the co-rater and I discussed them and the categories were edited accordingly. The co-rater was also provided a copy of the narrative and was asked to read and comment about the narrative to ensure that each case was accurately portrayed. In addition, each participant was given the case narrative about their individual case and was asked to read and comment about the narrative to ensure that their case was accurately presented. The participants’ feedback and any additional insights gathered from the replies were incorporated into the final case narrative.

A cross-case analysis was conducted and patterns and themes related to the research problem were generated across curriculum coordinators and science teachers. Similarities and differences were found in experiences with development and implementation of BOE science plans, as well as in the areas of perceived support for the process. From this cross-case analysis, a list of experiences was created. See Chapter 4 for the individual case narratives and the findings of the cross-case analysis.

Case Study BOE Analysis

The district BOE plans were collected from each of the six case study high schools in the study. The plans were reviewed to better understand assessment activities chosen by each district, and how they were implemented, analyzed and later used. A simple analysis was performed to code the different types of plans into similar groups
based on an analytic framework designed by Baxter and Glaser (1998). “One can conceptualize the task demands for content knowledge on a continuum from rich to lean. Similarly, the task demands for process skills can be conceived along a continuum from open to constrained. As shown in Figure 3, assessment tasks can involve many possible combinations of content knowledge and process skills.” (Baxter and Glaser, 1998, p38).

![Figure 3: Content-process Space of Assessment Tasks (Baxter & Glaser, 1998, p39)](image)

Although Baxter and Glaser intended this framework for analyzing task demands, I modified it by categorizing assessment activities within the BOE plans into one of four quadrants: I) content rich - process open, II) content lean - process constrained, III) content lean - process open, or IV) content rich - process constrained. The following are summarized descriptions provided by Baxter and Glaser.

Content rich – process open: assessments are problems that do not have a clean, simple solution and are rich with opportunities for high school students to apply their
subject-matter knowledge and process experiences. In this context, successful performance is dependent on adequate representation of the problem to be solved, sustained and systematic exploration strategies, ad explanation of the relationships observed and tested. The quality of student work reflects the depth of knowledge and process skills students bring to the situation.

Content lean – process constrained: tasks require minimal prior knowledge and limited process experiences for successful completion.

Content lean – process open: assessment tasks require students to coordinate a sequence of process skills with minimal demands for content knowledge.

Content rich – process constrained: tasks emphasize knowledge generation or recall, thus lacking opportunities for activities such as planning, selecting and implementing appropriate strategies or problem-solving procedures.

General information about each case study district’s BOE plan is presented in Chapter 4 to provide additional context for understanding the interview findings. In addition, categorization of the BOE plans are presented, using the Baxter and Glaser framework.

**Study Fidelity and Quality**

Triangulation of data collected offers a means of addressing internal validity issues present in using a mixed methods approach. Patton (2002) describes research and evaluation studies employing combinations of qualitative and quantitative data as common. Using differing sources and methods including the BOE survey, the interviews
of curriculum coordinators and teachers, and review of the BOE plans together provided an understanding of the research questions. The survey tapped a broader range of districts and respondents, and allowed me to generalize about BOE use in school districts. The interviews and BOE plan analysis allowed for the examination of the research questions on a more detailed level. “Triangulation strengthens a study by combining methods. This can mean using several kinds of methods or data, including using both quantitative and qualitative approaches,” (Patton, 2002, p. 247). Plano-Clark and others (2008) reiterate triangulation as a method to improve validity. The strategy of viewing data from different sources and perspectives results in a better understanding of the research questions. The complementary strengths of quantitative numbers and qualitative words offsets the weaknesses of the approaches used alone. The survey developed for this study allowed for larger numbers of respondents, a method that can be replicated by others. Qualitative interview methods and BOE plan review, although more difficult to generalize, allow researchers to capture unique perceptions about the research topic. They also allow the researcher to document unanticipated perspectives. This multiple perspective approach increases understanding and improves the quality of the study.

**Researcher Perspectives**

It is important to explicitly recognize that “The perspective that the researcher brings to a qualitative inquiry is part of the context for the findings. Self-awareness, then, can be an asset in both field-work and analysis” (Patton, 2002, p. 64). Thus, it is good practice to relay personal information about the researcher when discussing research
methodology. My experiences as a science teacher and my work with the BOE system provide me with perspectives that I bring to this study, so I have described those experiences here.

I have been a graduate student at Montana State University for the past six years working toward a doctoral degree while continuing to teach full time in Newcastle, Wyoming. I have nineteen years of high school science teaching experience. I have participated in a variety of professional learning experiences, ranging from one-day in-services to prolonged coursework and multi-year institutes. These experiences have contributed to my perceptions of effective professional development and support for teachers as new programs, such as the BOE, are implemented within a school, district or state. In addition, during this time I have participated in several years of both formative and summative assessment trainings. I have been involved with reviewing PAWS questions and data analysis, along with anchor pulling of student responses. I have written, piloted and revised BOE assessments at both the district and state levels. These experiences have shaped my views of the importance of assessment for learning and how it differs from assessment of learning. These experiences have also provided the background and interest needed to carry out this mixed-methods study of the Wyoming Body-of-Evidence Science Assessment Plan.

“For better or worse, the trustworthiness of the data is tied directly to the trustworthiness of the person who collects and analyzes the data – and his or her demonstrated competence” (Patton, 2002, p. 570). I have made every attempt to maintain intellectual rigor, professional integrity, methodological competence, and neutrality.
Qualitative inquiry, because the human being is the instrument of data collection, requires that the investigator carefully reflect on, deal with, and report potential sources of bias and error. Systematic data collection procedures, rigorous training, multiple data sources, triangulation, and external reviews are aimed at producing high-quality qualitative data that are credible, trustworthy, authentic, balanced about the phenomenon under study, and fair to the people studied. (Patton, 2002, p. 53)

As a researcher, I have interpreted the findings of my research as closely as possible to the meanings described by the participants in the study. I have worked with others to ensure the processes used to interpret the results are aligned with the questions being asked, thus creating credibility in the findings. Their recommendations limited the effect that my biases may have had when collecting data and reporting results. Although I am more than acquainted with some of the subjects in this study, Patton (2002) suggests that the perspective that the researcher brings to the study is part of the context for the findings. Throughout, I have strived to maintain trust through respect for participants’ ideas and opinions.

**Delimitations and Limitations**

Despite the measures taken to optimize study quality, there are several potential limitations inherent in the study. The survey used in this study involved current Wyoming school administrators, high school principals, and high school science teachers. There are 48 high schools in Wyoming, yet the number of respondents to the survey limited the study. Non-respondents are always problematic since their lack of participation can affect the conclusions drawn from the analysis of the data. Responses to survey questions were of concern since respondents can potentially answer questions
not as they see themselves, but as they would like to see themselves. Also, this process tends to obtain a sample that consists of pretty highly motivated individuals. The teachers who are not as motivated, may not bother responding to the survey. Without firsthand observations of the respondents’ practices, the survey data collected is subject to variance in accuracy. However, the interviews and BOE plans collected provided evidence to support the self-perceptions of BOE implementation and experiences.

Another limitation was the number of years respondents have been in their current position of employment or the extent of their experiences with the BOE process. Only two interview participants have been involved in the full implementation of the BOE system since its inception in 2000, and less than half of survey respondents had this level of experience with the system. Interview participants were selected by the researcher based on the demographic profile of the school district. However, care must be taken in generalizing the results because the final selection of interview participants was left to self-volunteering for the interview; thus participants may have uncommonly high interest in the BOE process. Since this study only collected data regarding the Wyoming BOE science assessment system, generalizability to other assessment systems, disciplines and states may be limited. However, this study can be used to inform a larger audience of program coordinators, assessment developers and science teachers about the strengths and weaknesses of implementing portfolio-like assessment plans.
CHAPTER 4

RESULTS OF THIS STUDY

Introduction

This chapter presents the results of the analysis of data collected in this mixed methods study. In keeping with the format of the discussion of the data collections procedures in Chapter 3, the following results are presented and discussed: the survey results obtained in the preliminary phase of the study, results from the district case-study interviews and analysis of BOE plans. The first section of this chapter presents the results of the quantitative BOE survey analysis, while the second section presents the results of the case study analysis in the form of six district case-narratives and a cross-case analysis. In the third section I discuss the themes that emerge and summarize the story these results tell collectively about the design and implementation of BOE science plans across the state of Wyoming, as they pertain to the research questions.

BOE Survey

This section describes the results of the analysis of the BOE Survey, which can be found in Appendix A. Appendix C maps the research questions, methods of analysis and the BOE survey questions. A total of 110 survey responses were received and analyzed. Participants self-categorized by selecting their main role within their school district as a high school science teacher, a high school administrator, a district curriculum coordinator, a district administrator, or another district employee. Sixty-six Wyoming
high school science teachers participated in the survey which represented 60.55% of the survey respondents. Ten district curriculum coordinators, eleven high school administrators, thirteen district administrators, and nine other school district employees (filling roles such as data coordinator and instructional facilitator) also took the survey.

The modal range of years in the current position across all respondents was 0-5 years of experience, the choice selected by 43 respondents. Although administrative personnel appeared to skew the mode toward the low end, two-thirds of the high school teachers responding had been in their present position longer than five years. Table 7 shows the distribution of survey respondents by professional categories and years of experience in their current position.

<table>
<thead>
<tr>
<th>Role</th>
<th>Number of Survey Participants (&lt;i&gt;n=109&lt;/i&gt;)</th>
<th>Years in current position</th>
</tr>
</thead>
<tbody>
<tr>
<td>High School Science Teacher</td>
<td>66</td>
<td>0-5 years: 20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6-11 years: 15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12-20 years: 17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Over 20 years: 14</td>
</tr>
<tr>
<td>District Curriculum Coordinator</td>
<td>10</td>
<td>0-5 years: 9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6-11 years: 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12-20 years: 0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Over 20 years: 0</td>
</tr>
<tr>
<td>High School Administrator</td>
<td>11</td>
<td>0-5 years: 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6-11 years: 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12-20 years: 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Over 20 years: 0</td>
</tr>
<tr>
<td>District Administrator</td>
<td>13</td>
<td>0-5 years: 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6-11 years: 7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12-20 years: 0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Over 20 years: 0</td>
</tr>
<tr>
<td>Other</td>
<td>9</td>
<td>0-5 years: 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6-11 years: 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12-20 years: 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Over 20 years: 1</td>
</tr>
</tbody>
</table>
BOE Participation

Thirty-nine percent of respondents reported having participated in a BOE peer-review at the state level. This reflects significant training and familiarity with the BOE framework as described in chapter one. Conducting a two-way contingency table analysis using crosstabs, we can see that more district administrators, district curriculum coordinators and high school science teachers have participated in the BOE peer-review process than have high school administrators and other district personnel who took the survey. These results are displayed in Table 8.

Table 8. Results Regarding Participation in BOE Peer-Review at the State Level by Respondent Role.

<table>
<thead>
<tr>
<th>Role</th>
<th>BOE Peer-Review Participation</th>
<th>No</th>
<th>yes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>high school science teacher</td>
<td>Count</td>
<td>51</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>% of Total</td>
<td>46.8%</td>
<td>13.8%</td>
</tr>
<tr>
<td>district curriculum coordinator</td>
<td>Count</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>% of Total</td>
<td>0.9%</td>
<td>8.3%</td>
</tr>
<tr>
<td>high school administrator</td>
<td>Count</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>% of Total</td>
<td>4.6%</td>
<td>5.5%</td>
</tr>
<tr>
<td>district administrator</td>
<td>Count</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>% of Total</td>
<td>3.7%</td>
<td>8.3%</td>
</tr>
<tr>
<td>Other</td>
<td>Count</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>% of Total</td>
<td>5.5%</td>
<td>2.8%</td>
</tr>
<tr>
<td></td>
<td>Total Count</td>
<td>67</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>% of Total</td>
<td>61.5%</td>
<td>38.5%</td>
</tr>
</tbody>
</table>
To what extent does a teacher’s understanding compare to others, in the school, district, or state? When asked if they had participated in developing a BOE science assessment plan in their current position, 87.5% of high school science teachers had, while only 53.8% of district administrators had taken part in such an endeavor. These responses reflect the distributed involvement of roles within the district. Table 9 summarizes these responses.

Table 9. Cross-tabulation Results of Role x Participation in BOE-science Plan Development.

<table>
<thead>
<tr>
<th>role</th>
<th>BOE-science plan participation</th>
<th>No</th>
<th>Yes</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>high school science teacher</td>
<td>Count</td>
<td>8</td>
<td>58</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>% within role</td>
<td>12%</td>
<td>88%</td>
<td></td>
</tr>
<tr>
<td>district curriculum coordinator</td>
<td>Count</td>
<td>3</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>% within role</td>
<td>30%</td>
<td>70%</td>
<td></td>
</tr>
<tr>
<td>high school administrator</td>
<td>Count</td>
<td>5</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>% within role</td>
<td>45%</td>
<td>55%</td>
<td></td>
</tr>
<tr>
<td>district administrator</td>
<td>Count</td>
<td>6</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>% within role</td>
<td>46%</td>
<td>54%</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>Count</td>
<td>4</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>% within role</td>
<td>44%</td>
<td>56%</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>Count</td>
<td>26</td>
<td>83</td>
<td>109</td>
</tr>
<tr>
<td></td>
<td>% within role</td>
<td>23.9%</td>
<td>76.1%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

However, when role is cross-tabulated with self-reported familiarity with the district science assessment plan matrix, science teachers fall just below high school administrators regarding familiarity about the alignment of their science assessments to the state science standards. It is noteworthy that 75 – 100% of respondents in all categories reported familiarity with the matrix. Curriculum coordinators and district
administrators indicated a 100% familiarity with this concept within their districts. Results are summarized in Table 10.

Table 10. Cross-tabulation of Role by Familiarity with District Science Matrix.

<table>
<thead>
<tr>
<th>Role</th>
<th>Count (% within role)</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>high school science teacher</td>
<td>7 (11.1%)</td>
<td>56 (88.9%)</td>
<td></td>
</tr>
<tr>
<td>district curriculum coordinator</td>
<td>0 (0.0%)</td>
<td>10 (100.0%)</td>
<td></td>
</tr>
<tr>
<td>high school administrator</td>
<td>1 (10.0%)</td>
<td>9 (90.0%)</td>
<td></td>
</tr>
<tr>
<td>district administrator</td>
<td>0 (0.0%)</td>
<td>11 (100.0%)</td>
<td></td>
</tr>
<tr>
<td>other (data coordinator &amp; instructional facilitator)</td>
<td>2 (25.0%)</td>
<td>6 (75.0%)</td>
<td></td>
</tr>
<tr>
<td>Total %</td>
<td>(10) 9.8%</td>
<td>(92) 90.2%</td>
<td></td>
</tr>
</tbody>
</table>

**BOE Activity Structure**

Are BOE science activities seamlessly incorporated into instructional plans, or are they stand-alone assessment practices? Looking into this research question, 71% of participants reported their district science assessments are integrated within classroom instruction, as illustrated in Figure 4. Answers of the nine respondents who selected “other” included “We try to integrate them as assessments within classroom instruction, but it doesn't work that way 100% of the time” and several stated “both,” indicating schools select a blend of instructional integration with stand-alone assessments.
Do science assessment plans provide multiple opportunities for students to retake individual assessment pieces? Participants reported “yes” at a rate of 79% as illustrated by Figure 5. As specified in statute W.S. 21-3-110(a)(xxiv) and Chapters 6 and 31 of the Current Rules and Regulations, to receive a diploma, students must meet both the Carnegie unit requirements and demonstrate proficiency on the Standards. The practice of multiple opportunities would lead to higher student proficiency rates on the Standards.
When asked if their BOE science plan included any BOE-consortium activities, as described in chapter 1, 72.0% of respondents indicated they used one or more consortium activities, 16.0% reported using no consortium activities, and 12.0% reported they were not familiar enough with their district science BOE plan to answer this question. The most popular consortium science activities were #8: Endangered Species, #5: Giants of Science, and #4: Scientific Inquiry as illustrated in Figure 6. The popularity of these activities is reflective of the broad contexts in which they can be applied.
Seventy-two respondents indicated their district uses one or more consortium activities in their science BOE assessment plan. The number of consortium activities used ranges from 1 – 13, with the most districts utilizing one or four activities. These responses are summarized in Table 11. One interesting insight within the responses to this question is that although 100% of administrators indicated being familiar with their district science assessment plan alignment to state science standards, four administrators answered this question with response C: I am not familiar enough with our district plan to answer this question. Is this a reflection that administrators task curriculum coordinators to verify alignment and then trust the results? More probing would be necessary to verify this assumption.
When it comes to looking at how districts analyze their proficiency rates, 60% of respondents reported their district uses cut scores, 20% reported using standard mastery, 9% used other methods. Setting cut scores is the process of determining the scores that divide various performance levels. Standards mastery means that pieces of knowledge must be learned and certain skills acquired at a level of proficiency as described in the Wyoming Science Content and Performance Standards. Respondents selecting “other” methods reported a mixture of cut scores, standard mastery and classroom grades. Eleven percent of respondents reported not knowing what their district used for proficiency criteria. Respondents were almost split 50/50 when asked if they had participated in setting cut-scores for the science assessment plan in their district, with 53% indicating they had participated in this process. The group with the largest participation rate was the district curriculum coordinators, but surprisingly, only 70% in this category had participated in this process. Perhaps that process had been accomplished before they took their position as curriculum coordinator.
Participants were asked to estimate how many hours they have spent working on their BOE science plan in the last three years. Choices provided were 0-8 hours, reflecting one day of professional development, 9-16 hours, representing two days of professional development, and so forth. The mean number of hours spent working on BOE science assessment plans in the last three years is 10.81 hours, just over one day of professional development time spent on this task. One and a half times more participants indicated spending less than 17 hours than indicated spending more than 17 hours. Seventeen hours was chosen as the cut-off for the survey, as it represented spending more than three days of professional development time, or an average of one day per year on BOE work.

**Likert Items**

The next section of the survey included three Likert items pertaining to satisfaction with their BOE experiences. The first Likert-scale question on the survey asked participants to rate their experience developing the BOE plan in their district. Means were based on a 5-point scale, 1 = displeased, 2 = somewhat displeased, 3 = neutral, 4 = pleased, and 5 = extremely pleased. Using the same scale, the next question asked participants to rate their experience implementing the BOE plan. The last question asked about impact on student learning taking place. Mean, standard deviation, skew, kurtosis, and Cronbach’s Alpha (if the item is deleted) are reported in Table 12 for each of the three scale questions. As evidenced by this data, participants are more displeased with BOE topics than pleased. For this data, positive skew values indicate the responses are clustered more to the left of the mean with extreme values trailing off far toward the
right side of the mean, and visa-versa for negative skew values. The negative kurtosis values indicate a curve with a broad, low peak, below that of a normal curve.

Table 12. Results of Likert-scale Questions about BOE Impact.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Mean (5-pt scale)</th>
<th>SD</th>
<th>Skew</th>
<th>Kurtosis</th>
<th>Cronbach’s Alpha if item is deleted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Developing BOE plan</td>
<td>2.84</td>
<td>1.114</td>
<td>-0.070</td>
<td>-0.898</td>
<td>0.792</td>
</tr>
<tr>
<td>Implementing BOE plan</td>
<td>2.78</td>
<td>1.098</td>
<td>0.038</td>
<td>-0.796</td>
<td>0.773</td>
</tr>
<tr>
<td>Impact on Student Learning</td>
<td>2.70</td>
<td>1.279</td>
<td>-0.130</td>
<td>-1.290</td>
<td>0.779</td>
</tr>
</tbody>
</table>

Note. Means based on 5-point scale (1 = displeased, 5 = extremely pleased).

The Likert items and survey responses are displayed in Tables 13-15. It is interesting to see how the mean changes with the employment roles. Curriculum coordinators rate their professional development experience with the BOE plan one and a half times higher than science teachers rate their experience. Perhaps this reflects a difference in who is providing the training to the different entities.

Table 13. Range of Responses to Experience Developing the BOE Plan.

<table>
<thead>
<tr>
<th>Survey Question: As a professional development experience, how do you rate your experience developing the BOE plan in your district?</th>
<th>Science Teachers (n=62)</th>
<th>Curriculum Coordinators (n=9)</th>
<th>High School Administrators (n=9)</th>
<th>District Administrators (n=10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 = displeased</td>
<td>13 (21.0%)</td>
<td>0</td>
<td>1 (11.1%)</td>
<td>1 (10.0%)</td>
</tr>
<tr>
<td>2 = somewhat displeased</td>
<td>20 (32.3%)</td>
<td>0</td>
<td>2 (22.2%)</td>
<td>1 (10.0%)</td>
</tr>
<tr>
<td>3 = neutral</td>
<td>16 (25.8%)</td>
<td>2 (22.2%)</td>
<td>3 (33.3%)</td>
<td>4 (40.0%)</td>
</tr>
<tr>
<td>4 = pleased</td>
<td>10 (16.1%)</td>
<td>7 (77.8%)</td>
<td>3 (33.3%)</td>
<td>4 (40.0%)</td>
</tr>
<tr>
<td>5 = extremely pleased</td>
<td>3 (4.8%)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Avg. Rating</td>
<td>2.52</td>
<td>3.78</td>
<td>2.89</td>
<td>3.10</td>
</tr>
</tbody>
</table>
Table 14. Range of Responses to Experience Implementing the BOE Plan.

<table>
<thead>
<tr>
<th>Survey Question: How do you rate your experience with the implementation of the BOE plan in your district?</th>
<th>Science Teachers (n=61)</th>
<th>Curriculum Coordinators (n=9)</th>
<th>High School Administrators (n=9)</th>
<th>District Administrators (n=10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 = displeased</td>
<td>12 (19.7%)</td>
<td>0</td>
<td>1 (11.1%)</td>
<td>2 (20.0%)</td>
</tr>
<tr>
<td>2 = somewhat displeased</td>
<td>19 (31.1%)</td>
<td>1 (11.1%)</td>
<td>3 (33.3%)</td>
<td>3 (30.0%)</td>
</tr>
<tr>
<td>3 = neutral</td>
<td>15 (24.6%)</td>
<td>3 (33.3%)</td>
<td>2 (22.2%)</td>
<td>4 (40.0%)</td>
</tr>
<tr>
<td>4 = pleased</td>
<td>14 (23.0%)</td>
<td>4 (44.4%)</td>
<td>2 (22.2%)</td>
<td>1 (10.0%)</td>
</tr>
<tr>
<td>5 = extremely pleased</td>
<td>1 (1.6%)</td>
<td>1 (11.1%)</td>
<td>1 (11.1%)</td>
<td>0</td>
</tr>
<tr>
<td>Avg. Rating</td>
<td>2.56</td>
<td>3.56</td>
<td>2.89</td>
<td>2.40</td>
</tr>
</tbody>
</table>

When asked to rate their experience implementing the BOE plan, again we see that curriculum coordinators have the most pleased view, yet only slightly above neutral with a mean of 3.56. However, this is still 1.4 times higher than the view of science teachers who are charged with implementation of the plans in their classrooms. Science teachers are more displeased with the implementation experience, as reflected in the mean of 2.56.
Table 15. Range of Opinion About Impacts on Student Learning as a Result of BOE Plan.

Survey Question: What is your opinion of the impact on student learning that is taking place in your classroom as a result of the BOE assessment plan your district has adopted?

<table>
<thead>
<tr>
<th></th>
<th>Science Teachers (n=61)</th>
<th>Curriculum Coordinators (n=9)</th>
<th>High School Administrators (n=9)</th>
<th>District Administrators (n=10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 = displeased</td>
<td>20 (32.8%)</td>
<td>0</td>
<td>3 (33.3%)</td>
<td>3 (30.0%)</td>
</tr>
<tr>
<td>2 = somewhat displeased</td>
<td>10 (16.4%)</td>
<td>0</td>
<td>0</td>
<td>1 (10.0%)</td>
</tr>
<tr>
<td>3 = neutral</td>
<td>13 (21.3%)</td>
<td>3 (33.3%)</td>
<td>4 (44.4%)</td>
<td>5 (50.0%)</td>
</tr>
<tr>
<td>4 = pleased</td>
<td>16 (26.2%)</td>
<td>5 (55.5%)</td>
<td>2 (22.2%)</td>
<td>1 (10.0%)</td>
</tr>
<tr>
<td>5 = extremely pleased</td>
<td>2 (3.3%)</td>
<td>1 (11.1%)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Avg. Rating</td>
<td>2.51</td>
<td>3.78</td>
<td>2.56</td>
<td>2.40</td>
</tr>
</tbody>
</table>

When surveyed about the impact on student learning, curriculum coordinators are again the most optimistic, with no respondents being displeased, and the highest mean of the categories polled. However, district administrators surveyed did not share this view, holding the lowest opinion of the four groups with a mean of 2.40 and 90% being neutral or being displeased with the impact on student learning.

One additional analysis for the Likert-type items was performed to look at the mean values from respondents with BOE peer review experience. With only one exception, every question and every role had a lower mean value, indicating respondents who have BOE peer review experience are more displeased with their experiences than their counterparts who have not participated in the state level peer review process. These results are displayed in Table 16.
Table 16. Means of Likert Items for Respondents whom had Peer Review Experience versus the total survey means.

<table>
<thead>
<tr>
<th>Item</th>
<th>Science Teachers (n=15)</th>
<th>Curriculum Coordinators (n=9)</th>
<th>High School Administrators (n=6)</th>
<th>District Administrators (n=9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 1</td>
<td>2.29/2.52</td>
<td>3.67/3.78</td>
<td>2.83/2.89</td>
<td>3.00/3.10</td>
</tr>
<tr>
<td>Item 2</td>
<td>2.43/2.56</td>
<td>3.33/3.56</td>
<td>2.67/2.89</td>
<td>2.14/2.40</td>
</tr>
<tr>
<td>Item 3</td>
<td>2.71/2.51</td>
<td>3.67/3.78</td>
<td>2.50/2.56</td>
<td>2.29/2.40</td>
</tr>
</tbody>
</table>

Open-Ended Responses

Sixty-six respondents shared additional thoughts about ways they see student learning taking place in the classroom as a result of the BOE assessment plan adopted by their district. Comments ranged from positive, “Students are actively engaged in learning through experimentation,” to negative, “The BOE assessments contribute little to student learning.” One comment, “Student learning in our district does not hinge on the BOE assessment plan, rather on our common district assessments,” exemplifies a misconception that the BOE is a separate issue from common district assessments. In reality, they should be referring to the same requirement. Twenty-six participants indicated they would be willing to be interviewed about the development of the BOE plan in their district.

Summary of Survey Results

When Wyoming teachers and administrators were surveyed about their experiences developing and implementing a BOE science assessment plan, the results were not positive about this system which has been required by state law for over ten years. The most discrepant difference in opinions about developing the BOE plan
occurred between teachers, who are generally more displeased with their experience developing the BOE plan, and curriculum coordinators, who are the most pleased of the groups surveyed. When surveyed about their experience implementing the BOE plan, the district administrators indicated the most displeasure of the groups. Although the consensus across the board was that respondents are generally not pleased with their BOE design and implementation experiences, the most neutral responses were expressed about the impact on student learning as a result of the BOE plan development.

Some errors affecting survey results are that respondents may not have answered all of the questions on the survey. Although 110 surveys were taken, there were not 110 responses for each question. One example is a teacher that skipped the question about years of experience.

According to Alreck and Settle (2004) the single biggest important limitation in surveying is a relatively low response rate. Response rates are often only 5-10 percent for mail surveys. At Michigan State University, Kaplowitz, Hadlock & Levine (2004) measured e-mail response rates at just below 21%. Wyoming high school science teachers responded at a 30% rate (67/222) to this survey, and 13 out of 48 district administrator responses made the response rate for that subgroup 27%. The lowest response rate came from curriculum coordinators with 10 of 48 replying to the survey, a response rate of 21%. Self-selection bias was a concern at the onset of the survey because it was thought that individuals who had participated in a BOE-peer review at the state level would show a high interest in survey participation. However, this was not the case, as only 38.5% of survey respondents indicated they had participated in a BOE-peer
review at the state level. The less than ideal response rate was probably more indicative of non-response bias. From the low Likert-scale means, one might presume that individuals who were disgruntled with the BOE process replied to the survey more frequently than those appreciating the process. However, this cannot be assumed to be universally true because many of the responses to the open-ended question indicated a better understanding of the standards is being addressed. For example, someone can be disgruntled with their experience during the design or implementation process, or even harbor a feeling that the BOE could be vastly improved, yet still feel that there were benefits for students. Thirty-five responses were positive comments to the question “In what ways do you see student learning taking place in your classroom, as a result of the BOE assessment plan adopted by your district?” as compared to 25 negative comments. Many participants left more than one comment and, of those, nine responses are classified as neutral comments.

**District Case Studies**

This section describes the development and implementation of six BOE Science Assessment plans through case narratives of both district curriculum coordinators and high school science teachers. Each participant was employed in his/her current position for at least two years, during the 2009-2010 and 2010-2011 school years, the period during which this study took place. The data in the narratives was gained from personal interviews and a BOE assessment plan, when provided. The data was analyzed and
synthesized across the six districts to provide rich context and was used to answer the research questions.

Each district narrative provides a description of the case participants’ (curriculum coordinators and a science teacher) perceptions about the BOE development and implementation in their current school district. A group of categories reappeared within and across cases during data analysis. Some corresponded to categories built into the interview questions, and others emerged through coding and analysis of the data. The embedded categories included BOE characteristic descriptions (processes and products, principles and philosophies, graduation requirements, and implementation), reflections on design principles (consistency, fairness, comparability, multiple measures, credibility and consequences), alignment to state standards, support, data reporting, and effects on student learning. Emergent categories revealed drift in the BOE process; as time moves further away from the initiation point, districts move further away from the original intentions of the BOE process and move toward individualized local content. Perceptions emerged about the Wyoming Science Standards being too vague to accurately define knowledge and skills being assessed. There were uneven perceptions of, as well as levels of support for, BOE planning, implementation and sustainability. Subcultures appeared to develop around the value assigned to the BOE process and varied within and among districts. These categories were used to organize the data.

Each district narrative describes participants’ experiences with the design and implementation of the BOE science assessment plan. Categories described within and across cases include BOE design and implementation, alignment to Wyoming state
science standards, descriptions of understandings about Wyoming BOE design principles, support for BOE processes, BOE data reporting and usage, and reflections about BOE impact on student learning. Quotes from the interviews and assessment plans were used to support the findings. All names of the districts and participants were changed for confidentiality. Districts were named after trees of the Western US and names of individual participants were randomly selected names of the researcher’s former students. Answers have not been adjusted to reflect grammar errors. Word substitutions for confidentiality or clarity are enclosed in parentheses.

Through the analysis of six Wyoming case studies, representing 12.5% of school districts, it was evident there is not consistency of understanding about rules and regulations governing BOE design and implementation. The interpretation of Wyoming BOE design principles varied somewhat from case to case. However, while all participating districts submitted district BOE plans that passed the state BOE-peer review, not all passed on the first submission. This indicates a lack of understanding about the BOE process the state requires, even after ten years of implementation. A review of terminology, which appears in Table 17 on the next page, will assist readers in interpreting interview comments.
<table>
<thead>
<tr>
<th><strong>Table 17. Wyoming BOE Design Principle Terminology.</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Alignment</strong></td>
</tr>
<tr>
<td>The combination of assessments that comprise the system are aligned with district content and performance standards so that the full set of standards, both in terms of content and cognitive complexity, are assessed. If the district is incorporating course-based components in the BOE system, it must demonstrate that the curriculum is aligned with both the standards and assessments. In order to meet the alignment criterion, the district must indicate how the assessment system reflects the district’s prioritization of the standards. Finally, multiple assessment measures and formats are employed in the system to maximize the alignment between standards and assessments.</td>
</tr>
<tr>
<td><strong>Consistency</strong></td>
</tr>
<tr>
<td>The focus of this evaluation should be concentrated on the system and should examine, for example, how different judges would evaluate the same set of data about a group of potential graduates. In order to satisfy this criterion, the district should also document that the results of the assessments are not overly influenced by error due to raters or the specific tasks/items used comprising the assessments. Individual assessments within the system need to be evaluated for consistency, in terms of error due to raters, tasks, administration conditions, and occasions.</td>
</tr>
<tr>
<td><strong>Fairness</strong></td>
</tr>
<tr>
<td>The assessment system shall be designed, implemented, and evaluated so that it is not biased against any groups of students. Appropriate accommodations need to be employed so students with disabilities and Limited English Proficient students have as fair a chance as possible to demonstrate what they know. Multiple assessment opportunities and formats should be used to maximize fairness. The results of the assessments comprising the system and the results of the system itself should be disaggregated to examine both the fairness of the assessment system and opportunities for all students to learn the standards.</td>
</tr>
<tr>
<td><strong>Standard-Setting</strong></td>
</tr>
<tr>
<td>The method for establishing cut-scores between various performance levels on the BOE should be based on a defensible methodology and the district should indicate a clear rationale for choosing one method over another. The method selected should incorporate clear descriptions of the performance levels and not be based on arbitrary performance distinctions (e.g., traditional percentages).</td>
</tr>
<tr>
<td>Table 17 continued</td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td><strong>Comparability</strong></td>
</tr>
<tr>
<td><strong>Multiple Measures</strong></td>
</tr>
<tr>
<td><strong>Credibility</strong></td>
</tr>
<tr>
<td><strong>Consequences</strong></td>
</tr>
</tbody>
</table>

**Cottonwood**

Cottonwood is a large school district in Wyoming. Cottonwood’s student enrollment is 27% minority, and 38% of students qualify for free and reduced lunches. Cottonwood employs an original district-developed BOE plan, and they do not show a two-year PAWS science-test score trend at the high school level. Cottonwood’s curriculum coordinator “Barry” has been in the current position three years, and the
science teacher interviewed “April” has been in the current position four years. When interviewed, both described their BOE system as district-developed quarterly tests that are given in three different courses, and then one performance district assessment is given each semester, for a total of 18 district assessments over three years. No Wyoming Body of Evidence Consortium activities are used in this school district. Barry described a student’s graduation requirements from Cottonwood that would include an 8th grade Earth & Space Science class, a 9th grade Physical Science class, a 10th grade Biology class and one other elective as a junior or a senior. April understood that students had to take 9th grade Physical Science, 10th grade Biology, plus one other high school science class in order to graduate.

**BOE Characteristics.** When asked to address how their BOE assessment plan was implemented, Barry described:

> Just this past year, we kind of solidified our guaranteed and viable curriculum and that curriculum for each science course has a very solid deadline for each assessment, so the teachers do not choose when they give the assessment. Every teacher in the district gives the assessment within about the same one or two week window without any exceptions.

How does this implementation in Cottonwood look from a teacher’s perspective?

April’s explanation included:

> …as far as I know, they do have what are called district assessments for the biology and physical science classes, and those assessments are given quarterly. And so, basically what it is, and they are driven by our curriculum, which has just been redone this summer, and will be applied starting on Wednesday, so what these assessments involve is several inquiry questions, as well as content questions for whatever the quarter unit happens to be, and so they are given every quarter. And then also once a semester, the students also have to do a performing district assessment, and with that, for example, in biology, they have to do a cell
transport for an assessment, so it’s total inquiry based, the students have to develop their entire lab, from start to finish, the teachers cannot help them, they can just provide them with materials, and so they have to do, they actually have to plan, conduct and analyze their experiments, write it up, and then teachers have to grade them according to a rubric everyone in the district uses, and then we also have to pick a sample, and we have to switch them with another person who also teaches the same course at a different school, so that we have the inter-rater reliability. So that we know that everybody is on the same page when it comes to grading.”

Similarities in Barry and April’s explanations include their understanding of the quarter tests administered to students in required science classes. These are common assessments that every teacher of physical science students gives at approximately the same time of year. April’s response focuses more on the performance assessment, probably from the consequence that it takes more of her instructional time to prepare and score than does the district-made quarter tests, which are multiple choice tests.

Barry and April had differing views about the principles or philosophies that guide the BOE assessment system in their district, a recurring difference identified in five out of six of the districts interviewed. April identified her understanding:

I would say that whatever is best for the students, whatever’s going to get them through, whatever gets to them, whatever is best for student achievement.

Barry’s interpretation was that it is more driven by standards:

I really think that the only principles that guide it, concept-wise is the Wyoming state standards, that’s the bible of everything we do and philosophy-wise it’s that students deserve multiple opportunities or we could just have a graduation test like they do in some states, but I think that Wyoming is very dedicated to the multiple opportunities of assessment, that students need to have many opportunities to demonstrate their proficiency, that drives our system.
Some differences are evident in these two responses. April mentions, “whatever is best for students,” yet it is unclear whether she is referring to student understanding or just enabling students to “get through.” Barry starts by citing the importance of the Wyoming science standards, but then he segues into the importance of not just having a graduation test like some other states, but giving students multiple opportunities for success. So, in the end, he appears to care not only about state standards, but also about students.

In an analysis of the assessment activities within Cottonwood’s assessment plan, using the Baxter and Glaser (1998) assessment task framework I categorized this plan as content-rich and process-open. Each quarter teachers administer a district-developed written assessment of content, and each semester students partake in a performance assessment using scientific inquiry skills. Some of the performance assessments are consortium activities, which makes this plan a blended plan, and not solely district-developed as might have been inferred from the interview with the curriculum coordinator.

Alignment to State Standards. In terms of the relationship between standards and assessment, alignment refers to the match between the items on the assessments and the knowledge and skills represented by the curriculum and the standards. Alignment is often misinterpreted because it is more than simply matching the content of test items with a standard. In a two-way alignment system districts must assure that all the standards are assessed, in terms of the performance levels established in the standards.
Both Barry and April indicated their district curriculum and assessments are fully aligned to the Wyoming State Science Standards. Barry explained:

We have a district curriculum that’s more specific than the state standards. The state standards were written in general terms on purpose, so that we still have some local control, but our curriculum is fully based on the Wyoming state science standards. And actually, we don’t have science curriculums in our district; we have a science curriculum in our district, and it takes students 13 years to complete. It’s fully aligned K to 12.

Barry did not refer to the two-way cognitive alignment to the standards; however, when prompted, April did:

We made sure that, not on the performing ones, there’s not really a whole lot of opportunity on that, but on the written ones, we made sure there’s a wide variety of questions that meet all the different levels of activity.

**Additional Wyoming BOE Design Principles.** During each interview, participants were asked to reflect on six of the BOE design principles as presented in the Wyoming Assessment Handbook: consistency, fairness, comparability, multiple measures, credibility and consequences. Although curriculum coordinators and teachers had different interpretations of how their BOE system addressed these requirements, all had a sense of the meaning of the terms.

Cottonwood has addressed the requirement of the consistency principle by assessing all the students in Physical Science classes across the district, with the same exams and performance assessments, administered within a two-week window. The same was described for all Biology classes. Fairness was interpreted from two different angles. Barry described fairness in the planning of assessments, by having special education instructors participate in the planning of the testing prompts. April saw
fairness as teachers being “on the same page when it comes to grading.” She indicated that her district does training on how to score the performance assessments and that teachers trade a few papers with one another to grade in order to help ensure grading is being done consistently across teachers.

In order to compare results from year to year, Barry identified Cottonwood’s strategy for addressing the comparability design principle:

We make sure we don’t make large amounts of changes all in one big jump, because then you can’t compare last year’s performance to this year’s students’ performance. And if you don’t have comparability then you can’t justify the decisions you’re making, so we always try to bite off little pieces at a time if we’re making changes.

April reflected that comparability relates to the training they participate in each year about how to grade the district assessments.

How is the multiple measures design principle addressed within Cottonwood School District? Are there multiple opportunities for students to master the standards?

From her experiences, April described frustration with the system:

Anytime, after they’ve taken it and their initial assessment has been graded, and they can see the results, and they have anytime after that, throughout the whole school year. They even have an opportunity to make-up those assessments for us during summer school, like at my school, we actually have an achievement center where they can do it in there too.

This system of multiple opportunities was confirmed by Barry, and sounds different from that described by any of the others schools in the study.

In order to establish credibility with stakeholders, which is the seventh BOE design principle, Barry described the process of inviting special education teachers to
participate in science committee meetings. April indicated she wasn’t involved with this and apparently misunderstood this requirement as she described seeing people walk around her school and talk to parents and students, but not being involved with it, she didn’t really know. This was not uncommon among the teachers interviewed. Two other teachers described school improvement meetings, and two different teachers described science-teacher only meetings.

What are the consequences for students at Cottonwood if they are not proficient on an assessment? Both Barry & April indicated that students could repeat the assessment in the achievement center, or summer school; or if needed, they could repeat the class again. Because students in Wyoming have to meet proficiency standards in five of nine content areas, they could graduate without being proficient in the science content area.

Support for BOE process. April was one of two of the teachers interviewed who had not participated in the scheduling of BOE assessments in her district, although she indicated she had helped design assessments for her district BOE plan. It appears that teachers are not involved in the process of setting cut-scores for the BOE plans, as April was one of four teachers who had not participated in the cut-score setting process either. She sees the support she receives in order to carry out her requirements with the BOE assessments as coming from her science curriculum coordinator as well as the feedback of results on the assessments also coming from Barry’s office. This is similar to one other science teacher interviewed. Curriculum coordinators seem to have a different lens of support. Like many of his peers interviewed, Barry sees support for the BOE process as
coming from both the district and the state level, albeit for different purposes. Three curriculum coordinators indicated financial support for the process occurs at the district level, while the state provides procedural support. Barry verbalized that he has participated in all levels of BOE planning, from design to scheduling, setting cut-scores and establishing feedback loops to his teachers and principals.

**Data Reporting and Usage.** Like three of her counterparts, April indicated she reports her BOE scores from classroom assessments to her curriculum coordinator. However, she did not know what happened to the data from that point on, a pattern that emerged in four other teacher interviews. Barry indicated that Cottonwood has a data coordinator that disaggregates the assessment results and compiles proficiency report cards that are mailed to parents.

**Reflection about Student Learning.** When asked to describe ways in which they saw students’ science learning affected by the assessment plan adopted in Cottonwood district, both April and Barry skirted the question, indicating there is not a strong connection between the two events. April indicated there is a disconnection for students between classroom grades and BOE assessment tests. Barry discussed alignment of curriculum:

> It’s getting to the point now that our assessment system is completely intertwined with our guaranteed and viable curriculum, fully aligned, and now that that’s happened then our assessment system and our curriculum will drive instruction because of the data that’s given to teachers, and our professional learning communities that we have set up in our schools meet twice a week and they can make decisions with each other on instruction, based on our curriculum and assessment, which is our assessment system that we do in our district. So I’m thinking that this year, we’re kind of on
the dawn of actually moving from a good system to a great system that moves for kids, that adjusts for kids, so that we can do what the kids need to learn science better.

Cottonwood personnel refer to their curriculum as guaranteed and viable, meaning every student will receive the same content and assessments regardless of which building they attend. Teachers in Cottonwood district have developed common quarter tests which are administered at the same time each year and are aligned to the Wyoming State Science Standards. In terms of the BOE design principles, Cottonwood interprets consistency, fairness, comparability and credibility similarly to definitions from the Wyoming Assessment Handbook. At Cottonwood, interpretation of multiple measures differs from that of other schools in the study and building credibility did not appear to involve community members. Relative to consequences of the BOE system developed in Cottonwood, the district’s interpretation of this principle, as with other districts in the study, related to students needing to establish proficiency in five of nine content areas in order to graduate.

From this pair of participants, there was agreement on two topics. The majority of BOE support is perceived as coming from within the district at Cottonwood. Both the curriculum coordinator and the teacher interviewed from Cottonwood indicated a disconnection between the BOE assessment plan and student learning that takes place in the classrooms.

Cypress

Cypress is another of the largest districts in the state of Wyoming. Cypress has a student enrollment that is 12% minority, and 34% of students qualify for free and reduced
lunches. In comparison to Cottonwood, Cypress also does not show a two-year PAWS science-test score trend at the high school level. Their district science PAWS scores at the high school level are flat, lying in the middle third of the state rankings. However, in contrast to Cottonwood, Cypress utilizes BOE-consortium activities within their BOE-science assessment plan. Cypress’s curriculum coordinator “Irma” has been in her current position four years, and the science teacher interviewed “Harry” has been teaching in his current position for seven years. When interviewed, although both described their BOE system as utilizing common assessments developed by the Wyoming BOE Consortium, there was a difference in opinion about the course of study students would follow to earn graduation credits in science. Irma stated that all students take 9th grade science in common, which is physical science in this district, and then must earn two more credits of their choice for graduation. Harry said students must take physical science followed by biology and then had a choice for their third science. Either way, students are required to take three science credits to meet graduation requirements.

**BOE Characteristics.** Irma describes her district’s BOE plan as a course-based performance-assessment system. Teachers have worked together to identify common assessments, selected from the consortium activities list, and administered in common courses. The BOE Consortium list of activities is presented by titles in the BOE-survey; see Appendix A. Irma described the use of the consortium activity on scientific inquiry. She explained that science teachers use the assessment rubric that accompanies the activity, but they might use it at different times of the year with different science content. Harry described that he is most familiar with the assessments in physical science because
that is the group of teachers he meets with most often. Harry explained how the physical science teachers had to redevelop the consortium activity Density Differentiation because, “it had a quadratic function in it, and they could not analyze that, so we recreated a kinetic and potential energy, one that is much better suited to 9th grade skills.” I recall that the need to adapt consortium activities for difficulty level was mentioned by several science teachers from across the state when I hosted a BOE round-table discussion at the 2010 Wyoming Math & Science Teachers Conference.

In an analysis of the assessment activities within Cypress’s assessment plan, using the Baxter and Glaser (1998) assessment task framework, I categorized this blended plan as content-rich and process-open. Teachers administer common written assessments of content, and each semester students partake in a performance assessment using one of the consortium activities selected by teams of content teachers.

Alignment to State Standards. Both interviewees described curriculum work that was started four years ago. Irma’s description of work included teachers from the four schools in the district:

We first of all looked at the cognitive level of the standards then looked for a match of the cognitive level of the activities. We’ve done some different things. We’ve added curriculum to the 9th grade, for example, physical science, we didn’t have anyone at high school, 9-12, addressing any of the earth science content, so we added the earth science to our physical science, so we made adaptations to the curriculum.

Harry described this work within his curriculum team:

All the sciences got together, and we sat down and came-up with, what’d they call it, guaranteed curriculum basically, so we lined out, basically lesson plan by lesson plan, content pieces, piece by piece, we went through and determined what was important and we lined it out with a
timeline of how much time you should spend on this concept, and that took us about a week, and we did that for physical science, also 9th grade physics, 9th grade earth science, biology and chemistry. And that curriculum was adopted by the district, and every science teacher has access to it, and it is aligned directly to the state standards.

These quotations show both Irma and Harry taking part in processes to foster alignment. Although there was commonality in a sense that both Irma and Harry believe Cypress’s curriculum is now aligned to the state science standards, there seemed to be a difference of opinion in how much work was left to be done. Harry indicated an assurance that they are directly aligned with the state science curriculum, whereas Irma stated, “I don’t think that we are really well aligned, and I would say I think it’s because the standards are pretty vague.”

**Additional Wyoming BOE Design Principles.** Irma and Harry both believed that their district met the consistency principle, yet had different views when it came to discussing what aspects of their BOE science plan fostered consistency, with Irma focusing on scoring processes and Harry focusing on administering the assessment in one’s classroom. In her role as curriculum coordinator, Irma explained views related to the intended meaning of the consistency principle as described in the Wyoming Assessment Handbook:

We have had trainings on the rubrics. We have gotten all biology teachers, for example, together with common work and have them do some common scoring. We have done some generalization studies on those common scores, to see whether or not we are within something that would be considered statistically reliable, and we’ve asked teachers and schools to do common scoring within their schools so that we have some evidence of consistency.
Whereas, Harry identified consistency from his teacher perspective, which doesn’t address how different teachers would evaluate the same set of work from students:

> When we do the common assessments with kids, the science teachers are, we are all constructivists, so we are very used to the idea that kids are going to have questions and we are not going to give them [answers] so, it is the kid’s work, and I know that is understood at Cypress. The students are to come up with their own design, they’re to request the materials and make sure the scope of the experiment makes sense, and then they are to proceed with their experiment, line their own experiment out, so I think as far as consistency, that’s pretty much across the board at Cypress.

Two different lenses emerged again when probed about fairness in their plans. Harry described providing equal opportunities to all students, which is a subset of the fairness principle as described in the Wyoming Assessment Handbook. Irma described her district’s bias review checklist. The bias review checklist is a checklist marked off by assessment review teams whereby a district assures that the assessment may be completed by ALL students regardless of gender, ethnicity, economic status and/or disability status; that the assessment is appropriate for students with disabilities, acknowledging that students with IEPs receive appropriate accommodations; that the assessment is appropriate for Limited English Proficient (LEP) student, acknowledging that students may receive appropriate accommodations; and indicating that disaggregated data for this test has been examined for potential bias. This bias review checklist was originally posted by the Wyoming Department of Education as an example; it has since been modified by districts to meet their individual needs.

When we discussed comparability of assessments from year-to-year, both individuals described the same process of scoring common assessments by looking at student work and comparing it to previous anchor papers. Irma and Harry both indicated
a desire to have more professional development time to do this work, as currently time is
being cut out of daily planning time to do such work. Similarly, both described their
BOE science plan as containing multiple opportunities for students to demonstrate
proficiency on the science standards over the required three years of science instruction.
Harry’s preferred method is to “look at the growth over time and see if the two’s are
turning into three’s and the three’s into four’s.”

In our discussion about the credibility design principle, Irma recalled from the
perspective of developing the BOE plan, about ten years ago: teachers and students were
asked for input on the consortium activities, but she didn’t recall that parents or
community members were asked for input. Harry had a different interpretation:

Our building administrators mandated that we put our assessment
measures on the actual syllabus, so we had to all independently do that.
So we sat down together, as a group, and came up with what those were.
We have[it] on every syllabus in the science department, those kids know,
and the parents know, and have to sign off on it, it’s building policy, they
have to sign off on the fact that that’s how they are going to be assessed.

This depicts the differences between a district perspective and a classroom teacher’s
perspective. I don’t believe the teacher’s statements actually refer to developing the BOE
plan, as intended. The Wyoming Assessment Handbook provides a description of
credibility as including the invitation to stakeholder groups to participate in discussions
about the assessment system and to share student work with members of the public.

How are the effects of not meeting proficiency, addressed by the consequences
principle, perceived by these two individuals? Irma stated, “I don’t think there are any
consequences that kids, or teachers, feel go with that.” Harry’s perspective was
different:
I think there’s a lot of re-teaching that happens when a kid [fails an assessment], so if this is the common assessment, then there are no repercussions for them not doing well on it, it’s just another grade. However, that being said, the common assessment is a measure for teachers to see where they are at on that particular content, so I think some re-teaching does happen because of that and I think that the repercussion for the kid, the consequence for the kid, is that it affects their grade.

Both of these comments reflect a similar misconception across the state.

The Wyoming Assessment Handbook describes consequences as intended and unintended results of the implemented assessment system, supported by data collected that would indicate positive or negative results as a consequence of the assessments. This is often misinterpreted, as the term “consequences” does not indicate results as seen by a collection of data, but is more often interpreted as a punishment.

**Support for BOE process.** When asked about the development of Cypress’s BOE science plan, Irma described her leadership role:

My role at that time was to teach teachers about what the Body of Evidence was, to lead large groups of teachers to decisions about, first of all, what courses would be the Body of Evidence courses, and then, which activities would go into each of the courses. I’ve also been instrumental in some of the training as well as some of the tweaking that we’ve done, some of the formal guidelines that were more comprehensive than the guidelines for administration that were given in each of the consortium items, and then also, if we have made decisions to change the assessments, those have gone through me as well. My understanding of standards is that they cannot be wholly assessed through paper-pencil tests, that we needed performance assessments to be able to measure how well someone knows something and can do something. So, it was my belief that if we really were going to measure proficiency on standards, we had to have teachers giving performance assessments.
Irma reported receiving support for this process from her district leadership team made up of an associate superintendent of curriculum, an executive director of curriculum, and an assessment director. She did think that the state supported the process several years ago by hosting trainings in Douglas, but then it was not supported for awhile under a recent state superintendent. “As a district we never said we weren’t going to do it, we just plugged along with the work that we had done, and continued to require it of our teachers and our students.”

Harry recalls that the BOE science plan was already implemented before he started teaching. Harry is one of the two teachers interviewed who has not participated in the development of the overall science BOE design in his district, although he has a sense that “the overarching theme is that they want kids to have multiple opportunities to present evidence that they have learned.” Harry did not know if there was a feedback loop whereby teachers could improve assessments, but he reported feeling able to re-teach and re-assess whenever he needs to. Both Irma and Harry indicated they wished there was more professional development time available for teacher collaboration on common scoring of assessments.

Data Reporting and Usage. Harry sees BOE data reporting as a requirement, “It is required. We record our grades, proficiency scores, and common assessments.” However, Irma doesn’t think that the data collected for the Body of Evidence gets used “except by teachers to make their cumulative decisions,” which would indicate it serves a function. Irma described an end-of-year score that teachers report regarding whether each student has met required standards. Cypress has chosen to use a scale of one to
eight, one is basic and eight is advanced, with varying levels between. This was
confirmed by Harry. Both indicated that students and parents can see these scores by
logging onto their online grade book system.

Reflection about Student Learning. When asked to describe ways in which they
saw students’ science learning affected by the assessment plan adopted in Cypress School
District, both curriculum coordinator and teacher mentioned pessimism about the state
assessment test (PAWS) in their responses. Harry said:

So there’s been a lot more student responsibility added. And I wouldn’t
say that was just necessarily because of this [the BOE], but that’s
definitely improved, because it forced us to define how we were going to
test kids, and we agreed. When they first took scores, we scored 14 points
higher than [our sister school in the district]. We’re expecting basically
the same [this year]. This might give indication as to the importance of
essential curriculum and state standards in our respective high schools.

Irma responded:

I wish I could say that there has been, that we’re producing more
proficient, more capable students. I don’t see that. And I’m basing some
of that on our PAWS scores. At one high school, 60% are basic and
below basic, and we are pushing the inquiry method of teaching, so we
ought to be able to have higher scores than that I think.

Cypress is one of the three schools interviewed that employs a combination of
Wyoming Body of Evidence Consortium activities with district-developed tests. In
contrast to Cottonwood, which administers common assessments at about the same time
each year, Cypress teachers have the flexibility to integrate the Consortium activities
when they see that it fits into their curriculum. Cypress teachers meet in content teams to
align course content with the Wyoming State Science Standards. The teacher and the
curriculum directors shared common visions about how this work was accomplished; yet
differed on opinions about how much work was left to accomplish. There appears to be work left to do in communicating how to interpret consistency, fairness, and credibility principles, as this teacher did not seem to have the sense that was intended in the Wyoming Assessment Handbook. Although Cypress has curriculum team to work on BOE procedures, personnel voiced a wish that more time was available for teacher collaboration on common scoring of assessments, which are currently viewed as just another requirement of the job. However, in reflection it was noted that the BOE process has helped to focus curriculum alignment and inquiry methods of teaching.

**Larch**

Larch is a medium-sized school district in Wyoming, with K-12 enrollment around 4,000 students. Larch School District has 24% minority enrollment and 31% of the student population qualifies for free and reduced lunches. Larch High School services students in grades 10, 11 and 12. In contrast to the larger districts in the state, Larch has scored in the top 10 (of 48 school districts) on the high school PAWS science test during the testing years of 2008, 2009 and 2010. Larch’s curriculum coordinator “Jennifer” has been in her current position for six years and has a science teaching background. A science background was common in half of the curriculum coordinators interviewed. The science department head-teacher interviewed “Kelly” has been in her current position for nine years. Both ladies stated that they do not have a required course-of-study to meet graduation requirements in the sciences, but are commensurate with the Hathaway scholarship requirements. Students must complete three years of science in grades 9 through 12 to earn a high school diploma as specified under Wyoming Statute
21-9-101(b). Kelly explained that students must have taken prerequisites in order to enroll in chemistry or physics.

**BOE Characteristics.** Both interviewees indicated no reluctance in describing their BOE science assessment system as one in a state of flux. Disgruntlement with a misinterpreted requirement to use consortium activities, their science department has proposed to use course grades as indicators of proficiency. As Kelly explained:

Historically we worked with the consortium activities, and we worked with these and the rubrics for quite a few years. What we found was that they were not meeting the needs of students, teachers, science departments, and that the ratings of proficiency were sometimes inappropriate based on teacher knowledge. So we have recently proposed, and have gone to leaving consortium activities in place for our teachers to use as they deem appropriate and as they fit into their curriculum, but we no longer use the consortium activities to determine proficiency. But, if what we have here are highly qualified teachers, then doing well in a course, taught by a highly qualified teacher, based on state standards, and has alignment and common assessments, then we think that the grade in the class means something. What we’ve proposed in our district, is that, given that the courses are standards based, and aligned, that if a student has, and I know that this would be very controversial, that if a student has a C or better in three year-long courses, so three full-year courses in the science department, plus proficiency or better on the PAWS, then we would determine that they would be proficient.

At the present time it is hard to categorize Larch’s assessment plan due to the state of revision they are lobbying for. For the assessment plan on record, I categorized this plan as content-rich and process-open because they employed several consortium activities which are all projects geared toward specific content. However, it appears from these interviews that teachers at Larch want to move towards a grade-based system, in which case this may constrain the use of the consortium activities they currently do not favor.
Alignment to State Standards. Jennifer and Kelly both indicated that their assessments are aligned to the Wyoming State Science Standards; however, their replies about how that was accomplished indicated an edge-of-defiance about the process. From Jennifer, “You speak as if the state assessment system in the area of science, which is markedly different between all 48 districts, that’s there’s any generalizability in there.” Kelly added:

Our alignment process, for example, my curriculum in chemistry, what I’ve done with other teachers who teach the same course, is to go through the state standards and make sure that all the general science standards are hit and the benchmarks specifically for chemistry are included in our course. We assess those through projects, laboratories, common assessments, unit tests, semester exams. They are aligned to the state standards and then they would be aligned to each other as well.

Jennifer mentioned they do a lot of backwards design mapping in their district; however, Kelly was unsure what it meant to be aligned at the cognitive level. Apparently the same language is not used commonly throughout the district when it comes to curriculum work across the disciplines.

Additional Wyoming BOE Design Principles. Jennifer mentioned that science was the core subject BOE plan she submitted to the state for the BOE peer review process. When talking about the design principles, she proudly indicated that they passed the peer review process held in the spring of 2009. When asked to address the consistency principle, she explained:

Common assessments are embedded in common indicator courses, and rubrics are developed for brief and extended responses; assessment blueprints are included in the book, a sufficient number of assessments using multiple formats are included in the BOE, and anchor papers and
student exemplars are used to ensure consistent scoring. And ours did pass in that area.

From a teacher perspective, Kelly indicated that teachers in the same curriculum plan together, and they have common assessments in place for all their content areas. Kelly stated that they use inter-rater reliability on different assessments to meet the fairness criteria. Although she confused the two ideas of fairness with consistency, she did go on to explain, “I work pretty closely with the other chemistry teachers and we will look at a certain number of each other’s student responses.” Kelly also stated that teachers did this on their own time, which makes one question how committed this district is to the BOE plan since they do not reserve professional development time for such intensive processes. Kelly did not mention disaggregation of assessment results to examine possible bias, although that is what Jennifer focused on when asked about the fairness design principle.

Jennifer indicated her district puts a lot of stock in academic review when it comes to students having multiple opportunities at an assessment. “The principal, myself, a counselor and a content person get together and determine if students need another chance or if they had a bad day.” Kelly collaborated, “In terms of our assessments, we don’t do test retaking necessarily, some teachers may do that.”

Both women talked fluently about comparability of the common assessments in their BOE system. Jennifer elaborated with details:

Comparability. Teachers at all cites are required to give the same assessments in indicator courses. Directions for the administration of the assessments are developed and distributed to schools and teachers. Anchor papers are used as guidelines. The processes for collecting assessments, scoring, reporting are included in the directions. Teachers
from a variety of classrooms and schools participate in the scoring of open-ended assessment formats, which we are kind of getting rid of because it’s just too subjective. Across the years, we have things called vertical content teams, to evaluate and improve assessments each year, or as the group decides.

In addition to comments made earlier about common assessments, from her teacher perspective Kelly summarized:

The knowledge, the educational background students get from 2008 to 2009 should be comparable to this year and the year after that. It might not be exactly the same, and certainly when we go through changes, for example we’re considering a new textbook for the next year, then obviously we still have a course based on the state standards, but there may be some adjustments in how that material is presented.

Credibility is the design principle most likely to be ignored from year to year, as Jennifer corroborated when she said, “we don’t revisit that every single year.” Kelly laughed when asked about inviting stakeholders to the BOE process, “You know, that sounds really nice, but it’s not that practical. Teachers don’t have endless amounts of time.”

Jennifer had much to say about the consequences design principle as it relates to the BOE assessment system in her district:

There are no consequences. We built an instructionally supportive system with multiple ways of students being successful, and it is. The consequence is a design principle that I am most committed to. A plan that is prescribed by a state legislature should not be built to be punitive. Can students not be proficient in science and not graduate? We will make them proficient in five of nine content areas. We don’t prescribe the five content areas, so they may actually graduate without being proficient in science, or language arts, or one of those, so we’ve moved it from being a consequence based system to being a system that supported learning. We have the highest percentage of students advanced in the state, in the area of science at grade 11, and so the consequence of the Body of Evidence system is that they are doing very well.
Kelly summarized a similar understanding about the lack of consequences for not being proficient on BOE assessments, “In truth, the consequences only become glaring if a student is not eligible for graduation.” As is the case in other districts, students can graduate with proficiency in five of the nine content areas, so unless they are deficient in five areas, with science being one of them, a student can still graduate without being proficient on science BOE assessments.

**Support for BOE process.** As a curriculum leader within her district, Jennifer described her belief that the assessment system is teacher driven. In an ideal situation, teachers are provided with adequate time to focus on assessment and given professional development time to do so.

If teachers spend lots of time looking at standards and looking at alignment and looking at the activities within their classroom then you are going to actually improve student achievement. And I happen to believe that teachers with good staff development in lots of areas, including assessment, given the time to teach, are ultimately the only thing that will improve performance.

With that said, she then indicated that there has been a lack of support for the BOE initiative at her district level and her teachers find it cumbersome.

As a science department leader in her district, Kelly has been involved with the BOE process in her district from many angles. She has participated in scheduling, designing, setting cut scores, and establishing feedback loops to administrators. Kelly described her overarching philosophy of the system as content based.

What are best venues to introduce, work with materials, provide conceptual, the possibilities to provide for students to understand conceptually, and the assessments of course are a piece of that. How do
you assess knowledge? Through either projects, laboratories or summative assessments.

Throughout this process, however, Kelly stated that she has never received support from the state or district on this. “It doesn’t even come from the administrative level of the district, it’s up to the teachers in the department.”

**Data Reporting and Usage.** Kelly reported that teachers in her district report BOE data to Jennifer’s office. “Her administrative assistant has a massive spreadsheet that keeps the data that we give to her.” Jennifer concurred, but then expanded on her frustration with the data reporting:

We keep a district BOE assessment system. And we have a person, the teachers determine the results, they do the calibration workshops, they do that, and the results are reported to my assessment secretary who enters them into the Excel, extremely efficient, spreadsheet, and keeps them until the state calls for them. It is used to generate and determine the transcript endorsement that ultimately shows up on their transcript when they graduate. I have a high percentage of students who attend out of state colleges, they don’t care about proficiency, they don’t care. They care about GPA, ACT, graduation rate, and I called them when we started this Body of Evidence thing, there isn’t another school who cares and in my particular educational context, that is something I thought about. That is their reality, so I couldn’t get a professor to come to a Body of Evidence meeting because it has no meaning in their world, so the only thing we can use it for is instructional purpose.

**Reflection About Student Learning.** Throughout the interview, one could tell that Jennifer was not enamored with the Body of Evidence system, a theme that was also evident in her last response when asked to describe ways in which students’ science learning is affected by the assessment plan adopted in the district. Yet she also voiced
the viewpoint that the BOE process triggered some positive changes in the science curriculum.

I’m pretty sure it’s because we’ve built an instructionally supportive system. Kids don’t wait three weeks to fail, it doesn’t have anything to do with the Body of Evidence. I believe our assessment plan was the beginning of a discussion about what true student inquiry is. And our training on the inquiry model, about how we question, craft, and allow students to explore the world, that’s made the difference.

Jennifer is commenting on a positive spinoff of the mandated assessment plan. As a result of the BOE mandate, curriculum discussions were initiated that might not have happened otherwise. Not as a result of BOE assessments, but as a result of curriculum discussions, Larch High School now has a plan for catching failing kids earlier in the semester and placing them into recovery classes. This tenor is resonated by teachers within the district, as evidenced by Kelly’s individual response:

I don’t think that an assessment plan imposed by the state, or by an outside; like the consortium activities used to be. [It] is going to have an effect, simply because I think that the teachers who are in place and the work between teachers, alignment of course, revision of what’s happening in terms of classes, is going to be more important than an assessment system that’s enforced from outside. If you have highly qualified teachers who are intent on doing what’s best for students and student learning, then that’s going to happen without imposing an assessment from higher up.

Larch High School was one of the outlier districts interviewed, in that its BOE assessment plan was comprised of strictly consortium activities. However, disgruntlement with this plan has placed teachers in negotiations with administrators to change the BOE plan. Throughout the interview, a sense of high teacher commitment and low district commitment surfaced. However, understandings about intentions from the Wyoming Assessment Handbook were not always clearly communicated, a commitment
to student success was strongly evident and agreements between teacher and administrator occurred on several design principles such as alignment, comparability and consequences. Although they are not enamored with the BOE system, both educators indicated there have been some positive changes to the curriculums as a result of the BOE process.

Linden

Linden School District is a medium-sized school in Wyoming, with a district K-12 enrollment close to 2000 students. However, Linden is 23 times larger than the smallest district in the state and six times smaller than the largest school in the state. Linden School District has the second highest minority enrollment in the state with 30% of enrolled students claiming minority status. In contrast, they have only 16% free and reduced lunch enrollment, the lowest in the state. Linden High School science PAWS scores were in the top10 in the state during the testing years of 2008 and 2009, but fell to the middle of the rankings in 2010. Linden’s curriculum coordinator, Nick, has been in his current position only two years. The Earth science teacher interviewed, Marlene, has taught Earth Science at Linden High School for four years. Although it was hard to deduce a required course of study in high school science at Linden High School, both participants referred to the state law, chapter 31, which requires three science classes to graduate. The general sequence both talked about was Earth science for freshmen, biology for sophomores, and chemistry for juniors. Marlene described there being “different levels of each of those courses” offered to students, some “general” and some “college prep.”
BOE Characteristics. When asked to describe the district Body of Evidence system, Nick animatedly indicated they do not use consortium activities.

Fundamentally, we have a standards based system. We do not use a course-based system, we used a standards-based system. So 30% of those three science courses, Earth Science, Biology, and Chemistry, 30% of those course grades are specifically Body of Evidence assessments. First and foremost is student performance on standards-based science assessments covers 30% of their grade, their overall grade in those categories, but then their performance within that 30% category is pulled out and combined with their performance over the three different courses to determine final levels of proficiency with regards to state standards. We assess, at the benchmark level, for proficiency of the overall standard. We average the performance at the benchmark level, so it’s compensatory at the benchmark level, and conjunctive at the standard level, meaning they must be proficient on all standards. But in theory, that wouldn’t mandate their proficiency on each individual benchmark. But they do have to be proficient on a majority of the benchmarks, and they must be proficient on ALL of the standards.

When asked if the 30% of the BOE science assessments mentioned would be district developed, consortium activities, or a blend of both, Nick responded:

They’re definitely not consortium activities. Our district, and myself, do not feel like those consortium activities, even though we disagree with the state, obviously the state likes those, and I know certain districts like them, and I’m not calling them bad assessments, but we can’t for the life of ourselves figure out how those consortium activities can be used to satisfy what we think are 28 aspects of the Body of Evidence model. One example of that would be multiple opportunities of multiple assessments, so we can’t figure that out. So instead of using those, what we do is district-developed, and it’s really department-developed assessments. But we do not use the consortium activities.

Upon inquiry into whether Nick would say that the district-developed assessments are project-based, or tests, or a blend of both, he responded:

They tend to be less project-based, and just a traditional type of formal assessment. It’s not that there aren’t any projects involved in them, but there’s definitely a strong tilt towards that traditional, formal assessment strategy. They do include some projects, and performance-based, and
informal assessments, but if I had to categorize the science assessments in particular, you’re going to be looking at a tilt, a little more towards traditional, formal assessments.

Marlene did not seem as confident in her description as she described the BOE science assessment plan at Linden as a blend of both consortium activities and district-developed assessment. Upon probing into which consortium activities she uses in earth science she replied, “We use a grizzly bear standard that was developed by the state,” and then she went on to describe the activity. Note that in the Wyoming Body of Evidence Consortium activity bank there is not an activity about grizzly bears specifically, although there is one about endangered species. It was also confirmed by the principal and a math facilitator in Linden that no consortium activities are utilized in their BOE assessment plans. A misconception was evident when Marlene said, “Students are told how it is graded and everything, so I think that the state said that at least 30% of their grade is in BOEs.” It is unclear whether miscommunication that fostered this misconception occurred between the state and the district, or between the district and the teacher.

In an analysis of the assessment activities within Linden’s assessment plan, using the Baxter and Glaser (1998) assessment task framework I categorized this plan as content-rich and process-constrained. Teachers administer district-developed written assessment of content, and as indicated by the curriculum coordinator, they lean toward traditional tests.
Alignment to State Standards. Nick and Marlene exemplified how there can be different interpretations of education jargon. When asked to describe her understanding of the alignment process, Marlene replied:

My understanding would be that we assess over what we teach, and what we teach is what we assess over, so it’s like a two-way street. I think we looked at this with our BOEs, looking at the standards at their core requirements first, and then looking at our curriculum to see that we cover that, and then building the assessments from there. So it’s kind of like going from the standard requirements, or the benchmarks or whatever, to our curriculum, to make sure we cover it, and then from our curriculum to the assessments. And then making sure the assessment is actually testing the content from the standard.

Marlene also indicated that in this process, earth science classes were added in her school in order to cover state standards that were not being addressed at the high school level. This is an account that has been reported in other districts throughout the state over the last five years, such as was also described by Harry in Cypress School District. Nick’s description of alignment included a detailed explanation of two-way alignment, from his BOE experience:

Two-way alignment is the perfect question to exemplify, and I’ll try to stay as apolitical as possible; my experience is that no one even knows what two-way alignment is, and so, including the state department, and including during the peer reviews. I think they [the Wyoming State Department of Education] have a really weird definition in the Wyoming Assessment Handbook, definitely not a clear definition, full of interpretation. We saw dozens of different plans get criticized for their lack of two-way alignment, and really what it seemed to us, is it was just the genuine lack of understanding of what two-way alignment actually is. Going right out of the Wyoming Assessment Handbook, page 17, under two-way alignment, it says quote/unquote, “a finer grain matrix should be developed for each assessment to document the match between assessment items tasks and the various benchmarks assessed.” And that’s where, and we hit, really a home run, on two-way alignment, by assessing at the benchmark level. As soon as you assess at the benchmark level, you really hit, you’ve really kind of attained the highest level of two-way
alignment. Without assessing at the benchmark level, I think there’s an argument to be made that you do not have two-way alignment, because you can’t ensure, first, defining two-way alignment would be first, that all the standards are represented and assessed in the assessments and secondly, the second part of that two-way alignment is that every item, [every] element of the assessment [is] aligned to a standard, and by going down to the benchmark level, and only assessing benchmarks on these assessments, I think we’ve created the ultimate example of two-way alignment. To ensure the cognitive depth, every benchmark has to have a minimum of four assessment items or tasks associated with it, each representing one of the four different cognitive depths.

Probably due to the in-depth alignment process used in their BOE plan, this district was invited to present at the school improvement conference held in the fall of 2009.

**Additional Wyoming BOE Design Principles.** Nick makes the case for consistency within Linden’s BOE plan with this description of the three schools in his district:

[The assessments are] given at the same time, the exact same numbers of days of instruction, graded in the same manner. The actual assessment is given in standardized classroom settings, so yes, it’s same same. Now we do have an alternative high school here that uses comparable assessments; as per chapter 6, they’re not identical to Linden Hole High School, but they are comparable. We also have an online school that once again uses comparable assessments; they’re not identical to those at Linden High School, but they’re comparable. Within the school sites we have 100% consistency.

Marlene collaborated with her description that among the teachers they are giving the assessments in the same format within the same time frame and setting. Marlene also stated that in terms of fairness, “the tests or projects would be equally accessible to all of the students.” Marlene mentioned there being “no bias” in our discussion about comparability and Nick mentioned reducing bias in his response to data disaggregation methods. To check for comparability, Marlene said the teachers traded papers to grade,
more frequently during her first year, but less frequently now. Nick interprets the multiple measures design principle in this manner:

Obviously the Body of Evidence law, statute, is open to interpretation, and one of them is with that word multiple. So what we know is that multiple means at least two, so it’s somewhere between two and infinity, we’ve always said that you have to have that many opportunities, and so we’ve, our district, adopted a very conservative interpretation of multiple, and we hold onto two. So we have a primary assessment, and an alternative assessment, so each common assessment there are two versions of each one assessment, they are identical in their alignment to content and skills, it’s just different context. So an example would be the primary assessment maybe has $5 + 5 = 10$, the alternative assessment will have $7 + 7 = 14$, so same addition concept, but delivered in a different way to avoid that assessment summary impact on reliability and validity. So they have an opportunity to take two, and if they failed to prove proficient on the primary assessment, they have one calendar week to retake the alternate assessment to prove final proficient.

Marlene’s understanding is that students have three days to take the alternative assessment, but she concurred that they do get a try at the alternative assessment if they didn’t do well on the first one given by the teacher. When asked about the credibility design principle, Nick described his familiarity with the requirements as follows:

I think you’re really specifically talking about the standard setting chapter, chapter 5. We followed the assessment handbook verbatim. Psychometricians struggle with this, I think it’s the most technical aspect of the Body of Evidence system, and really one of the most unrealistic components of it because of the expertise it takes to actually define cut scores, and especially incorporating the entire community in that process. So what we did is we just looked right to the assessment handbook, and followed that process to the letter, had parent nights, had business members, community leaders, come to the table over a period of, actually over a calendar year, this started, began starting in about 2007 and went through 2009, and as per the handbook, got their feedback, their opinions, what they thought about cut-scores, and standards, and what was good enough and what wasn’t, and yeah, we had an extensive community wide effort over a calendar year in trying to capture the community feedback on that. And once again, as per that model, began creating the assessments, began giving the assessments, we used the data, student performance data,
on those assessments, to continually inform our cut scores and our standard setting process. It was probably the most time-intensive and complex, and sometimes frustrating part, of the entire process.

Marlene remembers that in the process “anyone could come to any of our meetings” but that only science department members worked on the assessments. In her understanding of the consequence design principle, Marlene said that if a student is not proficient on the standards in a course then students are assigned an incomplete and they are assigned to summer school to finish the course. In summer school they work on the standards in which they are not proficient. Nick skirted the question and discussed how they set cut-scores for proficiency. However, he did verify that it would be possible for a student at Linden High School to pass a science class without being proficient on the standards.

Support for BOE process. The sense from both Nick and Marlene was that they did a lot of BOE work on their own. Linden High School builds four professional development days into the school year for working on BOE assessments and data. Marlene reported that she participated in the design and scheduling of BOE assessments, but believed setting cut-scores was just handed down to them by the administration. One can sense the frustration in Marlene’s response to the question:

It seems like what we were told we were allowed, or not allowed to do, from the administration, that changed every once in a while, like every day, so, kind of inconsistencies with directions were really unclear, and I think that came from the state level perhaps because every time the administrations would go and ask whoever, the state department about, what we were allowed to do or shouldn’t do, etc., those answers changed all the time and therefore what was given to us changed all the time.

In Nick’s report, he commented that the support from the Wyoming State Department of Education got them about halfway where they needed to be:
We leaned on the state as much as we could, we worked with Dr. Moore, and Tom Collins, and any number of the consultants and whatnot. We went to all, we were constant members of the Body of Evidence information sessions provided by the state on a regional basis. We were part of the Body of Evidence institute. I think Fremont County hosted it, and a group of different stakeholders in the Body of Evidence process. So I think we probably maximized every bit of state support that was out there. But not surprising, that wasn’t enough. And so we did a lot of internal research and learning on our own about validity and reliability, and standard setting. So I’d say it was kind of a 50/50, I’d say the state got us about half way to where we needed to be via the assessment criteria, and the institutes, and the workshops, and such, but that didn’t get us all the way across the finish line without our community. So it’s kind of like 50/50; we used everything the state offered, and then just did a lot of work on our own.

Nick reported that his feedback loop includes reporting assessment results to the community while Marlene answered that she only discusses them within her science department.

Data Reporting and Usage. As in other districts, Linden School District utilizes electronic grade-book programs in which parents and students can log-on and download progress reports. Although Marlene didn’t know who was looking at the data, Nick was more confident that it was being utilized. Nick included a brief description of the multiple measures matrix which is used to correlate the reliability and validity of each score, such as ACT score to grade point average to PAWS scores to BOE scores.

The Body of Evidence data is just one of approximately a dozen data streams that we’ve defined as a school for data-driven decision making, organization, district. We use Pearson and Fusion as well as Power Teacher in this case, to basically capture and drill down into, analyze, synthesize, extrapolate, just like we have an entire data-driven decision making process, and one of those data stream paths would be the Body of Evidence. So yes, at the end of every semester, the teachers use Power Teacher to capture their Body of Evidence data, they look at it as departments, they look at it between departments, and we look at it as an
entire school. And that’s one of our professional development days. Like I said, we have four days that’s developed, dedicated, to the Body of Evidence. We have a day before school starts, we have a day after school ends, and we have two days embedded into the school year in which Body of Evidence assessments and that data are utilized.

The difference in Marlene and Nick’s perceptions is striking. Without follow-up visits to the district, it would be hard to guess about which view represents reality. It could be that Marlene took part in these sessions and yet these sessions may not have been meaningful enough for her to remember that they took place.

Reflection About Student Learning. Like her counterpart Kelly, from Larch High School, Marlene did not see the BOE having an impact on student learning. There seems to be a disjunction between how some teachers perceive the use of developing assessments and a connection to results they see through student achievement. On the other hand, although he did not discuss student achievement, Nick reflected that he thought the biggest impact he saw the BOE having, particularly within the science department, was that of consistency.

The validity of assessments is like this continual goal, or journey; whereas what we’ve done is attained an immediate high degree of reliability, because we have huge levels of consistency, so what we know is that from classroom to classroom, from teacher to teacher, within the science department, we have a tremendous amount of consistency now as a result of all the efforts to increase reliability within the Body of Evidence, and so I think we see more consistent student achievement. I’m not going to say it’s had a tremendous impact on the level of science student achievement, but I do think that it has been an equalizing force as far as creating a higher degree of consistency which relates to student achievement. But we don’t see our top science students necessarily doing; [or] achieving anything bigger and better than they did before. I think what we see is all science students attaining a higher degree of proficiency.
The interview participants with the shortest tenure in their current positions were located in Linden School District. Linden is one of two school districts investigated that developed their own assessments and did not employ the use of any consortium activities. Linden personnel mirrored the emerging trend seen in several districts in which teachers and administrators did not always have the same sense about what the intent of the BOE design principles was. This was seen as differing ideas were voiced about the origin of the assessments used in the plan and in the design principle of alignment. Educators in this district felt they received support for the BOE process; however, they admitted it was labor intensive and that they had to do a lot of the BOE work on their own time. Although neither, the teacher or the curriculum coordinator sees the BOE as having a direct impact on student learning, they do admit the process has led to a high degree of reliability and consistency in the implementation of the curriculum among classrooms in the district.

Willow

Willow School District is a medium-sized school in Wyoming, ten times smaller than the largest schools in the state. With 28% minority enrollment and 47% of students on free and reduced lunches, one might not notice this district when looking at PAWS data. However, although there is not a distinct trend in the PAWS data from Willow School District, they were in the top 10 in high school science proficiency rates during the 2009 testing year, and have not been in the bottom ten during the three years of PAWS science testing. The curriculum coordinator (Eunice) has been in her current position for six years. This is Eunice’s second career in education, as she qualified for
retirement after being an elementary science teacher for many years. The science teacher (Debra) has been in her current position for five years. Debra moved to Willow School District after retiring as a middle/high school science teacher from a nearby state. Combined, the two women have over 60 years of science teaching experience. As per state requirements, students graduating from Willow High School must have earned three science credits. The district’s requirements include physical science or earth science, biology, and then students may choose the third from chemistry (level I or II), physics (level I or II), advanced biology, or an anatomy/physiology class.

**BOE Characteristics.** The BOE science assessment plan in Willow High School consists of common semester tests and projects. The projects selected are consortium activities, such as Giants of Science, Carmaliticus (an imaginary organism), Endangered Species, Careers in Science, Let It Swing, and How Things Work. Although there are four teachers administering common assessments, they do the projects on their own timelines. Eunice described the underlying philosophy:

We like to focus on authentic learning, where students would be involved with direct results, like from a science activity they would do, like through the consortium. We do follow the protocol of Body of Evidence and making certain that we align with both the content standards and the performance standards, and also some of our activities do not access the advanced levels of the performance standards as well as they should and that’s going to be our target this year, to make certain that the levels of performance are evaluated appropriately.

Debra’s perspectives on the underlying principles that guide their planning are generally as broad, but still a work in progress.

I think that perhaps two things, we need to produce students who are college bound that need to have a good background, and be able to be
successful in their college career, but I think we also need to have students that can communicate well, read and understand things like newspaper articles so that they can make informed decisions whether they go to college or not. And I think that the idea that all of us have is that science should be a useable, a lifelong useable tool for them. That they should be able to, when they get done with our classes, be able to assess things in their own life. And not, it isn’t necessarily a rote memorization, it is a useable tool. I’m not sure we are there yet. But I think that, at least the four of us, the ideas that we have when we talk together. And honestly, Wyoming just does things so much differently than [where I was at] for 30 years by myself. So my whole idea of how I presented [to students] and was able to do things has changed a lot because I now have to comply with standards in a more rigid manner and I have to comply with what my other three peer teachers are doing. It’s an interesting switch. The philosophy of all of this I think is there, it’s just that getting to that point isn’t probably solidified yet.

Willow’s BOE science assessment plan includes two consortium activities per year that engage students in rich content and a project utilizing science inquiry skills.

These are used along with district-developed common semester tests. In an analysis of the assessment activities within Cottonwoods assessment plan, using the Baxter and Glaser (1998) assessment task framework I categorized this plan as content-rich and process-open.

Alignment to State Standards. As curriculum coordinator, Eunice works closely with the teachers in her district. It is evident, from the nearly ten minutes she talked about the alignment process, that it is one of the features of curriculum work she enjoys.

We’ve got two-way alignment in trying to mesh what the standards says with what the assessment is actually doing, and I believe that a two-way alignment [the] Body of Evidence is talking about [is] at the standards level. And then we even go a third step down to the benchmark level, to make sure that we have covered that, to make sure that the benchmarks align to the particular piece, and that the assessment piece is aligned to the benchmark. Once we have looked at the content alignment, then we try to match-up the level, and we’ve been using Norm Webb’s depth of
knowledge, although we’re moving to the new Bloom’s because some people are old Bloom’s people and never did care for Webb, although Webb allows for a lot more flexibility, but then what they [the teachers] do is, once they have their assessment design, once they’ve come up with the content alignment, then they go back and identify the level of complexity that they’re looking at. And it’s a real eye-opener to some of our teachers because some of them realize that some of this stuff that they’re assessing in is barely at the knowledge, comprehension level, which is some areas is very basic, and they never move to the proficiency, so that kind of opens up their eyes so that they go back and revisit the assessment to make certain that they’re looking at proficiency and advanced. If we look at the term curriculum, the curriculum is not WHAT, it is HOW. The standards are saying what, and when we look at curriculum, that’s more matched with the performance standards, the how piece. And so, we try to use those performance standards as to how the students are going to show their knowledge base, and that also moves into instructional strategies. So, what we would say is the standards are aligned to our curriculum, but our curriculum is really saying HOW all of this is going to take place, and HOW the students are going to, in the assessment part, in a balanced assessment system, how the students are going to show those performance levels.

Debra concurred that she had recently done some assessment alignment:

We have just gone through a process of taking every item on those standardized tests and going through a list of bias level of question whether it’s a total recall, it’s an inquiry, or it’s a critical thinking question [and] how to categorize that. Part of that is, I think understanding what you are asking the student, that makes it aligned with whether it’s part of the benchmarks or not. So we try to make sure that we’re assessing what the state wants us to assess. Basically I guess we are trying to align with the benchmarks. And then I think that the other part of that is the knowledge and skills that are represented through consortium activities, through those presentations that they give us, through, to papers that they write which would indicate a higher level of thinking and a higher level of knowledge and understanding rather than just recall. So I think we align with those two parts that way.

Additional Wyoming BOE Design Principles. Debra described her understanding of consistency as that of administering the same semester tests and using the same rubrics to grade consortium activities. Eunice described the statistical analysis she applies by
using Cronbach’s alpha to analyze the close-ended assessments for consistency. Eunice also described teachers using the same rubrics to score other assessments. Debra’s understanding of the fairness principle included allowing students on IEP’s (individual education plans for special education students) to receive accommodations such as extra time. By her description, one can sense Eunice’s affinity for fairness.

Fairness is pretty cool. We do a fine job on that. Once the assessment is written, then two other staff members, and they don’t have to be within the science department, read through the assessment, and carefully take a look at, if their close-ended questions, making certain that they are bias-free. We have a really nice little handbook our district has developed to use to follow that, and it’s kind of a checklist, looking at gender, looking at ethnicity, looking at use of words, looking at the level, of depth of the types of words that are used. If it’s an open-ended assessment, then they also take a look at the rubric, to make sure the rubric is clear, and not biased toward maybe the sharper student or whatever. Because the rubric is actually shared with the students so the students understand what they need to do to score on that.

Discussing comparability, the curriculum coordinator and teacher perspectives were slightly different from each other, each describing one strand from the design principle as described in the Wyoming Assessment Handbook on page 23. Debra addressed the common assessments and the awareness by teachers that they all need to present the same material, even though each teacher presents materials through individual strengths. Eunice described being able to compare data from year to year and not allowing staff to change more than 25% of assessment questions each year.

Eunice did not directly discuss the design principle addressing multiple measures, but inferred it was addressed, “What the staff does is they take a look at each one of the benchmarks in the standards and they design their particular assessment to cover, making certain that each of those benchmarks are covered a multiple of times.” Debra stated,
“There are alternate forms of the same test given,” making it hard to determine exactly how multiple measures are addressed in Willow’s assessment plan.

Both women agreed that stakeholders were invited to the process of BOE plan development, but that they did not see much input outside of the individual content departments. Eunice mentioned that they “got dinged” last year in the peer review for this design principle, so she scheduled more meetings and “invited lots of people in” to share the data with the community. “Actually, our outside resource stakeholders prefer that [we authenticate student learning via] basic knowledge, comprehension type of assessments.”

Eunice and Debra both addressed the knowledge that students can graduate by demonstrating proficiency in only five of nine content areas. Debra explained, “They can fail all the science standards and still graduate. They have to pass the science classes, but they could get D’s, not be proficient in science standards, and yet still graduate.” Eunice described that an additional opportunity to work on proficiency in standards is offered during summer school.

Support for BOE process. As a curriculum coordinator, Eunice has been very involved with every aspect of the BOE plans, from the development to setting cut-scores and establishing feedback loops, although she leaves the scheduling of the assessments up to the principals and teachers. Eunice admitted it’s been “pretty cool” to work on the science assessments since her background is in science and that it has been a “huge learning curve in the other areas.” Debra has participated in designing assessments and schedules when she administers the common assessments in her classroom. Debra’s
understanding of cut-scores is that they were already set within the school grading scale with 85% being advanced, 70% being proficient and 60% being basic. She feels the feedback loop is still in development, but she likes to share results with her students.

I keep some of those activities, in fact I keep all their tests and all their activities in a file for them for the year, and when we go over them, take a look at them at the end of the year, I say “Hey, look what you wrote, look at that lab that you wrote in September, and look at this lab that you wrote in May. Isn’t this, this is fantastic, look at the growth you’ve had.

Eunice confesses she has had obstacles:
I had a lot of obstacles, one of which happened to be the administration, because they felt that the BOE was a waste of time. What I did was I tapped into some of those very conscientious teachers who kind of got it, the idea, and then I found that my job became to try to make the system as painless as possible and have it make sense. Once it started to make sense to the teachers, then they got on board. The parts that didn’t make sense, I told them I would handle those. For example, the Cronbach’s alpha, we still haven’t figured out why that’s important, but (laugh) it does give us an idea of the fact that the students, that certain questions they frequently will get wrong. But it still doesn’t give us the answers that we want to get on consistency and what’s reliable, but we’re still working on it. We’re still trying to find ways. The handbook that Scott Marion developed, I think it’s a marvelous document. I think it’s extremely helpful, but, some of the areas I wish there were a better explanation for why certain things are done. The alignment piece is pretty clear, fairness is pretty clear, consistency is not so clear, more toward open-ended I can see, but the closed-ended it’s difficult for us to totally understand that piece, but all the criteria, it hits onto what we need to do for those kids, and it makes us a lot more accountable about the assessments we give our kids.

**Data Reporting and Usage.** Debra reports that her BOE results are turned over to Eunice’s office and that “all we really know from that is that we’re doing well.” Eunice admits that she could do a better job of disaggregating BOE assessment results because their large Hispanic population is a concern for her.

In disaggregating, we’re not finding a huge difference with our different subgroups. We’re finding a concern for special education, we’re finding a
concern with our ELL (English language learners) students. We’re finding our girls outperform our guys in some areas, like in the closed-ended assessments. In the open-ended assessments the guys outperform the girls, so we’re trying to use that disaggregated data to see if it makes sense, but it’s not making a lot of sense to us right now.

Debra believes students are given a report at least once a year on their standards-based progress. Not having a child in school, she was uncertain about the frequency of the reports. They use Infinite Campus as their grading program and parents have access to that online.

Reflection About Student Learning. As with responses from colleagues around the state, Debra sounded unconvinced that the BOE assessment plan was responsible for the student growth she sees throughout a school year.

Whether it’s due to just the curriculum we implement, maybe it’s just giving credit to those students themselves and their growth, I don’t know. I can’t even imagine how a first year teacher walks into this system because it’s tremendously harder [than where I was at previously]. I’m torn whether the accountability is necessary or not.

Eunice was not sure either, and she focused on teacher accountability in her reflection.

I don’t know if it’s the students at this point that are directly affected that much. What it has done, it’s given those teachers a heads-up to kind of dump the old traditional style of teaching and move into more authentic learning, and so, that’s where the students are beginning to reap the benefits. Where the kids are enjoying more of the hands-on activities instead of the old traditional lecture-type programs.

Educators interviewed in Willow School District seem much more at peace with the BOE process than the teachers in other districts investigated. It is interesting how different cultures appear to develop around the BOE in different districts. And although these educators don’t have all the answers yet, they don’t seem frustrated by that, but
instead seem to view it as a natural part of the process. As an articulate advocate for the BOE process, the curriculum coordinator in Willow School District appears to have accomplished the goal of making the system make sense to teachers in her district as there was agreement between coordinator and teacher on many design principles. The BOE plan in Willow School District consists of a blend of consortium activities and district-developed tests. Admittedly, one area for growth is in using the data to influence instruction. This is a pattern that can be identified in all districts interviewed.

Tesota

Tesota School District is located near the Wind River Indian Reservation, the seventh largest Indian reservation, by area, in the US. The Wind River Indian Reservation is shared by the Eastern Shoshone and Northern Arapaho tribes of Native Americans. Although the district student enrollment is closer to 1500 than it is to 2000, the district is considered a middle-sized district by Wyoming enrollment numbers. Twenty-eight percent of the enrollment is of minority status, making it the fourth highest minority enrollment of the school districts in the state. In contrast, Tesota is 17th on the list of 27 school districts in Wyoming with respect to students qualifying for free and reduced lunch, with 33% of its students qualifying. According to the 2010 census the median income for a household in the county is $32,503. Tesota High School science students have scored in the middle of the PAWS science results for three consecutive years. The curriculum coordinator, Gary, and the science teacher, Francis, are the most tenured of the professionals interviewed. Gary has held his current position for 11 years, and Francis has been a science teacher at Tesota High School for 14 years. Gary
indicated that most students take physical science as freshmen and then biology as sophomores. Students choose from several electives to complete the third science credit required for graduation. Francis agreed, although she said that some students skip physical science and start high school as freshmen in biology so that they can take more upper-level electives before graduation.

**BOE Characteristics.** Both participants described Tesota’s BOE science assessment system similarly, as a blend of both district-made assessments, mostly tests, and consortium activities embedded into science classes. The way this district calculates BOE proficiency scores is the most mathematical of the districts interviewed, as described by Gary:

Our BOE system starts collecting evidence at the 9th grade level in the science program, and that’s pretty much through the other eight areas also. There are courses in the science program, in the science department, that are identified as what we call indicator courses, and those are courses that we have designated as providing evidence to the high school graduation requirements for the Body of Evidence. Those indicator courses are science courses that we feel address the Wyoming standards and the level called for in the performance standards. So they tend to be around the 11th grade level of challenge. So, as an example, some of our AP courses, AP biology, AP environmental science, are not indicator courses because we feel they are above that 11th grade graduation standard level. We collect, within each of the indicator courses there are an identified set of assessments that every student who takes that course takes those assessments, they are assessments that are, if you will, embedded into the course content and those assessments serve the purpose of, they serve both the course grade purpose as well as serving the purpose for the Body of Evidence. Those sets of courses, those sets of assessments, pardon me, make up I think it’s 80% of the course grade. From that, and then there’s another 20 or 30 %, it’s either 70/30 or 80/20, from those courses then we, at the end of the course, at the end of each semester in each course, we harvest the % that the student has in that course. So, for an example, biology is an indicator course. The student would take first semester biology, and at the end of that first semester of biology they would have
taken those identified assessments, and that makes up, I think it’s 70%, of the course grade. The other 30% would be, or can be unique to the particular teacher. So then, the student at the end of the first semester would have a percentage, and let’s just say it’s a 94%, we harvest that out of the system, we apply that to a scale, and that percentage is then used to determine whether the student, what level of proficiency the student was performing at in that course, in that semester of that course. But because we want to give students multiple opportunities each semester counts, for the indicator course, so then we might somehow combine multiple semesters together, and then at the end of that the student will have a certain percent average and that’s what we use to determine whether they are proficient or advanced or not. That’s kind of a long rambling way of describing it.

Francis provided a similar, yet shorter, description:

One thing that we took a lot of time to create is common assessments. So if a kid is in my class taking biology, or Mr. X’s or Mrs. A’s biology, they all have the same assessments. And when we created these assessments, we looked at the standards to determine if they were aligned, the questions were aligned to standards and benchmarks, and also we made sure that we tried to get the same level of complexity that the standard was trying to address. And I bet you we spent four years refining our assessments to get to a point where we felt comfortable with them. The way it works in our grade-book is that 70% of a students’ grades are these common assessments, 30% of their grade in the class is all other assessments that a teacher gives, up to each teacher as to what they assign. Since they have 3 credits in science that they have to have to graduate, in our school, they have to have four semesters of proficiency in science. And proficiency is, I believe, is exemplified by three different measures. One is the PAWS, for juniors, another is what we call teacher judgment, and that’s a grade within a course, and for our cut-score to be proficient they have to have a 71% or better, in a course, and they need four semesters, or two full years of proficiency.

In an analysis of the assessment activities within Tesota’s assessment plan, using the Baxter and Glaser (1998) assessment task framework I categorized this plan as content-rich and process-constrained. Common assessments are all textbook administered tests of content, and the two consortium activities students partake in are both research reports.
Alignment to State Standards. Gary and Francis both talked about their alignment to state standards and benchmarks with both content features and cognitive complexity indicators. Because no prompting was necessary to elicit cognitive complexity from these participants, the sense is that they have a genuine feel for the two-way alignment process, as indicated by Francis in the above description. Gary continued talking about alignment:

We spent a lot of time on alignment, in terms of staff time in terms of groups and committees checking alignment. In addition to checking each item for fairness and bias, we also checked each item for an alignment to the standard and a check on cognitive complexity, you know, difficulty. Two-way alignment sounds complicated, and people get it really confused, but at least in my mind, it’s pretty straight forward. The first way of the two is to say, what does, take a particular test item and say what benchmark standard does this test item address? And so, for item number one, test question number one, which standard, which benchmark does it address. And I do that for every test question. So I’ve done one-way alignment I know going from my test question what standards and benchmarks are addressed. The second-way alignment, the second piece of two-way alignment, is to essentially go the reverse of that and say, for standard one, benchmark one, how many test questions do I have addressing that, and I just simply go back through my document notation and count them up. And then I go standard one, benchmark two, how many test questions do I have address that? And, at the end of the day, if I have a pretty solid system that has two-way alignment, every test question is aligned to a standard and benchmark and every standard is assessed with multiple items, many multiple items, and every benchmark, every benchmark doesn’t have to be assessed, but you have to have a pretty good sampling of all those benchmarks. And so then you can look back and say ok, here are three benchmarks that I never assessed at all, is that ok? And when you’re doing that second part, what you’re looking for is not really how many items were addressed to the benchmark, but another cut on that is, does the cognitive complexity match for a particular standard and benchmark? Let’s just say that the benchmark is cognitive complexity three, all my items were one, I don’t have a match, and I’m going to have to address that somehow. I think we have real good alignment in terms of coverage, I think we’re real solid. In terms of cognitive complexity we’re probably relatively weak. We probably have more items at a knowledge
level, recall level, and fewer at higher levels than probably the standards and benchmarks would imply we should.

From her teacher perspective, Francis reported:

When we were developing these common assessments, actually, after we developed the common assessments, and we looked at student data, that’s when we really started looking at fairness, looking at cognitive complexity, and matching up each and every question with a standard. That’s when it showed us how we needed to improve the assessments, and we spent a lot of time working with that. We’ve also spent quite a few years developing lessons using UBD, Understanding by Design, and putting that on a curriculum map so that everybody within the district could actually take a look at my lessons and see what standards were being addressed and whether it has the right level of complexity with the skills and assessments. [With] all the work that we’ve done, we’re pretty well aligned. The only thing that concerns me is that they’re going to change the standards and then we’re going to have our work cut out for us again.

Additional Wyoming BOE Design Principles. Gary and Francis varied in their reports about consistency of their assessments. Francis’s understanding is that they worked collaboratively for four years to obtain consistency in assessments so that they wouldn’t have to be changed again for several years. Gary described consistency as inter-rater reliability that should be done every year, but that “the amount of time available to do those in a real thorough manner just doesn’t exist.” One piece of information that probably contributes to consistency in an unreported sense is that Tesota High School has not had any teacher turnover in their science department in eight years. Gary feels they’ve done substantially better on fairness.

We have panels of teachers that we use, that we pull together who take the assessments and went through them item by item and compared the assessment items to a description of what’s fairness, and we did some training with those panels prior to that and so then we’ve done, pretty much every item, every assessment checked by those fairness panels,
bias review panels, and then we document all that. So I think we’re, at least at kind of a gross level, pretty good there.

Francis concurred:

We had a guide to check for fairness. Was it culturally biased, gender biased, was it using words that the student may not be familiar with? We have about 30% Native American, so we were really careful that we had gender and cultural biases erased from those assessments. So we did go through a process to check for fairness.

Both Gary and Francis openly displayed confidence in the comparability of their assessment system. Gary described:

There’s a comparability between teachers, comparability over time, comparability between schools, all that stuff, pretty solidly because our system is built on the idea that there’s common assessments regardless of the teacher, regardless of the year. And so, we’ve got those assessments identified through the kind of annual training with teachers to make sure they know what those are and have them readily available and all that kind of stuff. So teacher-to-teacher we’re pretty good at, pretty comparable there, we document that. Over time, the good news is we tend to stick with the same assessments which would make them comparable. The bad news is there’s a fairly, pretty dang tough to meet all those technical requirements, so there’s an incentive not to change our assessments. There’s kind of an incentive to say well that’s good enough we’ll just leave it there. So the assessments stay the same over time.

In his description of the multiple measures design principle, Gary referenced an idea Francis also mentioned, that of combining two years of scores to determine student proficiency.

We combine four semesters to come up with the final score. And if you take, just say you take six semesters of science, which you would, we will take your high four of those six, so you’re going to take multiple assessments within a semester and then you’re going to have multiple semesters, and then we combine them together, and we’re going to take the high four, high four of your six. And if you take eight, we’ll take the high four of your eight. The other piece is we have, not so much of a multiple measures, but a variety of measures in the sense that we have
both performance stuff, that’s lab based, as well as knowledge level tests.

The curriculum coordinator and teacher views of how credibility was established in the assessment system varied. Gary established that they put their system up for open public comment and review with a series of meetings. Francis recalled that it was just all teachers working on the assessments.

Although the school offers many opportunities for remediation through after-school tutoring, night school and summer school, both curriculum coordinator and teacher felt the consequences of not obtaining science proficiency were pretty minor. Gary described from his perspective:

It ultimately boils down to, if I’m not proficient in science, and if that’s the only one, or one of the few that I’m not proficient in, there really is minor consequences because the cut-scores are such that you can actually pass the course, in other words, the cut-scores that we have are not the same as letter-grade cut scores. So, you can actually pass the class, and not be proficient. You can earn your credits to graduate without being proficient. And that’s kind of OK, because you only have to be proficient in five of nine, and science is not one of those, science does not have to be one of the five. So the consequences of actually not being proficient in science, really is fairly minor. I mean, there may be a little, kind of “I don’t feel good about it,” but past that, there really are no substantial consequences.

Support for BOE process. The indication from Gary and Francis is that Tesota School District provided much support for the BOE development but recently there has been dwindling support beyond the implementation process. Both Gary and Francis have been involved in all stages of the BOE process, participating in design, scheduling, developing cut-scores and building feed-back loops. This is not surprising given their tenure within the district. Francis commented that they had PLC time to meet and have
critical discussions. Gary reported that some of the work time occurs during the summer and some during professional development days built into the school year, but that it’s all paid time. They provide an orientation at the start of each school year, for all teachers, to walk through the assessment system. When asked about what support systems were available when questions arose, Gary reported:

Way back when, the Department of Education, the Wyoming Department of Education had some pretty knowledgeable people that [we] would use them for a resource, and then as you are probably very aware, we didn’t have any resources for a couple of years, because there either were, [pause] the expectation for doing it changed remarkably for a couple of years and the people who were pretty knowledgeable in the process simply were not around. And then, we kind of figured it out on our own there for a while, and then recently, in the last couple of years, there has been some support from the department of education, but it’s still kind of a shadow of what it was eight years ago.

Data Reporting and Usage. Francis reported that teachers in her district report their BOE data to Gary’s office and his secretary inputs it into spreadsheets. Beyond that she indicated she did not know what was done with the data after that point.

You know, I do not know. I do know that they generate some pretty cool graphics (laugh) to hand to the kids and to the parents to show what courses they’re proficient in and how many more proficient they need to get in order to graduate. In fact I was just asking, today, our curriculum person, OK, I have all of this assessment information, all of their BOE stuff from last year, what do you want me to do with it? Because what we typically do is periodically we’ll go back through those, and we’ll look for an advanced, you know, kind of an anchor paper. Here’s my advanced, here’s my proficient, here’s my below proficient and then novice and then we kind of look at those, and kind of, how should I say, calibrate ourselves, so that we can stay consistent in our assessments.
Gary concurred that teachers send their results to his office. In addition he reports results to the state once a year. Gary shared an example from his district of a highly charged situation that occurred when disaggregated results were reported to students and parents.

A couple of years, probably four years ago now, (pause) the high school principal chose to have some charts made that showed disaggregated results. And our primary ethnic subgroup would be Native American, and so we had Native American results contrasted with white results and so forth. And it wasn’t just Native American and white, but all ethnic groups, but the two big ones for us are Native American and white. And he posted those, and that generated strong backlash, to the point of he opted to take them down. And they got (laugh) hidden away in a closet somewhere. So, even though all those disaggregated sub-results are available, and they’re on the website, if you make them, when we have prominently displayed them in the building so that students could see them, it doesn’t work. It didn’t work. It generated anger, it did not generate concern, or understanding. When it hit us was, they vandalized it. At least, from what you could kind-of see from the vandalism, they felt that by reporting disaggregated sub-groups performance it was disrespecting the groups that performed lower.

Reflection About Student Learning. Francis reflected that the BOE assessment system has focused teaching on the standards.

I’ve certainly seen a difference in the number of proficiencies, in other words, they’ve gone up. Mostly because we’re carefully making sure we’re teaching what we’re suppose to be teaching, and assessing students knowledge based on the standards we’re supposed to be assessing. One thing that we’ve found, there’s a bunch of stuff that we used to love to teach, but it didn’t match any of our standards, so it’s streamlined what it is we’re supposed to teach and it’s actually given us more time in order to teach those standards so that’s why I think we have greater proficiency. And we also, like I said, we’ve done a lot of work trying to make sure that our assessments are quality assessments that each kid gets in order to show their proficiency.
Gary feels that the process has moved science teachers to requiring more higher-level thinking from students by using more project-based assessments. When asked if he thought that would not have happened without the BOE plan he reflected:

I wouldn’t go that far. And here’s why I say. If we had not done BOE, then that would have given us, however many hours of professional development time, hours of staff work time, to have done something else. Who knows what that would have been. I mean, we’ve been doing BOE roughly now for ten years; if Scott Marion wouldn’t have come along with the idea of BOE we’d have done something else with all these hours these 10 years. And who knows what that might have been. So, the other thing that we might have done could have equally done the same thing, or maybe even better, we don’t know.

Tesota personnel are both comfortable and confident in explaining their understanding of the BOE plan in their district. There is a sense of multilevel communication as both described a genuine feel for two-way alignment and a similar train of thought in describing multiple measures across years. In response to the intricacy and workability of their BOE system, Gary was asked, by the Wyoming Department of Education, to present the approach he developed to setting cut scores at the 2009 fall school improvement conference. Tesota’s BOE plan is a blend of consortium activities and district-made assessments. It is also one of the exemplar plans posted on the Wyoming Department of Education website.

Although there was varied understanding about how often consistency measures, to ensure that assessments yield consistent decisions about student performance, and credibility measures, should be implemented, revisited, or revised, whether it is once a year or every four years, there was agreement that the amount of time to do these every year just doesn’t exist among all the other reporting that must occur to keep the system up
to date. As seen in other cases, teachers at Tesota lack an understanding of what happens to the BOE scores once they are reported to curriculum offices. However, there is a sense that the BOE system has focused teaching more on standards and assessment than would be done without its presence.

Cross-Case Analysis

Analysis and synthesis across the six cases was conducted to further examine the development and implementation of Body of Evidence science systems within the state of Wyoming. From this analysis, several themes emerged that are relevant to the questions of this study.

What design elements and implementation processes characterize BOE science assessment plans across Wyoming School Districts?

What support do teachers receive to develop a BOE science plan, align the plan to the state science education standards, and implement the intended science plan?

How do districts handle this opportunity to design and implement a locally useful assessment?

The first set of themes reveals the ways BOE assessments and curriculums are aligned to state science teaching standards. These six case studies revealed three distinct categories of BOE development, with associated strategies through which they meet the requirements of the BOE as set forth by the Wyoming State Legislative Act of 2001. The second set of themes depicts the nature of district support that is provided for the BOE development and implementation. Here evidence indicates an emerging continuum from limited support to comprehensive support. Finally, because BOE is an ongoing process,
a third theme was identified that discovered factors that affect the communication and support needed to sustain BOE implementation. A detailed description of each set of themes follows along with the research questions that the description attempts to answer.

**Domains of BOE Development and Implementation.** *What design elements and implementation processes characterize BOE science assessment plans across Wyoming school districts?* Across the districts, the great range of approaches in BOE plans was discovered to be not particularly coherent. However, the findings suggest that across districts three types of BOE assessment plans emerge: 1) plans that utilize a combination of Wyoming BOE consortium activities along with district-developed assessments; 2) plans that utilize only district-developed assessments; and 3) plans that are in a state of revision, with the eventual blend of activities still unknown.

**Table 18. Domains of BOE Development and Case Study Classifications.**

<table>
<thead>
<tr>
<th>Domain of BOE Development</th>
<th>Case-Study District</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOE plans that utilize a combination of Wyoming BOE consortium activities along with district-developed assessments.</td>
<td>Tesota Willow Cypress Cottonwood</td>
</tr>
<tr>
<td>BOE plans that utilize only district-developed assessments.</td>
<td>Linden</td>
</tr>
<tr>
<td>BOE plans that are in a state of revision.</td>
<td>Larch</td>
</tr>
</tbody>
</table>

No district interviewed used strictly Wyoming BOE consortium activities. Several comments were made by participants about those assessments taking too much time or being better teaching tools than assessment pieces. As evidenced in Table 16,
one district is in a state of revision with its BOE assessment plan. In spite of the variety of BOE plan designs, from interviews, I sensed that most participants have a foundational understanding of the intent of the BOE design principles. However, if the initial intentions of a state-wide accountability system were to establish comparability across district assessment plans, and thus equity for students across the state, then that vision drifts as plans become localized to reflect community values and practices.

**BOE Plans that Utilize a Combination of Wyoming BOE Consortium Activities along with District-Developed Assessments.** Gary and Francis both described Tesota’s BOE science assessment system similarly, as a blend of both district-made assessments, mostly tests, and consortium activities embedded into science classes. Francis stated they do two to four consortium activities a year, depending on the class. Gary’s description included:

> For the most part, those [district assessments] would be tests. Cognitive complexity probably would be lower on most of the district-developed stuff and where we needed performance-based stuff or projects we tended to use the consortium stuff.

The BOE science assessment plan in Willow High School consists of common semester tests and projects. The projects selected are consortium activities, such as Giants of Science, Carmaliticus, Endangered Species, Careers in Science, Let It Swing, and How Things Work. Then, the four teachers took a class and worked together to develop common unit plans under the mantra of *Understanding by Design*. Debra admitted the semester tests are primarily low on Bloom’s taxonomy level, “Primarily they are multiple choice but we do have some short answer and essay questions on the tests.”
Cypress High School also employs a BOE assessment plan utilizing selected consortium activities and district developed tests. Their most utilized consortium activity is #4: Scientific Inquiry, whereby all science teachers use the assessment rubric that accompanies the activity, but they might use it at different times of the year with different science content.

There did not seem to be a common theme among responses when asked about the philosophies toward the development of the BOE plans. Among the six individuals interviewed from these three districts, the responses varied from focusing on preparing students for college and developing usable life skills to being fair to all students, providing rigor and using performance assessments.

Cottonwood’s assessment plan includes district developed quarter tests. The format of these tests mirrors the format of the PAWS science test that students take as juniors. Although Barry stated they do not include consortium activities in their BOE assessment plan, what science teacher April described (below) sounds very much like two consortium-developed activities.

And then also once a semester, the students also have to do a performing district assessment, and with that, for example, in biology, they have to do a cell transport for an assessment, so it’s total inquiry-based, the students have to develop their entire lab, from start to finish, the teachers cannot help them, they can just provide them with materials, and so they have to do, they actually have to plan, conduct and analyze their experiments, write it up, and then teachers we have to grade them according to a rubric everyone in the district uses.

Further investigation into specifics of Cottonwood’s assessment activities reveal three consortium written assessments.
"BOE Plans that Utilize only District-Developed Assessments." Linden’s curriculum coordinator, Nick, was adamant in mentioning three times that they do not use consortium activities. Although, there seemed to be some confusion from Linden High School science teacher, Marlene, as she described “the grizzly bear standard” assessment piece as a consortium activity. It was then confirmed by a math coordinator at Linden High School that there are no consortium activities in their BOE assessment plan within any discipline. Perhaps in Marlene’s mind, any assessment piece that had collaborative input from teachers was a “consortium” of teachers. Marlene seemed unfamiliar with the concept of the Wyoming BOE Assessment Consortium. Upon inquiry into whether Nick would say that the district-developed assessments are project-based, or tests, or a blend of both, he responded:

They tend to be less project-based, and just a traditional type of formal assessment. It’s not that there aren’t any projects involved in them, but there’s definitely a strong tilt towards that traditional, formal assessment strategy. They do include some project, and performance-based, and informal assessments, but if I had to categorize the science assessments in particular, you’re going to be looking at a tilt, a little more towards traditional, formal assessments.

Both of these cases (Linden and Cottonwood) indicate a familiarity with the Wyoming BOE Consortium activity format, and the principle that assessments need to be more than tests alone. In rebutting the application of the consortium activities, there is a sense that local control is preserved and that they either 1) do not want to participate in a consortium of schools or 2) that there is a loss of
autonomy if they select assessment not solely developed by teachers in their district.

Although neither district provided an explanation for their differing paths and choices, there are advantages and disadvantages of such choices. One advantage is that producing district-developed assessments gives teachers opportunities to experience the process and intricacies of piloting assessments and collecting data to determine if an assessment is reliable, valid, and aligned to state standards. This process does not occur for all teachers in a district if a district chooses consortium activities which have already been piloted and checked for bias, reliability and validity. However, this can be a disadvantage for smaller schools that may not have enough students to do statistical checks with such small numbers. For small schools, consortium activities would provide reliance that chosen assessments already meet these criteria.

**BOE Plans that are in a State of Revision.** Kelly and Jennifer both indicated no reluctance in describing Larch’s BOE science assessment system as one in a state of flux. They want to go to a system of using classroom grades as the foundation for their BOE assessment plan. This is the feeling of their science department chair:

If what we have here are highly qualified teachers, then doing well in a course, taught by a highly qualified teacher, based on state standards, and [which] has alignment and common assessments, then we think that the grade in the class means something.

Although they used consortium activities for several years at Larch High School, their science teachers feel they take too much class time to implement, class time that could
otherwise be spent teaching content instead of assessing. As described by their curriculum coordinator:

> I know that particularly in the area of science, they wish to go to grade-based. First they have ensured that there is common, essential learning between courses that are similar, biology-biology, physics-physics, that kind of thing. Teachers do that work. Then, they would like to make the case that an A is advanced, and a B is proficient, that kind of thing. They would like to go to grade-based, I’ve kept them from using grades for the last four years so that they could standardize their practice and their expectations and essential learning. I know that is where they want to go.

Whether Larch is able to standardize this practice enough to pass the state-wide peer review process of the BOE assessment system is yet to be seen. Perhaps standardizing instruction for teachers across a district will be enough to meet the demanding criteria represented by the BOE design principles.

Passing the peer review process is not just an issue for this district, but for all districts. One of the disadvantages of the BOE peer review process is that as few as four hours may be spent by the reviewers using a five-page rubric (see Appendix D) to evaluate the complexities of a district’s BOE plan. Needless to say, all reviewers have to go on is what is written in the plan. Through this in-depth analysis I have observed variability in a number of aspects of BOE plans. In my experience with the peer-review process it is difficult to get a handle on the complexities of individual plans and to determine the degree to which plans are aligned with the required principles. There is a lot to synthesize in the small amount of time devoted to the peer-review meeting.

**Nature of District Support.** *What support do teachers receive to develop a BOE science plan, align the plan to the state science education standards, and implement the*
intended science plan? Regarding training and support provided during BOE
development and implementation the data reveals similarities and differences in the ways
in which participants feel they receive support for the BOE process. From survey results,
high school administrators indicated spending more time working on BOE science plans
than did other groups surveyed. Categories of support range from limited to
comprehensive with perceptions varying about whether the support is coming from the
district level, state level or both.

Table 19. Participant Engagement in BOE Development & Implementation Activities.

<table>
<thead>
<tr>
<th>Participant</th>
<th>BOE Activity Engagement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Participated in</td>
</tr>
<tr>
<td></td>
<td>schedule of BOE</td>
</tr>
<tr>
<td></td>
<td>assessments</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Curriculum</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Coordinators:</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Cottonwood</strong></td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Cypress</strong></td>
<td>No</td>
</tr>
<tr>
<td><strong>Larch</strong></td>
<td>No</td>
</tr>
</tbody>
</table>
### Table 19 continued

<table>
<thead>
<tr>
<th></th>
<th>No</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes, to community</th>
<th>state &amp; community</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linden</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>state &amp; community</td>
</tr>
<tr>
<td>Willow</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>state</td>
</tr>
<tr>
<td>Tesota</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes, to teachers</td>
<td>district &amp; state</td>
</tr>
</tbody>
</table>

**Teachers:**

<table>
<thead>
<tr>
<th></th>
<th>No</th>
<th>Yes</th>
<th>No</th>
<th>Yes, to science coordinator</th>
<th>science coordinator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cottonwood</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes, to science coordinator</td>
<td>science coordinator</td>
</tr>
<tr>
<td>Cypress</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes, to students</td>
<td>science department</td>
</tr>
<tr>
<td>Larch</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes, to administration</td>
<td>none</td>
</tr>
<tr>
<td>Linden</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes, within science department</td>
<td>science department</td>
</tr>
<tr>
<td>Willow</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes, to students</td>
<td>district</td>
</tr>
<tr>
<td>Tesota</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes, within science department</td>
<td>district</td>
</tr>
</tbody>
</table>

**Comprehensive Support.** April and Barry indicated Cottonwood school district fully supported the professional time it takes to design and implement a BOE plan.

Cottonwood district supports teachers in this endeavor by providing PLC time to work during the school year and pays teachers for their work time during organized after-school and summer meeting. Barry indicated teachers prefer the latter arrangement as they do not have to design lesson plans for substitute teachers via this schedule. Linden staff also stated they had PLC time for curricular work as well. Linden’s curriculum coordinator describes:
When we kind of redid our system a couple of years back, our district actually spent, collectively, over 30,000 professional development hours dedicated to Body of Evidence, development of these tests, fine tuning of them, the maintenance of the Body of Evidence system, disaggregation of data, inner-rater reliability, all of those various aspects. Now where we are at is we’re kind of on a maintenance track, we’re not creating or doing a lot of modifications, assessment creation design, we’re just kind-of maintaining the system; so what it looks like now is that the teachers tend to incorporate, they communicate with students and with parents about what the Body of Evidence system is.

Willow and Tesota school districts also support teachers with PLC work time on BOE-related activities. Both curriculum coordinators felt they received support from the state during these processes. At the same time, both teachers felt their district supported their efforts to participate in BOE building-level activities.

**Marginal Support.** Neither of Cypress’s interview participants indicated they had participated in scheduling the BOE assessments, indicating a lack of understanding of how this process developed. What struck me as I analyzed this case, was the different perspectives from a person who works at the district level, and who has no role in the implementation of the BOE plan and from a teacher who is responsible for not only delivering the classroom consortium assessments, but also making decisions about student proficiency in science. From her perspective, the curriculum coordinator indicated she felt supported from both the state and district level in fulfilling requirements of the BOE system. From his teacher perspective, Harry felt the teachers supported each other as they had to fulfill many of their BOE requirements on their own time.
Severely Limited Support. Both Jennifer and Kelly indicated that there has been a lack of support for the BOE initiative at Larch School District and the teachers find it a cumbersome process. Although they fulfill their obligatory responsibilities of administering and reporting assessment results, they do not feel designing and implementing this assessment system has improved student learning in science.

Communication Factors. How do districts handle the communication and support needs associated with BOE implementation? Although all participants indicated using some form of technology to collect, analyze, disaggregate and disseminate assessment results, no one interviewed sounded enthusiastic about this process. An emergent pattern in all districts studied was that it is compulsory to collect the assessment data and pass it on to the next level, whether the recipients are curriculum coordinators, data coordinators, administrators, or the Wyoming State Department of Education. No district described a comprehensive way of utilizing the data in discussions among teachers, administrators, parents and state department personnel. Only one district (Tesota) described posting the data publicly, and this resulted in backlash from a vulnerable population that felt this data should remain confidential. Using the data, to influence instruction, looks like an area of growth for every district in this study.

Communication obstacles appeared to fall into two categories: 1) between the Wyoming Department of Education and districts and 2) within districts. Participants from Cypress think the standards are pretty vague and can be interpreted in different ways. If a state wants districts to align curriculum and assessments to the standards, then the standards need to be as clear as possible about content and cognitive intentions.
The Wyoming Department of Education attempted to communicate intentions about BOE design principles when they published the Wyoming Assessment Handbook and supported face-to-face training sessions at different locations around the state. At the start of the reform there were people and resources to accomplish these trainings. Then, as state superintendents changed, priorities changed and people were reassigned positions or lost their jobs. This resulted in uneven perceptions of support about the BOE process. This tendency for reforms to be well-supported at first and then marginally supported later on, is not uncommon, but can be problematic. If the system is supposed to grow and change with the times, continue to push proficiency to higher levels, then having state-level support primarily during start-up is not enough. When there is buy-in for the BOE process, coordinators interpret the intent of the BOE design principles as intended in the assessment handbook and there is continuity of flow between development, implementation, and support. When teacher/coordinator agreement about intent of design principles occurs, one can deduce that communication within the district is functioning effectively.

However, a system breakdown is evident when what is required at one level has little significance at other levels. This is evident within several interviews in this study. None of the districts interviewed understood the consequences design principle as intended by the state department of education. Every district interviewed mentioned either a lack of consequences or consequences to students if they failed to meet proficiency in five of nine content areas. No district interpreted the consequences design principle with a positive spin. As described in the Wyoming Assessment Handbook, “the
district should evaluate whether the stated purposes are being served by the system, districts should also collect data to document that students are not being systematically denied opportunities to learn the standards and meet graduation requirements,” (p. 25-26). Perhaps consequence was the wrong term to designate this particular design principle.

Although there was varied understanding about how to address the credibility design principle, whether it should include community members or only school personnel, there was generally a lack of follow through seen on this principle. Every district interviewed reported taking care of this principle in the initial planning of the BOE, but no district mentioned annual displays of student work to the public for critiquing, or semiannual public review of the assessment pieces in the plan.

Larch’s curriculum coordinator eloquently articulates her view that this time-consuming and costly required reporting system is irrelevant to many students and to higher educational institutions who will work with the students next. The feeling of futility comes through clearly in interviews with educators in Larch. From their perspective, the system almost becomes “much ado about nothing,” unless those involved can give the process local meaning. In many cases, such as Willow and Tesota, the district is striving to give the process value by building better instruction through assessments. One positive spinoff of the mandated assessment plan appears to be that the BOE supports instruction through initiating curriculum discussions that might not happen otherwise.
Comparison of Findings Across Quantitative and Qualitative Methods

The results presented from both the survey and case study interviews indicated that there is a great variation in how this state-mandated assessment requirement is implemented and sustained across school districts in Wyoming.

There were three broad areas that were analyzed across all data sets in an attempt to answer the research questions in this study. The first was to identify what design elements and implementation processes characterize BOE science assessment plans. Results from the survey indicated most districts integrate assessments within classroom instruction and provide multiple opportunities on individual assessment pieces. These trends were supported by interview responses. Although, through interviews it was discovered that different districts interpreted *multiple opportunities* in two different ways: 1) students are provided with multiple chances to demonstrate understanding of the same content using different versions of an assessment; or 2) students are provided with repeated chances to succeed on the same assessment. Regarding the development of the assessments, it was initially discovered through the survey that a majority of Wyoming schools use consortium activities in their science assessment plans, as did four of the six districts interviewed. Regarding scoring of the assessments, from survey responses it was seen that more districts use a cut-score method of determining proficiency than use other methods; however, through interviews it was uncovered that this is not a standardized process and that there are a variety of methods used to set cut-scores. About half of survey responses indicated participating in setting cut scores; whereas eight of the dozen
interview participants partook in this endeavor. The survey revealed that nearly all respondents were familiar with their blueprint, normally in matrix form, showing assessment alignment to state standards, a fact supported through interviews, as all participants indicated their assessment plans are fully aligned to state science standards, and often mentioned their own involvement in working on this process.

The next broad area examined support received to develop, align, and implement the intended BOE science plan. Through interviews it was determined that there are several levels and types of support provided for the different BOE processes: support from state-level personnel, support from district-level resources, and support within teacher cadres. However, it was quite obvious that the perceived levels and categories of support are different both from district to district and within districts. This range was also evident in responses from the survey, and may have been indirectly reflected in the spread of responses regarding how many hours were spent working on BOE plans over the last three years. One method of support that appears under-utilized is learning about the BOE by participating in the peer-review process as a reviewer for another district, as less than half of survey respondents have participated in this reflective process.

The last broad area analyzed was how districts handle the communication and support needs associated with BOE implementation. One recurring theme was that collaboration needs more support. Although strong working relationships between teachers and curriculum coordinators within a district surfaced during certain interviews, there was not a sense that teachers in one district know what teachers in other districts do, and a similar sense was observed among curriculum coordinators. Collaboration time
among consortium schools has not occurred for some years, and lack of participation in peer reviews leaves schools in the dark about how other districts interpret design principles. Although all districts interviewed had someone devoted to managing the bulk of BOE data collect and reporting it to the next level, one wonders if the product (a single student proficiency score for each content area) justifies the number of person-hours it takes to sustain the BOE process. Certainly, there are some questions of the cost/benefit balance given the large – maybe even excessive – time spent in individual districts to meet the BOE demands. Is the BOE an optimal use of resources? And if so, how might the data gathered be processed and adapted to further help a district or school evaluate how it is doing?

**Summary of Findings**

This mixed-methods research study explored the patterns of development and implementation of Body of Evidence science assessment systems throughout the state of Wyoming. The survey and six case studies reveal that there are a variety of differences in BOE development and implementation plans across the Wyoming school districts. Each of the case study participants brought a unique context and perspective through which to view the Wyoming BOE assessment system. Districts follow different avenues toward fulfilling the BOE requirements as laid out by the Wyoming State Department of Education in the Wyoming Assessment Handbook. While the survey data established that there is a low perceived level of satisfaction with experiences in developing and implementing the BOE assessment system, the in-depth case studies provided more
insight into the intricacies of these perceptions. Social connections were a strong influence for all of the teachers interviewed. Opportunities to collaborate with colleagues contributed to the clarity of the assessment plan for teachers who are responsible for the implementation of the BOE plan. The interview data indicates that support within districts is district specific and levels of teacher engagement in the BOE processes are not reflective of the patterns of district support provided. The case studies also provided more insight into the processes of applying the design principles as called for in the Wyoming Assessment Handbook. In conclusion, the findings from this study reveal, from the perspective of curriculum coordinators and teachers, that this comprehensive assessment system is perceived as a time consuming and a laborious process; and that is does not necessarily provide commensurate benefits with respect to district, classroom, or student decision-making actions. From training to support, there are many facets to consider in sustainability of a comprehensive BOE assessment system that is perceived at the local level as truly useful; that adapts to changing conditions in the post-startup period; and that is sustainable given the lower levels of direct state support available over time.
CHAPTER 5

FINDINGS, CONCLUSIONS, AND IMPLICATIONS

Introduction

The purpose of this chapter is to discuss the implications of the findings and how the findings may be used to assist in improvement of instructionally supportive assessment systems, as well as recommendations for further research. The first section is a summary of the study that includes a review of the problem and the research questions. A brief overview of the literature review is included. The data collection methods are also reviewed in this section. The next section summarizes the findings of this study and includes a review of the statistical analysis and qualitative analysis. The conclusion section presents evidence to inform each of the research questions. The implications section addresses how this study can inform the work of science teachers, district curriculum coordinators, district administrators, and state assessment system coordinators. The final sections offer suggestions for future research and a look ahead at the future of the BOE system in Wyoming.

Review of the Study Context

Many Wyoming educators will recall the debate that occurred when Body of Evidence was first discussed, and then enacted as Wyoming law in 2001. Many factors lead to the requirement for Body of Evidence. As specified in statute W.S. 21-3-110(a)(xxiv) and Chapters 6 and 31 of the Current Rules and Regulations, each Wyoming
school district is required to develop and implement a Body of Evidence (BOE) system to
determine if students have met the Wyoming Content and Performance Standards and
Common Core of Skills. To receive a diploma, students must meet both the Carnegie unit
requirements and demonstrate proficiency on the standards. Wyoming is under strict
scrutiny from the Wyoming Supreme Court to assure that our students receive both an
adequate and equitable education. The Body of Evidence requirement exists in order to
assure that every child has equal access to the common core of knowledge and skills
found in state law and that a graduation diploma represents the same high standards from
district to district. Alternatives to Body of Evidence, such as a high stakes graduation
examination, have not been popular in Wyoming. The philosophy at the heart of the
Wyoming Body of Evidence system is to provide multiple measures to assess student
mastery of the content standards; in this way, no single assessment can disqualify a
student from graduation. Thus, the overarching purpose of this study is to explore the
patterns of development and implementation of BOE science systems throughout the
state of Wyoming. Although literature exists describing characteristics of high quality
assessment systems (Afflerbach, 2008; Baron, 1998; Bertenthal et al., 2008; Darling-
Hammond et al., 2008; Davies, 2007; Falk, 2000; Gallagher, 2007; Gorin, 2006; National
Research Council 2006; Paris, 1998; Pellegrino et al., 2001; Popham, 2003; Popham,
2008; Wiggins, 1998) very little research exists documenting such a system as it exists in
the real world. Therefore, this study also attempted to add to this limited research base.
Specifically, this mixed-methods study, utilizing survey and case study data, sought to
answer the following research questions:
1. What design elements and implementation processes characterize BOE science assessment plans across Wyoming school districts?

2. What support do teachers receive to develop a BOE science plan, align the plan to the state science education standards, and implement the intended science plan?

3. The BOE approach was originally envisioned by the 2000 WY state legislature as a way to provide districts with a measure of decision-making autonomy, participation and ownership within the structure of state-mandated assessment. How do districts handle this opportunity to design and implement a locally useful assessment?

The intent of this study was to examine these issues. The literature outlined stated that quality assessment includes four components: 1) attention to the developmental perspective of the student, 2) a match between instruction and assessment, 3) generation of quality evidence, and 4) management by instructors to allow appropriate feedback and follow-up (Wilson, 2009).

Findings of the Study

This section addresses each of the research questions posed in this study. The literature base that surrounds each question is presented when appropriate. The questions are also addressed in terms of which sources of data contributed to answering these questions.

Research Question 1

What design elements and implementation processes characterize BOE science assessment plans across Wyoming school districts?
a) What are the characteristics of the BOE systems being implemented today within and across districts?

b) How are the design principles presented in the Wyoming Assessment Handbook integrated into district BOE science plans?

c) What processes do different districts utilize to ensure that assessments are coordinated with state science standards and benchmarks?

Wyoming Department of Education personnel are trying to do as well as possible with student assessment by developing an alternative assessment system, the Body of Evidence. In an effort to guide districts toward meeting legislative mandates, while allowing flexibility for local decision making, the Wyoming Department of Education developed the *Wyoming Assessment Handbook*, originally published in 2001 with revisions made in 2003 and in 2008. Through the survey conducted for this study it was discovered that seven out of 10 schools integrate assessment within classroom instruction, two in 10 schools administer stand-alone assessments, and one in 10 choose a blend of the two strategies. These trends were supported by interview responses. Three-fourths of survey respondents indicated their BOE science assessment plan provides students with multiple opportunities on individual assessment pieces. However, through interviews it was discovered that different district interpreted *multiple opportunities*, as well as certain other design requirements presented in the *Wyoming Assessment Handbook*, in different ways. While these varied interpretations of key principles are not necessarily a negative circumstance, further sharing and consensus building among districts, or sense-making regarding the variations by the state, would be in order. When demands run high for local control there is much difficulty with implementation and consistency beyond multiple choice tests.
Regarding the development of the science assessment plan, addressed by sub-question b, it was initially discovered through the survey that a majority of Wyoming schools use between 1-13 consortium activities in their science assessment plans, as did four of the six districts interviewed. All interview participants were familiar with the majority of design principles from the Wyoming Assessment Handbook although, as noted above, their interpretations of key principles sometimes varied. As expected, the curriculum coordinators discussed more fluently explanations of how these principles are integrated into their assessment plans than did the teachers interviewed.

In reference to sub-question c, the survey revealed that nearly all respondents were familiar with their district’s BOE blueprint, normally in matrix form, showing assessment alignment to state standards. This fact was supported through interviews, as all participants indicated their assessment plans are fully aligned to state science standards, and often mentioned their own involvement in working on this process.

Referring back to the assessment triangle introduced in Chapter 2, it appears that the BOE assessment system has the potential to meet the three criteria of an effective assessment system: 1) be grounded in a theory about how people learn, 2) be used to elicit demonstrations of important knowledge and skills, and 3) be used to interpret evidence to draw meaningful inferences about what students know and can do (National Research Council, 2001). It is the degree to which each of these criteria is implemented that holds the power of the BOE. It appears that each district employs these criteria to different degrees within the BOE plans.
Based upon participants’ statements during interviews, it appears that three of Wilson’s principles for quality assessments (Wilson, 2009, p. 718), were met to a fair degree. These included: “(1) developmental perspective, (2) a match between instruction and assessment, and (3) the generating of quality evidence.” Wilson’s fourth principle, “management by instructors to allow appropriate feedback, feed forward and follow-up” appears to be less valued or more difficult to achieve. Without focused leadership regarding utilization of the data gathered, schools may be data rich and information poor.

Research Question 2

What support do teachers receive to develop a BOE science plan, align the plan to the state science education standards, and implement the intended science plan?

a) What support do teachers receive while developing, aligning and implementing district BOE science plans, and how does this vary across districts?

b) Is there a relationship between the extent or nature of the activities within the BOE science plan and the level of support teachers receive to develop, align, and implement the intended plan?

c) Is the support teachers perceive they receive in a given district connected to the support curriculum directors perceive they provide in developing and sustaining a district’s BOE plan?

From interview responses, teachers feel most supported in the various stages of the BOE processes by having professional time set aside for BOE work. These hours are frequently scheduled into the school calendar from year to year; however, the majority of teachers interviewed indicated there was never enough time set aside each year for the amount of BOE work required. From the survey responses, sixty-five percent of participants indicated spending less than seven hours per year working on BOE science plans. This seems strikingly low for a legislatively mandated assessment system that is
supposed to provide data upon which to make instructional decisions. In addition, during interviews, all curriculum coordinators indicated adequate professional development time was reserved for just such activities. This is contrasted with teacher interviews, indicating instructors are expected to perform such duties, as cross-score analysis, on their own time. One wonders where the reality between these discrepant viewpoints lies. In all interview cases, one gets the sense that each party feels they are doing the best they can with the time and resources provided and a sense of lack of support at some level or another is almost universal.

Results of a meta-analysis by Cimbricz (2002) contend that state-mandated testing influenced teachers and their work negatively. Although the BOE wasn’t accompanied by massive negative impacts as cited by Cimbricz, only about half of survey respondents wrote about benefits associated with the system such as improved comparability, improved consistency and increased teacher awareness about the standards to be addressed in classrooms.

Research Question 3

The BOE approach was originally envisioned by the 2000 WY state legislature as a way to provide districts with a measure of decision-making autonomy, participation and ownership within the structure of state-mandated assessment. How do districts handle this opportunity to design and implement a locally useful assessment?

a) What are the different ways districts utilize BOE data?

b) How do districts disaggregate assessment results?

c) In what ways do teachers, curriculum coordinators, and administrators see student learning taking place as a result of BOE assessment plans?
In Wyoming, school districts have been given an opportunity to design and implement a locally useful assessment system, a role mostly nonexistent in other states. The benefit is that they get some autonomy, but the disadvantage is that they have a lot of added responsibility – no doubt without a lot of extra resources. So how do districts handle this mixed blessing? Through interviews it was determined that each district has either a local data expert, whether it is someone with a background in data analysis, or a curriculum coordinator who has picked up the skills by being immersed in the BOE requirements. In every district interviewed, design fell to a team of teachers, most commonly headed by a curriculum coordinator. Implementation strategies varied slightly from district to district; common policies were assessments integrated within classroom instruction and stand-alone assessment tasks. Some district produce an assessment timeline and others allow teachers freedom to administer assessments as they fit into the curriculum.

The second component of Question 3 concerns how districts use BOE data. All district personnel interviewed indicated collecting BOE assessment data, but the uses of the data varied widely. One teacher from a large school indicated she just reported the results while another teacher in a similarly large school reported that “the data is always available to us. I don’t know that we have really gotten to the point where we are using the data to make decisions.” Although four teachers interviewed remembering using data initially to look at fairness policies in their BOE plans, no teacher indicated using the data on an ongoing basis for instructional improvement or knowing what the curriculum coordinators do with the data once it leaves the teacher’s grade book.
The third part of Question 3 concerns disaggregation of BOE assessment results. In the interviews, only the curriculum coordinators were able to describe disaggregation of assessment data at any depth, although none talked about it enthusiastically. As the curriculum coordinator in Linden High School said, “we’ll disaggregate data at the end of body of evidence assessment data, at the end of every semester, and we do it based on gender, age, ethnicity, special education, or not, and socioeconomic status, so you can see us disaggregating student performance data with an eye on fairness.” This was echoed by Willow High School’s curriculum coordinator:

> We disaggregate all of it, in every content area. And, but we don’t do a very good job of that. We kind of take a look at it and say, hey, because we have a large Hispanic population, and that’s one of our areas of concern. But we’re finding, in science, in disaggregating, we’re not finding a huge difference with our different subgroups. We’re finding a concern for special education [students], and we’re finding a concern with our ELL students.

Finally, the fourth part of Question 3 addresses the ways in which the BOE assessments affects student learning. From both survey responses and interviews there is a vast array of opinions about the student learning that takes place, or lack there-of, as a result of the BOE assessment system. Comments left by administrators in the survey include “The BOE has become more of an obstacle to learning in that it requires too many resources to develop, implement, and sustain,” and “Students and parents only care about the credits. It does not matter how many times you stress the BOE parents / students do not understand it,” to “Students are actively engaged in learning through experimentation.” Only three curriculum coordinators left a comment, “Students are applying their knowledge and not simply answering questions,” “Instruction is intentional and centered around essential learning curriculum and BOE,” and “There is
more awareness of and focus on teaching standards instead of chapters”. This is not universal however, some faculty are more concerned about this, some less.” At the end of the survey, teachers were more varied and expressive in their reflections, with 35 leaving positive comments, 25 negative comments, and 9 neutral comments. Positive comments include:

I see that integration of the assessments into instruction provides a great foundation for the students that aren't test takers to show their knowledge about a subject. In classes where there have typically been little in the way of labs, this has provided a hands-on lab experience for students through the implementation.

From another teacher:

For those students who take the assessments seriously, the results in the clarification of specific concepts have improved their understanding. Unfortunately, freshmen do not always understand the importance of finishing a task and doing it well. The BOE activities are used as supplements for our other assessments to indicate if they are proficient in some of the benchmarks. As always, there are some students who do better with the BOE type of assessment that do not do well on the usual assessments.

Negative comments regarding the impact of the BOE on student learning include:

Far too much time is taken up doing consortium activities. The suggested time frame is unrealistic (too short). Students must work on these projects at school. They are not allowed to take them home. My school does not have any type of study hall, thus the work must be done in the 52-minute chunk of time allowed for class. Each consortium activity takes an average of two weeks from start to finish. This, coupled with the PAWS, ACT prep, etc. takes away a huge amount of instructional time. I feel we are doing our students a grave disservice with all of the assessments that take place.

From another teacher, “Students see it as an annoyance and just try to get it done...they don't use it as a learning experience because they have encountered and assimilated the
displeased attitude and annoyance that the faculty have for the process.” A somewhat neutral stance was seen in this teacher’s reply:

The students learn how to design and run an experiment. They receive a lot of inquiry instruction during this time, but very little content for the course. It does not make sense to judge their proficiency based on a consortium activity with VERY little content included.

A mixed comment from another teacher:

Particularly in those classes taught in a semester block, the assessments really eat into instructional time. While I believe that the consortium assessments are valuable learning tools and they are integrated into our curriculum, I have a hard time justifying high stakes pieces for underclassmen.

There are problems that accompany almost all large-scale, standardized assessments and limit their usefulness in the classroom (National Research Council, 2001). Benefits of the BOE system include: 1) multiple opportunities which reduce the high-stakes pressure on students, and teachers, of a single assessment; 2) utilization of common assessments which increases the generalizibility of results across classrooms; and, in theory if not in practice 3) providing immediate information for teachers to use in planning lessons for the next day, next week, or next month.

Implications

High school reform has risen to the top of the education improvement agenda. Improving assessment policy requires the participation of all stakeholders, in particular: teachers, administrators, and policymakers. Although American education has been in the throes of reform for decades, building consensus around the need and direction for
change requires effective leadership, a shared understanding of the issues, and potential solutions. Areas for system-wide action include:

- Reviving policies to support improvement.
- Ensuring high quality work for all students.
- Eliminating the achievement gap.
- Increasing the knowledge and skills of high school teachers.
- Enlisting broad support for positive change.
- Finding, or reallocating, resources to implement and sustain necessary changes.

Designing a state-wide authentic assessment plan, which adheres to the recommendations in the literature, requires vision and commitment from all levels of the educational system.

**Recommendations for Policy Makers**

Spend some time in your constituent schools. Get to know their assessment and improvement plans. Become familiar with the consequences of high stakes testing. Many researchers (Glaser, 1991; Linn & Herman, 1997; Paris, 1998; Haertel, 1999; Pellegrino, 1999; Popham, 2001; Cimbricz, 2002; Jones, et. al., 2003; Hammerman, 2005, Gorin, 2006; Nichols & Berliner, 2008; Bertenthal, et. al., 2008) summarize concerns including reduced graduation rates, narrowing of the curriculum and neglect of performance skills. Teachers must be helped – not just commanded – to teach in new, different and better ways. Darlington-Hammond (1997) writes that in order for teachers to help students reach higher goals will require “most teachers to move far beyond what they themselves experienced as students” (p. 319). She emphasizes that teachers learn, just as students, “by studying, doing, and reflecting; by collaborating with other teachers; by looking closely at students and their work; and by sharing what they see” (p. 319).

This will mean supporting effective professional development that analyzes differences
between standards for student learning and standards for student performance of those standards. “If schools are to teach for understanding and hold all students to high standards, policies must support such teaching practice,” (Shannon & Bylsma, 2006, p. 137). However, keep in mind, that as noted by Stigler and Hiebert in the TIMMS 1999 Video Study, the teaching methods in each of the high-achieving countries looked quite different from each other. “In other words, it appears that there is not one way to teach effectively, but many,” (Stigler & Hiebert, 2009, p. 185).

There is normally a formal or informal cost-benefit analysis performed for all negotiations with state assessment contractors, and as this occurs policymakers should be asking themselves: What benefit is being served by spending an estimated 10 million dollars per year on the PAWS test design and administration? Could the educational system be better served by a less expensive test, as well as utilization of the extra capital to develop in-house capacity for data collection and analysis?

**Recommendations for State –Level Leadership**

The State of Wyoming, through its requirement of a Body of Evidence system, has created a need for teachers to gain assessment literacy and to see the connections between instruction and assessment. There have been some efforts in Wyoming to explain to teachers the BOE assessment system. There remains misunderstanding and confusion about the role these assessments play in determining graduation proficiency. One issue state-level leaders may need to consider is: To what extend do the variations documented here fall into line with the state’s original vision of the BOE? If these visions are not consistent, what can be done to realign the systems? A reviving of the
BOE consortium work would require training and support in using data for instructional decision making. There needs to be clarity and focus to help districts align the curriculum and local assessments with the state learning standards. Schools may require periodic training and continual support in using data for instructional decision making. Continue a strong level of support for the sustainability of the BOE plans, as follow-through must be as rigorous as the initial implementation support.

Reallocation of existing resources will undoubtedly be needed to implement major reform and improvements in high schools across the state. As pointed out by Shannon and Bylsma in 2006, sustaining improvement efforts require strategies that provide on-going feedback, opportunities for reflection and self-assessment, and effective analysis and use of data. In 2005 the Denver Commission on Secondary School Reform produced the report, *Not a Moment to Lose*, in which it outlined 21 attributes of a high quality high school. Four of these attributes dealt directly with assessments and data analysis of assessments. Expectations from the standards, assessment and graduation requirement section could have been taken directly from the Wyoming Assessment Handbook, as this section describes using a variety of assessments based on common criteria to measure student proficiency and demonstrate mastery, including projects, portfolios and presentations.

As specified in Chapter 31, Section 10(e)(iii), the BOE Peer Review will serve as the update of each district’s BOE documentation. The BOE systems were reviewed for accreditation in 2003 and again in 2009. Due to the turn-over rate in educational positions, maybe once every six years is not frequent enough to sustain the vision of the
BOE over time. This amount of lag-time may be too much for systems in a state of flux to receive useful feedback. A few minor adjustments could improve the integrity of the peer review process. Perhaps a rotational basis whereby different districts and different content areas submit plans for peer review could be established, thus removing any temptation for district’s to cherry pick their best two BOE plans for review as allowed under the current system. Working collaborations – led by exemplary curriculum coordinators, teachers and school leaders whose districts actively use BOE results to improve student learning – could be fostered across the state. These collaborators could reinvigorate the BOE development process across the state. This group could also help to form a pool of trained reviewers prepared to participate in BOE peer review across the state. If policymakers are concerned about holding districts accountable for the rigor of assessment plans, districts could be required to annually submit their BOE plans to the Wyoming Department of Education, or post them on their district websites for public viewing. Because the peer review process is lengthy, it would be sensible to continue evaluation of only two content areas at each peer review setting, although these could be randomly selected to encourage districts to do an excellent job on all nine plans.

This next recommendation could be taken on at the state or district level. Offer assessment literacy programs for all teachers and administrators (Popham, 2001). Make assessment literacy a goal for everyone involved in education, so that informed discussions can occur about the direction of assessment design.

The Wyoming Department of Education should be applauded for their systematic efforts of hosting state-level item, bias, and data reviews each summer. But now let’s
make the process a little more transparent and public reveal the findings of the meeting. Package the reports and submit the findings to all relevant constituencies, including the media, the educational committees of your state legislature, members of district school boards, and posted in archives that can be retrieved from the Wyoming Libraries Association (Popham, 2001).

Recommendations for District Level Coordinators

The National Science Teachers Association (NSTA) supports that teachers must have access to professional programs and activities that reinforce the assessment strategies advocated for students in the Standards. “Professional development should be integrated and coordinated with other initiatives in schools and embedded in curriculum, instruction, and assessment practices (NSTA, 2006, p.1).” Systems for collecting, analyzing, and using data in schools must be refined. One hope is that teachers will be given more opportunities to reflect on their experiences with common assessments in and within classrooms and to collaborate on developing and scoring common assessments.

School districts have a responsibility to raise expectations and provide coherent professional development to assist teachers in planning and improving their instruction. There appear to be discrepancies in knowledge levels and understanding of policies across the state. In order to require high quality work from students, educators need to reflect and examine the quality of work they require and the learning activities they provide. This cannot be accomplished in isolation. Professional learning communities should focus on this time-consuming work. As other studies have determined, planning is time-consuming and it never stops. When good programs are being implemented, they
are simultaneously being redesigned. (Loucks-Horsley et al., 2003). Shannon and
Bylsma (2006) postulate that positive professional learning communities build collective
responsibilities for student learning which results in higher student engagement and
motivation to learn. Teachers are students of instruction, they also need opportunities to
learn and grow in helping students to monitor, evaluate, and guide their own thinking.

Although there is a sense of skepticism about the value of the BOE to student
learning, Wyoming educators should embrace this responsibility to develop and
implement local assessments as part of a sanctioned statewide assessment system. This
opportunity is rarely provided to individual districts in other states. One thing that would
give the local BOE value is the use of collected data to inform curriculum decisions. By
looking at the whole picture of a BOE plan, districts should use the information, not only
for state proficiency reports, but for school improvement processes as well. Also, as
district leaders, make sure you are taking advantage of the support offered from the state
level. Support personnel to attend BOE workshops and peer-reviews. Build
collaborations; take advantage of your colleagues around the state and learn what is going
on in other districts. It is not a bad thing to seek support.

Districts are at the forefront to provide assessment literacy programs for parents
and community members, assuming first and foremost that your staff has been given the
tools to become assessment literate. There are a lot of parents who are willing to take
action to improve schools, but they too need to be educated about the issues in order to
speak knowledgeably and articulately about these assessment topics.
Recommendations for Teacher Leaders

A high percentage of teachers were involved in BOE development, yet they don’t perceive a connection to student learning in their classrooms. A lack of buy in from teachers in general surfaced in this study, and there are thus implications in this study for teacher leaders. One major implication is that the support for on-going collaboration among all teachers, including teacher leaders and their many colleagues, is essential to building and maintaining a rigorous assessment system. To teachers reading this manuscript, I would note that this ability to retain local control of a portion of the statewide assessment is vitally important. Even though BOE plans are mandated by the state, this should be a multi-directional process with local educators as well as the state contributing and reaping benefits. Teacher leaders need to move things in this direction, making sure we are not just going through the motions of sending the data to the next level and then forgetting it. Recognize that the data collected through the BOE, if carefully selected with teachers’ input, can comprise a rich data source that helps teachers to do a better job with their students.

Teachers need to be informed about the issues and the ways the data can be used in the classroom. Push for making the data useful; you are the professionals where the rubber meets the road. Perhaps the consortium activities need to be reviewed and revised every five years or so. This could be a task spearheaded by members of the Wyoming Science Teachers Association.

Read the Wyoming Assessment Handbook and see if you think your district assessment plan is aligned with the design principles of the BOE intent. If it is, great!
Please present at conferences and workshops, and tell others about the processes you utilize. And, if it’s not, take an opportunity to discuss with administrators in your district how to revitalize your assessments and what the next steps should be. Look at some exemplars on the Wyoming Department of Education website and investigate some processes other districts are using to build successful assessment plans.

As teacher leaders consider the kinds of assessments concerned educators should support they could use the knowledge of the Baxter and Glaser (1998) assessment continuum (referred to in Chapter 3) in order to promote assessments that are both rich in content and rich in scientific processes. Another way to advocate for meaningful assessments is to review the actual items on the PAWS test, or comparable assessments in your state. Become involved in your state’s test item review process, the bias review process, or the data review process. Check with your state science teachers’ organization or state education agency to learn how to become involved in these processes.

Suggestions for Further Research

Multiple opportunities for further study arose from this research project. Large-scale paper and pencils tests, typically in a multiple-choice format, have been used for most of the past century. According to existing research literature (Messick, 1994; Haertel, 1999; Popham, 2001; Cimbricz, 2002; Jones, Jones, and Hargrove, 2003; Gallagher, 2007; Darling-Hammond, et. al., 2008), the results of these tests do little to help teachers improve classroom instruction. A study that provides an in-depth understanding of the relationship between mandated assessment and instructional practice
in an actual school setting would add to the limited body of knowledge currently available. If a study could show that a state-mandated assessment system, consisting of engaging tasks, could change teaching practices, then it would add to the body of knowledge being explored by practitioners and researchers regarding the melding of curriculum, assessment and instruction. Another possible question for study would be to investigate if professional development opportunities offered about using the BOE Consortium Assessment Activities would cause more buy-in to using authentic assessment than district-led professional development.

Little research exists which investigates the relationship between student performance in science content on district-created assessments and their performance on state science assessments. Because districts have the luxury of designing and using a variety of assessments in the BOE plans to monitor the learning of standards, it is postulated that this will also guide the instructional process. While intriguing, these two investigations were beyond the scope of this study and leave open areas for investigation. In addition, Haertel (1999) postulates that tested proficiency fails to generalize across school disciplines and into real-world contexts. Future researchers could investigate the degree to which the proficiency levels established through BOE results can be used to predict student success across school curriculums or extend beyond high school.

Studies of Nebraska’s STARS system, where districts utilize their own assessment plans, show that their own assessment is not having a detrimental impact on students’ performance on tests designed to measure academic achievement relative to that of their national peers (Gallagher, 2007). Although it is fair to say these are positive
signs for assessment reform, the long-range sustainability remains a question. One suggestion would be an in-depth analysis of the amount of time, funding and other resources it takes to sustain innovative large-scale assessment program statewide. Such programs, which have benefits, may also be difficult to sustain without an ongoing input of resources. It would also be interesting to compare these issues between large schools and small schools, to see how the needs are similar and different. Also, to follow a whole district through sustainability for multiple years would yield interesting findings about the culture of this mandated educational reform effort. Further studies could seek to elucidate the complex dialectic between students, parents, teachers, administrators, and state department personnel that must occur for such a system to be successful.

Higher-order thinking and complex problem-solving are focal points in many educational reform efforts. While much is known about effective classroom assessment, psychometric practice is not rich in procedures for evaluating higher-order understanding, reasoning, and thinking in large-scale, high-stakes assessments. An important endeavor here is translation of accountability at the classroom level to reports that can be used at the state and/or national level. More research is needed between cognition theory research and large-scale assessment, and between psychometrics and classroom-based assessments.

Looking Ahead

As of June 2011, Wyoming joined the Smarter Balanced Assessment Consortium, one of two multistate consortia awarded funding from the U.S. Department of Education to develop an assessment system based on the new Common Core State Standards. As
the consortia work to develop a comprehensive accountability system that includes computer adaptive assessments and performance tasks, it is unclear where this will leave the future of the BOE assessment system in the state of Wyoming. In the 2011 Wyoming legislative session, the BOE was identified to be reexamined during the 2012-2013 school year. As of publication time, the future for this innovative assessment system is unclear.

Summary

This chapter reviewed the findings of this study and conclusions drawn were discussed. Implications for reforms in comprehensive statewide assessment systems in science were explored, and recommendations for policy makers, for state-level leadership, for district-level coordinators, and for teacher leaders were made. Finally, this chapter closed with suggestions for further study and a brief glance at the future of the BOE system.
REFERENCES CITED
REFERENCES CITED


APPENDICES
APPENDIX A

BOE ACTIVITY ALIGNMENT
Science Assessment Activity #1: Exploring Genetic Engineering – Where to Draw the Line

Standards and Benchmarks Science Assessment

An “A” in the table below indicates the standards and benchmarks in this assessment activity that have the potential to elicit evidence of student learning. An “I” indicates that instructional strategy that is assumed, but not assessed. An “A*” indicates the standards and benchmarks that are assessed only by the optional component. This activity has been recoded to the revised Wyoming 2003 Standards by members of the Wyoming Body of Evidence Activities Consortium.

11.1 CONCEPTS AND PROCESSES

In the context of unifying concepts and processes, students develop an understanding of scientific content through inquiry. Science is a dynamic process; concepts and content are best learned through inquiry and investigation.

UNIFYING CONCEPTS AND PROCESSES

Concepts in LIFE SYSTEMS and EARTH, SPACE, and PHYSICAL SYSTEMS are taught within the context of the following Unifying Concepts and Processes of Science:

- Systems, classification, order and organization
- Evidence, models, and explanations
- Change, constancy, and measurement
- Evolution and equilibrium
- Form and function

LIFE SYSTEMS Benchmarks

11.1.1 The Cell: Students explain the processes of life, which necessitates an understanding of relationship between structure and function of the cell and cellular differentiation. They identify activities taking place in an organism related to metabolic activities in cells, including growth, regulation, transport, and homeostasis. Students differentiate between asexual and sexual reproduction.

A 11.1.2 Molecular Basis of Heredity: Students demonstrate an understanding that organisms ensure species continuity by passing genetic information from parent to offspring. They utilize genetic information to make predictions about possible offspring. Students apply concepts of molecular biology (DNA and genes) to recent discoveries.

11.1.3 Biological Evolution: Students explain how species evolve over time. They understand that evolution is the consequence of various interactions, including the genetic variability of offspring due to mutation and recombination of genes, and the ensuing selection by the environment of those offspring better able to survive and leave additional offspring. Students discuss natural selection and that its evolutionary consequences provide a scientific explanation for the great diversity of organisms as evidenced by the fossil record. They examine how different species are related by descent from common ancestors. Students are able to explain how organisms are classified based on similarities that reflect their evolutionary relationships, with species being the most fundamental unit of classification.

11.1.4 Interdependence of Organisms: Students investigate the inter-relationships and interdependence of organisms, including the ecosystem concept, energy flow, competition for resources, and human effects on the environment.
### Science Assessment Activity #1: Exploring Genetic Engineering – Where to Draw the Line

| 11.1.5 | Matter, Energy, and Organization in Living Systems | Students describe the need of living systems for a continuous input of energy to maintain chemical and physical stability. They explain the flow of energy and organic matter through a series of tropic levels in living systems. Students investigate the distribution and abundance of organisms in ecosystems, which are limited by the availability of matter and energy and the ability of the system to recycle materials. |
| 11.1.6 | Behavior and Adaptation | Students examine behavior as the sum of responses of an organism to stimuli in its environment, which evolves through adaptation, increasing the potential for species survival. They identify adaptation as characteristics and behaviors of an organism that enhance the chance for survival and reproductive success in a particular environment. |

### EARTH, SPACE, AND PHYSICAL SYSTEMS Benchmarks

| 11.1.7 | Geochronal Cycles | Students describe the Earth as a closed system and demonstrate a conceptual understanding of the following systems: geosphere, hydrosphere, atmosphere, and biosphere. Students explain the role of energy in each of these systems, such as weather patterns, global climate, weathering, and plate tectonics. |
| 11.1.8 | Origin and Evolution of the Earth Systems | Students investigate geologic time through comparing rock sequences, the fossil record, and decay rates of radioactive isotopes. |
| 11.1.9 | Origin and Evolution of the Universe | Students examine evidence for the Big Bang Theory and recognize the immense time scale involved in comparison to human-perceived time. They describe the process of star and planet formation, planetary and stellar evolution including the fusion process, element formation, and dispersion. |
| 11.1.10 | Structure and Properties of Matter | Students describe the atomic structure of matter including subatomic particles, their properties, and interactions. They recognize that elements are organized into groups in the periodic table based on their outermost electrons and these groups have similar properties. They explain chemical bonding in terms of the transfer or sharing of electrons between atoms. Students describe physical states of matter and phase changes. Students differentiate between chemical and physical properties, and chemical and physical changes. |
| 11.1.11 | Chemical Reactions | Students recognize that chemical reactions take place all around us. They realize that chemical reactions may release or consume energy, occur at different rates, and result in the formation of different substances. They identify the factors that affect reaction rates. |
| 11.1.12 | Conservation of Energy and Increase in Disorder | Students demonstrate an understanding of the laws of conservation of mass and energy within the context of physical and chemical changes. They realize the tendency for systems to increase in disorder without an input of energy. |
| 11.1.13 | Energy and Matter | Students demonstrate an understanding of types of energy, energy transfer and transformations, and the relationship between energy and matter. |
| 11.1.14 | Force and Motion | Students develop a conceptual understanding of Newton’s Laws of Motion, gravity, electricity, and magnetism. |
Science Assessment Activity #1: Exploring Genetic Engineering – Where to Draw the Line

11.2 SCIENCE AS INQUIRY

Students demonstrate knowledge, skills, and habits of mind necessary to safely perform scientific inquiry. Inquiry is the foundation for the development of content, teaching students the use of processes of science that enable them to construct and develop their own knowledge. Inquiry requires appropriate field, classroom, and laboratory experiences with suitable facilities and equipment.

<table>
<thead>
<tr>
<th>SCIENCE AS INQUIRY Benchmarks</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A 11.2.1</strong> Students research scientific information and present findings through appropriate means.</td>
</tr>
<tr>
<td><strong>A 11.2.2</strong> Students use inquiry to conduct scientific investigations.</td>
</tr>
<tr>
<td>• Pose problems and identify questions and concepts to design and conduct an investigation.</td>
</tr>
<tr>
<td>• Collect, organize, and analyze and appropriately represent data.</td>
</tr>
<tr>
<td>• Give priority to evidence in drawing conclusions and making connections to scientific concepts.</td>
</tr>
<tr>
<td>• Clearly and accurately communicate the result of the investigation.</td>
</tr>
<tr>
<td><strong>A 11.2.3</strong> Students clearly and accurately communicate the result of their own work as well as information from other sources.</td>
</tr>
<tr>
<td><strong>A 11.2.4</strong> Students investigate the relationships between science and technology and the role of technological design in meeting human needs.</td>
</tr>
<tr>
<td><strong>A 11.2.5</strong> Students properly use appropriate scientific and safety equipment, recognize hazards and safety symbols, and observe standard safety procedures.</td>
</tr>
</tbody>
</table>

11.3 HISTORY AND NATURE OF SCIENCE IN PERSONAL AND SOCIAL DECISIONS

Students recognize the nature of science, its history, and its connections to personal, social, economic, and political decisions. Historically, scientific events have had significant impacts on our cultural heritage.

<table>
<thead>
<tr>
<th>HISTORY AND NATURE OF SCIENCE Benchmarks</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A 11.3.1</strong> Students examine the nature and history of science.</td>
</tr>
<tr>
<td>• As scientific knowledge evolves, it impacts personal, social, economic, and political decisions.</td>
</tr>
<tr>
<td>• The historical misuse of scientific information to make personal, social, economic, and political decisions.</td>
</tr>
<tr>
<td><strong>A 11.3.2</strong> Students examine how scientific information is used to make decisions.</td>
</tr>
<tr>
<td>• Interdisciplinary connections of the sciences and connections to other subject areas and career opportunities.</td>
</tr>
<tr>
<td>• The role of science in solving personal, local, national, and global problems.</td>
</tr>
<tr>
<td>• The origins, limitations, and conservation of natural resources, including Wyoming examples.</td>
</tr>
</tbody>
</table>
APPENDIX B

BOE SURVEY
BOE Survey

1. Introduction

Dear Participants,
First of all, THANK YOU for your contribution in sharing your experience, ideas, and opinions by responding to this survey. This is a study about the development of BOE science assessment plans in Wyoming. I am requesting your help to make this descriptive research effort a success. Please note that your response is vital for my research.

It takes about 15 minutes to finish this questionnaire. If you are over 18 years old and you are willing to complete this survey, mark your choices after careful consideration. Your responses are strictly confidential and anonymous. Your completion of this questionnaire indicates your personal willingness and implies your consent to use your responses for the research purpose.

By clicking on NEXT and providing your email address below, you agree to participate in the survey and for data to be collected. Thank you again, in advance, for your help.

Sincerely
Shara Ensminger
dowdeng@wooton1.k12.wy.us
Doctoral Candidate, Science Curriculum & Instruction
Montana State University

demographics

1. My primary role is:
   ○ high school science teacher
   ○ district curriculum coordinator
   ○ high school administrator
   ○ district administrator
   ○ Other (please specify) __________________________

2. Number of years in current position:
   ○ 0-5 years
   ○ 6-11 years
   ○ 12-20 years
   ○ more than 20 years

3. Have you participated in a BOE peer-review at the state level?
   ○ Yes
   ○ No

4. Have you participated in developing a BOE science assessment plan in your current district position?
   ○ Yes
   ○ No
**BOE Survey**

3. **page 3**

1. Are you familiar with your district science assessment plan blue-print (sometimes called a matrix) of alignment to state science standards?
   - Yes
   - No

2. Are your BOE science assessments integrated as assessments within classroom instruction, or administered as stand-alone assessment pieces?
   - Integrated as assessments within classroom instruction.
   - Administered as stand-alone assessment pieces.
   - Other (please specify)

3. Does your BOE science assessment plan provide multiple opportunities for individual assessment pieces?
   - Yes
   - No
   - I don’t know
BOE Survey

4. Does your BOE science assessment plan include any consortium activities? If so, please check all that apply:

☐ #1: Genetic Engineering
☐ #2: Fast Plants
☐ #3: Don’t Turn Out the Lights
☐ #4: Scientific Inquiry
☐ #5: Giants of Science
☐ #6: Ontology Differentiation
☐ #7: Dormilicious
☐ #8: Endangered Species
☐ #9: Careers in Science
☐ #10: In the Bag
☐ #11: WY Plants
☐ #12: Many Sides to an Important Story
☐ #13: The Great Race
☐ #14: Micro organisms, Small but Important
☐ #15: Let It Swing
☐ #16: Trolleys to Hybrid
☐ #17: It’s In the Dirt
☐ #18: Connecting Science Through Time
☐ #19: The Power of an M&M
☐ #20: How Things Work
☐ #21: Rodent Roadsters
☐ #22: An Analogy of Cells to Organized Systems
☐ #23: Miracle Drugs vs. the Super Germs
☐ #24: A Big Hunka Hunka Lithosphere
☐ #25: Field of Dreams
☐ #26: Survival at the Moon Colony
☐ #27: Bigbottle
☐ #28: Children of the Corn

☐ Our district plan does not include any of these consortium activities.

☐ I am not familiar enough with our district plan to answer this question.
5. Does your BOE science assessment plan provide multiple opportunities for individual assessment pieces selected above? If so, for which ones?

☐ #1: Genetic Engineering
☐ #2: Fast Plants
☐ #3: Don't Turn Out the Lights
☐ #4: Scientific Inquiry
☐ #5: Giants of Science
☐ #6: Density Differentiation
☐ #7: Carrotlicious
☐ #8: Endangered Species
☐ #9: Careers in Science
☐ #10: In the Bag
☐ #11: WY Plants
☐ #12: Many Sides to an Important Story
☐ #13: The Great Race
☐ #14: Micro-organisms, Small but Important
☐ #15: Let It Swing
☐ #16: Evolving to Hybrid
☐ #17: 1% in the Diet
☐ #18: Connecting Science Through Time
☐ #19: The Power of an M&M
☐ #20: How Things Work
☐ #21: Rodent Roadsters
☐ #22: An Analogy of Cells to Organized Systems
☐ #23: Miracle Drugs vs. The Super Germs
☐ #24: A Big Hunka Hunka Lithosphers
☐ #25: Field of Dreams
☐ #26: Survival at the Moon Colony
☐ #27: Biobattle
☐ #28: Children of the Corn

☐ Our plan does not include any of these consortium activities.
☐ I am not familiar enough with our district plan to answer this question.
6. What criteria for proficiency does your BOE science plan adopt, standard mastery or cut scores?

- Standard Mastery
- Cut Scores
- I'm not sure
- Other (please specify)

7. Have you participated in setting cut-scores for the science assessment plan in your current district position?

- Yes
- No
BOE Survey

4. page 4 - PD experiences

1. In the last three years, approximately how many hours have you spent working on your BOE science assessment plan?
   - 0-8 hours
   - 9-16 hours
   - 17-21 hours
   - more than 21 hours

2. Please answer the following questions as displeased, somewhat displeased, neutral, pleased, extremely pleased.

<table>
<thead>
<tr>
<th>Question</th>
<th>Displeased</th>
<th>Somewhat displeased</th>
<th>Neutral</th>
<th>Pleased</th>
<th>Extremely pleased</th>
</tr>
</thead>
<tbody>
<tr>
<td>As a professional development experience, how do you rate your experience developing the BOE plan in your district?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How do you rate your experience with the implementation of the BOE plan in your district?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What is your opinion of the impact on student learning that is taking place in your classrooms as a result of the BOE assessment plan your district has adopted?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. In what ways do you see student learning taking place in your classroom, as a result of the BOE assessment plan adopted by your district?

   [Blank space for response]

4. THANK YOU again for your time! Would you consent to an interview about the development of the BOE plan in your district? If so, please leave your email address, and I will email you to set up a time for an interview.
   - Yes
   - No

   If yes, please provide your email address here:
APPENDIX C
INTERVIEW QUESTIONS
Curriculum Coordinator Interview Questions

Today you will be participating in an interview focusing on your experiences with the science BOE assessment plan within your current school district. The overarching purpose of this study is to explore the patterns of development and implementation of Body-of-Evidence (BOE) science systems throughout the state of Wyoming. I would like to put you on speaker phone and tape-record this interview so that I may analyze the transcripts at a separate time. Is this agreeable with you?

Thank You. Please be aware that your answers are voluntary and confidentiality will be maintained throughout the study. You may withdraw at any time by notifying the author of the study (that would be me, Sharla Dowding). For the purpose of identifying this transcript at a later date, I will begin by asking some demographic questions. Please state your name. Please state where you are employed.

And how long you have been in this position?

Does your school district have a required course of study for graduation requirements in science?

1. Please describe your district’s BOE science assessment system.
   
   1a: What are the features, processes or products that characterize this system?
   
   1b: Would you describe your BOE science assessment plan as district developed, consortium activities, or a blend of both?
   
   1c: Would you describe the nature of your BOE assessments as projects or tests?

2. Please describe how the BOE system is implemented by the science instructors in your district.

   2a: Does this differ throughout your school? (or district?)

3. The Wyoming Assessment Handbook presents eight design principles to guide the development of districts’ Body of Evidence systems. As I mention some of these, please describe your understanding of the ways your science system attempts to integrate these design principles within the BOE.

   3a: Consistency

   3b: Fairness

   3c: Comparability
4. The Wyoming Assessment Handbook states, “In terms of the relationship between standards and assessment, alignment refers to the match between the items on the assessments and the knowledge and skills represented by the curriculum and the standards” (p. 14).

4a. Please describe to me your understanding of the two-way alignment process and how you see it used in your district’s science BOE plan.

4b. Are there ways in which your district’s BOE plan is not well aligned with state standards? If so, please explain.

4c. How about alignment with your district curriculum?

For each item 7 – 10 use the following probes:

-- If so, please describe the process.

-- Whom did you turn to, or what resources did you use, if questions or obstacles arose? -
- What kinds and levels of support did your district provide? Can you give an example?

5. Have you participated in developing the overall design, schedule or specific assessments for your BOE plan? -- If so, please describe the process.

5a. Please describe the principles or philosophies that guide this system.

5b. If questions or obstacles arose, whom did you turn to, or what resources did you use?

5c. What kinds and levels of support did you receive? Can you give an example?
- Was this support from your district level, or from the state?

6. Have you participated in setting cut scores for the science assessment plan in your district?

-- If so, please describe the process.

-- Whom did you turn to, or what resources did you use, if questions or obstacles arose? -
- What kinds and levels of support did your district provide? Can you give an example?
7. Have you participated in determining the frequency or scheduling for the BOE science assessments in your district?

--- If so, please describe the process.

--- Whom did you turn to, or what resources did you use, if questions or obstacles arose? -
- What kinds and levels of support did your district provide? Can you give an example?

8. Have you participated in establishing feedback loops to enable you, or others in your district, to use BOE science assessment results to make instructional decisions or to modify the assessments?

--- If so, please describe the process.

--- Whom did you turn to, or what resources did you use, if questions or obstacles arose? -
- What kinds and levels of support did your district provide? Can you give an example?

9. To whom do you report the BOE data gathered in your district?

10. What are the different ways you utilize BOE data?

11. How does your district disaggregate BOE science assessment results?

   11a. Are the BOE science results reported to students?

   11b. To parents? If yes, please describe the process.

12. Please describe any ways in which you see your students’ science learning affected by the assessment plan adopted by your district.

I would like to interview a high school science teacher in your district who teaches a required course. Do you have a teacher you would recommend?

At the end of the interview, I would ask if they had taken my online survey about the BOE. If not, would they mind going to the website (I’ll email them the link) and taking the survey. I will also make them aware of the session I will be presenting at the school improvement conference and invite them to attend if they are at the conference. Then I
would thank them for their time and assure them that if at any time they have follow-up questions or comments for me they may contact me at my email & phone number.

Teacher Interview Questions

Today you will be participating in an interview focusing on your experiences with the science BOE assessment plan within your current school district. The overarching purpose of this qualitative study is to explore the patterns of development and implementation of Body-of-Evidence (BOE) science systems throughout the state of Wyoming. I would like to put you on speaker phone and tape-record this interview so that I may analyze the transcripts at a separate time. Is this agreeable with you? Thank You. Remember that your answers are voluntary and confidentiality will be maintained throughout the study. You may withdraw at any time by notifying the author of the study (that would be me, Sharla Dowding). For the purpose of identifying this transcript at a later date, I will begin by asking some demographic questions. Please state your name. Please state where you teach and how long you have been in this position. Please describe your teaching assignment. Does your school district have a required course of study for graduation requirements in science?

1. Please describe your district’s BOE science assessment system.
   
   1a: What are the features, processes or products that characterize this system?
   
   1b: Would you describe your BOE science assessment plan as district developed, consortium activities, or a blend of both?
   
   1c: Would you describe the nature of your BOE assessments as projects or tests?

2. Please describe how the BOE system is implemented by you as an individual instructor.

   2a: Does this differ throughout your school? (or district?)

3. The Wyoming Assessment Handbook presents eight design principles to guide the development of districts’ Body of Evidence systems. As I mention some of these, please describe your understanding of the ways your science system attempts to integrate these design principles within the BOE.

   3a: Consistency
3b: Fairness
3c: Comparability
3e: Multiple Measures
3f: Credibility (inviting stakeholders)
3g: Consequences

4. The Wyoming Assessment Handbook states, “In terms of the relationship between standards and assessment, alignment refers to the match between the items on the assessments and the knowledge and skills represented by the curriculum and the standards” (p. 14).

4a. Please describe to me your understanding of the two-way alignment process and how you see it used in your district’s science BOE plan.

4b. Are there ways in which your district’s BOE plan is not well aligned with state standards? If so, please explain.

4c. How about alignment with your district curriculum?

For each item 7 – 10 use the following probes:

-- If so, please describe the process.

-- Whom did you turn to, or what resources did you use, if questions or obstacles arose? -
  - What kinds and levels of support did your district provide? Can you give an example?

5. Have you participated in developing the overall design, schedule or specific assessments for your BOE plan?

5a. Please describe the principles or philosophies that guide this system.

6. Have you participated in setting cut scores for the science assessment plan in your district?

-- If so, please describe the process.

-- Whom did you turn to, or what resources did you use, if questions or obstacles arose?

-- What kinds and levels of support did your district provide? Can you give an example?
7. Have you participated in determining the frequency or scheduling for the BOE science assessments in your own classroom, school or district?

-- If so, please describe the process.

-- Whom did you turn to, or what resources did you use, if questions or obstacles arose? -
- What kinds and levels of support did your district provide? Can you give an example?

8. Have you participated in establishing feedback loops to enable you, or others in your school or district, to use BOE science assessment results to make instructional decisions or to modify the assessments?

-- If so, please describe the process.

-- Whom did you turn to, or what resources did you use, if questions or obstacles arose? -
- What kinds and levels of support did your district provide? Can you give an example?

9. To whom do you report the BOE data gathered in your classroom?

10. What are the different ways you use BOE data?

10a. Does this use differ throughout your school? (or district?)

11. How does your district disaggregate BOE science assessment results?

11a. Are the BOE science results reported to students?

11b. To parents? If yes, please describe the process.

12. Please describe any ways in which you see your students’ science learning affected by the assessment plan adopted by your district.

At the end of the interview – remember to ask them if they have any questions for me.
APPENDIX D

RESEARCH QUESTION MATRIX
### Research Question Matrix

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Method for collection</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. RQ#1: What design elements and implementation processes characterize BOE</strong></td>
<td></td>
</tr>
<tr>
<td>science assessment plans across Wyoming school districts?</td>
<td></td>
</tr>
</tbody>
</table>
| 1a: What are the characteristics of the BOE systems being implemented today within and across districts? | Survey item #2.3: Does your BOE science assessment plan provide multiple opportunities for individual assessment pieces?  
Survey item #2.4: Does your BOE science assessment plan include any consortium activities? If so, please check all that apply.  
Survey item #2.5: Does your BOE science assessment plan provide multiple opportunities for individual assessment pieces selected above? If so, for which ones?  
Survey item #2.6: What criteria for proficiency does your BOE science plan adopt, stand mastery or cut scores?  
Collection of evidence from Body of Evidence plans from several selected districts.  
Interview Questions: Tell me about your BOE science assessment system. Probing Questions may be needed: What are the characteristics that you feel define your system? |
| 1b: How are the design principles presented in the Wyoming Assessment Handbook integrated into district BOE science plans? | Survey item #2.2: Are your BOE science assessments integrated as assessments within classroom instruction, or administered as stand-alone assessment pieces?  
Interview Questions: The WY assessment handbook indicates 8 design principles should guide the development of districts’ Body of Evidence systems. I’d like to name these and have you describe to me your understanding about the ways your science system attempts to integrate these |
### 1c. What processes do different districts utilize to ensure that assessments are coordinated with state science standards and benchmarks?

<table>
<thead>
<tr>
<th>Question</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survey item #2.1: Are you familiar with your district science assessment plan blueprint (sometimes called a matrix) of alignment to state science standards? Interview Questions: “In terms of the relationship between standards and assessment, alignment refers to the match between the items on the assessments and the knowledge and skills represented by the curriculum and the standards” (WY assessment handbook, p14). Please describe to me your understanding of the two way alignment process and how you see it used in your science BOE plan.</td>
<td></td>
</tr>
</tbody>
</table>

### RQ#2: What support do teachers receive to develop a BOE science plan, align the plan to the state science education standards, and implement the intended science plan?

<table>
<thead>
<tr>
<th>Question</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>2a: What support do teachers receive while developing, aligning and implementing district BOE science plans, and how does this vary across districts?</td>
<td>Survey item #2.7: Have you participated in setting cut-scores for the science assessment plan in your current district position? Interview Question: Have you participated in developing an assessment for your BOE plan? If so, please describe this process. Whom did you turn to when you had questions about developing the assessment? Interview Question: Have you participated in setting cut-scores for the science assessment plan in your district? If so, please describe this process. May need probing questions for this one: What type of assessment were you developing, a multiple choice test, a project? Is this assessment used in more than one classroom? Interview Question: Tell me how BOE assessments are implemented in your</td>
</tr>
</tbody>
</table>
| 2b: Is there a relationship between the extent or nature of the activities within the BOE science plan and the level of support teachers receive to develop, align, and implement the intended plan? | Survey item # 3.1: In the last three years, approximately how many hours have you spent working on your BOE science assessment plan?  
Survey item #1.3: Have you participated in a BOE peer-review at the state level?  
Survey item #1.4: Have you participated in developing a BOE science assessment plan in your current district position?  
Interview Question: Would you describe your BOE science assessment plan as district developed, consortium activities, or a blend of both? Would you describe the nature of your BOE assessments as projects or tests? |
| 2c: Is the support teachers perceive they receive in a given district connected to the support curriculum directors perceive they provide in developing and sustaining a district’s BOE plan? | Survey item # 3.1: In the last three years, approximately how many hours have you spent working on your BOE science assessment plan?  
Survey item #3.2a: As a professional development experience, how do you rate your experience developing the BOE plan in your district?  
Survey item #3.2.b: How do you rate your experience with the implementation of the BOE plan in your district?  
Interview Question: Describe the support you receive from your district administration to develop and/or to sustain your BOE assessment plan in science. Whom do you see this support coming from? May need probing questions here: In what areas do you feel most supported, developing or implementing assessments? In what ways would you like to receive more support?  
Interview Question (for CC): Describe the |
support you provide teachers to develop and/or to sustain the BOE science assessment plan in your district. May need probing questions here: In which area do you feel you provide the most support to teachers, in developing or sustaining the BOE plan? In what ways would you like to provide more support?
RQ#3: The BOE approach was originally envisioned by the 2000 WY state legislature as a way to provide districts with a measure of decision-making autonomy, participation and ownership within the structure of state-mandated assessment. How do districts handle this opportunity to design and implement a locally useful assessment?

| 3a. What are the different ways districts utilize BOE data? | Teacher Interview Question: What are the different ways your district utilizes BOE data? To whom do you report your BOE data?  
CC Interview Question: What are the different ways your district utilizes BOE data? To whom do teachers report their BOE data? |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3b. How do districts disaggregate assessment results?</td>
<td>Interview Question: How does your district disaggregate BOE science assessment results? Are the results reported to students, to parents?</td>
</tr>
</tbody>
</table>
| 3c. In what ways do teachers, curriculum coordinators, and administrators see student learning taking place as a result of BOE assessment plans? | Survey item #3.2c: What is your opinion of the impact on student learning taking place in your classroom as a result of the BOE assessment plan your district has adopted?  
Survey item #3.3: In what ways do you see student learning taking place in your classroom, as a result of the assessment plan adopted by your district? (I would ask this question as an interview question also.) |
APPENDIX E

BODY OF EVIDENCE PEER REVIEW RUBRICS
# BODY OF EVIDENCE PEER REVIEW RUBRIC

## ALIGNMENT

<table>
<thead>
<tr>
<th>Meets Criteria (ALL bullets)</th>
<th>Evidence in Plan to Support Criteria</th>
<th>Does Not Meet Criteria (Highlight bullet(s) that apply)</th>
<th>Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>- There is documentation of adequate sampling of benchmarks as well as all the standards within the two representative content areas.</td>
<td>- Assessment samples for the representative content areas (1 core &amp; 1 non-core) are included.</td>
<td>- The district provides little, incomplete, unclear or no evidence of adequate sampling.</td>
<td>- The district provides little, incomplete, unclear or no evidence of adequate sampling.</td>
</tr>
<tr>
<td>- There is evidence of a two-way alignment process: all assessment items and tasks align to standards and are represented in the assessments within the two representative content areas.</td>
<td>- Blueprints for the assessment samples are included in the plan.</td>
<td>- The district provides little, incomplete, unclear or no evidence of two-way alignment.</td>
<td>- The district provides little, incomplete, unclear or no evidence of two-way alignment.</td>
</tr>
<tr>
<td>- The assessments from the representative content areas reflect the cognitive depth of the content standards and the types of student performance described in the performance standards.</td>
<td>- Matrices indicating all the assessments in the representative content areas (1 core &amp; 1 non-core) and the standards and benchmarks assessed by each are included.</td>
<td>- The evidence that the assessments reflect the cognitive depth of the content standards and the types of student performance described in the performance standards.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- The processes used by the district to ensure alignment of current standards and benchmarks as well as future changes are described.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- If the district Body of Evidence system includes course-based information (e.g., grades), the process for assuring alignment among the course curriculum, standards, assessments, and grading practices are described and appropriate polices included.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Evidence of the processes used to ensure alignment of assessment items/tasks to the levels of cognition called for in the performance standards is present.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Evidence of “think aloud” protocols and/or careful examination of student work is used to evaluate/document, and revise, if necessary, the alignment of its standards and assessment system.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Wyoming Department of Education
BOE Peer Review Rubric Spring 2008
### CONSISTENCY

<table>
<thead>
<tr>
<th>Meets Criteria</th>
<th>Evidence in Plan to Support Criteria</th>
<th>Does Not Meet Criteria</th>
<th>Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>(First two bullets, and 3rd bullet only if teacher judgment is used in plan)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• For <strong>open-ended assessments</strong>, the district plan describes clear procedures to be used to ensure inter-rater reliability and defines a desired, acceptable rate. Data are presented that support implementation of the stated procedures.</td>
<td>□ The procedures used to ensure inter-rater reliability on open-ended assessments are described.</td>
<td>□ The district provides little, incomplete, unclear or no evidence of procedures to be used to ensure inter-rater reliability on <strong>open-ended assessments</strong>.</td>
<td>• The district provides little, incomplete, unclear or no evidence of procedures to be used to ensure inter-rater reliability on <strong>open-ended assessments</strong>.</td>
</tr>
<tr>
<td>• For <strong>closed-ended assessments</strong>, the district plan describes clear procedures to be used to ensure reliability and defines a desired, acceptable rate. Data are presented that support implementation of the stated procedures.</td>
<td>□ Inter-rater reliability data that meets acceptable rates (inter-rater reliability to meet or exceed 80% exact agreement and 98% exact + adjacent agreement) is included.</td>
<td>□ The district provides little, incomplete, unclear or no evidence of procedures to be used to ensure reliability on <strong>closed-ended assessments</strong>.</td>
<td>• The district provides little, incomplete, unclear or no evidence of procedures to be used to ensure reliability on <strong>closed-ended assessments</strong>.</td>
</tr>
<tr>
<td>• If <strong>teacher judgment</strong> is part of the plan, the plan describes procedures to ensure reliability of judgment across assessments within a course &amp; across teachers. There is clear documentation that judgment is anchored to the performance standards. Data are presented that support implementation of the stated procedures.</td>
<td>□ The procedures used to ensure reliability of teacher judgment across assessments within a course and across multiple teachers are described.</td>
<td>□ Reliability data on closed-ended assessments (to meet or exceed average reliability coefficients greater than 0.85) is included.</td>
<td>• The district provides little, incomplete, unclear or no evidence of <strong>teacher judgment</strong>.</td>
</tr>
<tr>
<td></td>
<td>□ Reliability data of teacher judgment is included.</td>
<td>□ Procedures used to ensure reliability of teacher judgment across assessments within a course and across multiple teachers are described.</td>
<td>• The district provides little, incomplete, unclear or no evidence of acceptable rates of reliability being defined.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ The district provides little, incomplete, unclear or no evidence of <strong>data</strong> that supports implementation of the stated procedures.</td>
<td></td>
</tr>
<tr>
<td>FAIRNESS</td>
<td>Evidence in Plan to Support Criteria</td>
<td>Does Not Meet Criteria (Highlight bullets that apply)</td>
<td>Recommendations</td>
</tr>
<tr>
<td>----------</td>
<td>-------------------------------------</td>
<td>-----------------------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>□ Meets Criteria &lt;br&gt; (ALL bullets)</td>
<td>- There is evidence the district uses procedures or tools to ensure that assessment items/tasks are not biased against subgroups of students. &lt;br&gt; - There is evidence the district uses accommodations and alternate assessments appropriately. &lt;br&gt; - A variety of assessment formats and strategies are included in the system. &lt;br&gt; - The district disaggregates assessment results (i.e. ethnicity, gender &amp; socio-economic status) and the results are used to search for possible bias in the system. &lt;br&gt; - Relevant district data are presented to document that participation rates are at least 95% for all subgroups.</td>
<td>- The procedures (e.g., bias committees) used to ensure that items and tasks are not biased against any subgroups of students are described. &lt;br&gt; - Sample forms and/or notes from bias review committee meetings are included. &lt;br&gt; - Policies and procedures for ensuring fair participation of all students in the system (e.g. students with disabilities or English language proficiency) are evident. &lt;br&gt; - There is evidence that illustrates accommodations and alternate assessments are used. &lt;br&gt; - There is evidence that the district system provides students with multiple opportunities, using different formats and strategies, to demonstrate their knowledge and skills. &lt;br&gt; - The plan includes disaggregated assessment results by identifiable subgroups (i.e. ethnicity, gender &amp; socio-economic status) and describes how the district uses the information to make decisions. &lt;br&gt; - There is evidence that disaggregated assessment results are used to search for potential bias in the assessment system. &lt;br&gt; - The plan includes participation rates data for the content area assessments submitted.</td>
<td>- The district provides little, incomplete, unclear or no evidence of plans, procedures, or tools to ensure that assessment items/tasks are not biased against any subgroups of students. &lt;br&gt; - The district provides little, incomplete, unclear or no evidence that accommodations and alternate assessments are used appropriately. &lt;br&gt; - The district provides little, incomplete, unclear or no evidence that multiple assessment opportunities are provided. &lt;br&gt; - The district provides little, incomplete, unclear or no evidence that a variety of assessment formats and strategies are included in the system. &lt;br&gt; - The district provides little, incomplete, unclear or no evidence of a process being used to disaggregate assessment results and the results are being used to search for possible bias in the system. &lt;br&gt; - The district provides little, incomplete, unclear or no evidence that participation rates are at least 95% for all subgroups.</td>
</tr>
</tbody>
</table>

Wyoming Department of Education  
BOE Peer Review Rubric Spring 2008
<table>
<thead>
<tr>
<th>STANDARD SETTING</th>
<th>Evidence in Plan to Support Criteria</th>
<th>Does Not Meet Criteria</th>
<th>Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ Meets Criteria (G.L. bullets)</td>
<td>☐ The rationale and the standard-setting method used for determining proficiency at the content level is describec.</td>
<td>☐ The district provides little, incomplete, unclear or no evidence of a rationale and a defensible method of standard-setting which describes how the determination of level of proficiency is made at the content level.</td>
<td>□ The district provides little, incomplete, unclear or no evidence that cut scores are clearly tied to performance standards.</td>
</tr>
<tr>
<td>☐ Meets Criteria (G.L. bullets)</td>
<td>☐ The cut scores used for each level of proficiency in the representative content area are included in the plan.</td>
<td>☐ The district plan provides little, incomplete, unclear or no evidence of the cut scores used in each content area.</td>
<td>□ The district plan provides little, incomplete, unclear or no evidence that cut scores are clearly tied to performance standards.</td>
</tr>
<tr>
<td>☐ Meets Criteria (G.L. bullets)</td>
<td>☐ The levels at which the cut scores have been set are clearly tied to the performance descriptors for the representative content areas.</td>
<td>☐ The district plan provides little, incomplete, unclear or no evidence that cut scores are clearly tied to performance standards.</td>
<td>□ The district plan provides little, incomplete, unclear or no evidence of a timeline showing adequate notification to students on progress toward proficiency in each content area.</td>
</tr>
<tr>
<td>☐ Meets Criteria (G.L. bullets)</td>
<td>☐ How and when individual scores are aggregated to make &quot;graduate/not graduate&quot; decisions are explained.</td>
<td>☐ The district plan provides little, incomplete, unclear or no evidence that cut scores are clearly tied to performance standards.</td>
<td>□ The district plan provides little, incomplete, unclear or no evidence that key stakeholders have been involved in the standard-setting process.</td>
</tr>
<tr>
<td>☐ Meets Criteria (G.L. bullets)</td>
<td>☐ The plan includes the timeline the district uses for student notification process.</td>
<td>☐ The district plan provides little, incomplete, unclear or no evidence that key stakeholders have been involved in the standard-setting process.</td>
<td>□ The district plan provides little, incomplete, unclear or no evidence that key stakeholders have been involved in the standard-setting process.</td>
</tr>
<tr>
<td>☐ Meets Criteria (G.L. bullets)</td>
<td>☐ The plan describes how key stakeholders are involved in the standard-setting process.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Wyoming Department of Education
BOE Peer Review Rubric Spring 2008
## Comparability

<table>
<thead>
<tr>
<th>Meets Criteria</th>
<th>Evidence in Plan to Support Criteria</th>
<th>Does Not Meet Criteria</th>
<th>Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐</td>
<td>□ There is documentation of ongoing district-wide trainings, common rubrics, the use of “seeded” papers, and common administration guidelines used to ensure comparability.</td>
<td>☐ The district provides little, incomplete, unclear or no evidence that specific procedures are in place for ensuring comparability of assessments for all students in a given year.</td>
<td>☐ The district provides evidence that specific procedures are in place for ensuring comparability across years.</td>
</tr>
<tr>
<td></td>
<td>□ The district has a process for ensuring the assessments are administered similarly from year-to-year.</td>
<td>☐ The district provides little, incomplete, unclear or no evidence that specific procedures are in place for ensuring comparability across years.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>□ There is evidence that the district ensures that assessments are scored the same as in previous years (e.g., the use of anchor papers and common scoring rubrics, and scoring workshops for new teachers).</td>
<td>☐ The district provides little, incomplete, unclear or no evidence that specific procedures are in place for replacing assessment tasks/items with comparable tasks/items in terms of content, focus, and cognitive demand.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>□ The plan includes evidence of procedures for replacing assessment tasks/items such as the use of assessment blueprints and protocols.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- The district provides evidence that specific procedures are in place for ensuring comparability of assessments for all students in a given year, regardless of classroom, program, or school in the district.
- The district provides evidence that specific procedures are in place for ensuring comparability across years.
- The district provides evidence that specific procedures are in place for replacing assessment tasks/items with comparable tasks/items in terms of content, focus, and cognitive demand.
APPENDIX F

EXCEPRT FROM NCLB LAW
“PART A—IMPROVING BASIC PROGRAMS OPERATED BY LOCAL EDUCATIONAL AGENCIES ‘‘Subpart 1—Basic Program Requirements ‘‘SEC. 1111. STATE PLANS. ‘‘(2) ACCOUNTABILITY. — ‘‘(A) IN GENERAL.—Each State plan shall demonstrate that the State has developed and is implementing a single, statewide State accountability system that will be effective in ensuring that all local educational agencies, public elementary schools, and public secondary schools make adequate yearly progress as defined under this paragraph. Each State accountability system shall— ‘‘(i) be based on the academic standards and academic assessments adopted under paragraphs (1) and (3), and other academic indicators consistent with subparagraph (C)(vi) and (vii), and shall take into account the achievement of all public elementary school and secondary school students; ‘‘(ii) be the same accountability system the State uses for all public elementary schools and secondary schools or all local educational agencies in the State, except that public elementary schools, secondary schools, and local educational agencies not participating under this part are not subject to the requirements of section 1116; and ‘‘(iii) include sanctions and rewards, such as bonuses and recognition, the State will use to hold local educational agencies and public elementary schools and secondary schools accountable for student achievement and for ensuring that they make adequate yearly progress in accordance with the State’s definition under subparagraphs (B) and (C).” (NCLB, 115 STAT. 1446)
APPENDIX G

ADEQUATE YEARLY PROGRESS
Appendix G – Adequate Yearly Progress

‘‘(B) ADEQUATE YEARLY PROGRESS.—Each State plan shall demonstrate, based on academic assessments described in paragraph (3), and in accordance with this paragraph, what constitutes adequate yearly progress of the State, and of all public elementary schools, secondary schools, and local educational agencies in the State, toward enabling all public elementary school and secondary school students to meet the State’s student academic achievement standards, while working toward the goal of narrowing the achievement gaps in the State, local educational agencies, and schools.” (NCLB, 115 STAT. 1446)

In 115 STAT. 1446 AYP is defined:

‘‘(C) DEFINITION.—‘Adequate yearly progress’ shall be defined by the State in a manner that— ‘‘(i) applies the same high standards of academic achievement to all public elementary school and secondary school students in the State; ‘‘(ii) is statistically valid and reliable; ‘‘(iii) results in continuous and substantial academic improvement for all students; ‘‘(iv) measures the progress of public elementary schools, secondary schools and local educational agencies and the State based primarily on the academic assessments described in paragraph (3); ‘‘(v) includes separate measurable annual objectives for continuous and substantial improvement for each of the following: ‘‘(I) The achievement of all public elementary school and secondary school students. ‘‘(II) The achievement of— ‘‘(aa) economically disadvantaged students; ‘‘(bb) students from major racial and ethnic groups; ‘‘(cc) students with disabilities; and ‘‘(dd) students with limited English proficiency; except that disaggregation of data under subclause (II) shall not be required in a case in which the number of students in a category is insufficient to yield statistically reliable information or the results would reveal personally identifiable information about an individual student; ‘‘(vi) in accordance with subparagraph (D), includes graduation rates for public secondary school students (defined as the percentage of students who graduate from secondary school with a regular diploma in the standard number of
years) and at least one other academic indicator, as determined by
the State for all public elementary school students; and ``(vii) in
accordance with subparagraph (D), at the State’s discretion, may
also include other academic indicators, as determined by the State
for all public school students, measured separately for each group
described in clause (v), such as achievement on additional State or
locally administered assessments, decreases in grade-to-grade
retention rates, attendance rates, and changes in the percentages of
students completing gifted and talented, advanced placement, and
college preparatory courses.

“(D) REQUIREMENTS FOR OTHER INDICATORS.—In
carrying out subparagraph (C)(vi) and (vii), the State— ``(i) shall
ensure that the indicators described in those provisions are valid
and reliable, and are consistent with relevant, nationally recognized
professional and technical standards, if any; and ``(ii) except as
provided in subparagraph (I)(i), may not use those indicators to
reduce the number of, or change, the schools that would otherwise
be subject to school improvement, corrective action, or
restructuring under section 1116 if those additional indicators were
not used, but may use them to identify additional schools for
school improvement or in need of corrective action or
restructuring.” (NCLB, 115 STAT. 1446)
APPENDIX H

WYOMING HOUSE BILL 0078
(a) A statewide task force is created to:

(i) Investigate, assemble and recommend necessary accountability processes and systems to assist state efforts in addressing education accountability requirements of the federal No Child Left Behind Act of 2001 (NCLB), while maintaining uniformity and quality of the statewide education program standards and state student content and performance standards as required under W.S.21-2-304. Recommendations under this paragraph shall specify a plan to implement sufficient accountability processes and systems to enable the implementation of rewards and sanctions as specified by the federal NCLB Act based on the educational performance and progress of each school district and each school within a district.

(b) The statewide task force created under subsection (a) of this section shall be comprised of the following:

(i) One (1) member who is a school district superintendent appointed by the state superintendent of public instruction;

(ii) One (1) member who is a secondary school principal from a Wyoming school district appointed by the Wyoming association of secondary school principals;

(iii) One (1) member who is an elementary school from a Wyoming school district appointed by the Wyoming association of elementary school principals;

(iv) One (1) member who is a teacher in a Wyoming school district appointed by the state superintendent of public instruction;
(v) One (1) member who is a special education program or service instructor and provider for children with disabilities appointed by the state superintendent of public instruction;

(vi) One (1) member who is serving as a trustee on a Wyoming school district board of trustees appointed by the Wyoming school boards association;

(vii) One (1) member serving on the Wyoming state board of education elected by the board as representative;

(viii) One (1) member who is a school district curriculum director appointed by the state superintendent of public instruction;

(ix) One (1) member appointed by the governor to represent private business;

(x) One (1) member of the Wyoming appointed by the president of the senate;

(xi) One (1) member of the Wyoming house of representatives appointed by the speaker of the house; and

(xii) Two (2) members who are parents of school children appointed by the governor.

(c) Appointments specified under subsection (b) of this section shall be made by not later than April 1, 2003.

(d) Recommendations developed under paragraph (a)(ii) of this section shall:

(i) Integrate federal and state accountability requirements which shall be based upon the statewide student assessment system established under W.S.21-2-304(a)(v), statewide education program standards and student content and performance standards;

(ii) Develop a system of rewards and sanctions applicable to all schools and school districts which is consistent with school accountability and school improvement provisions specified by the federal NCLB Act and which requires corrective action for those schools not meeting established progress levels;
(iii) Specify a level for and measure annual progress of schools and school districts, including all schools receiving state assistance;

(iv) Be developed in sufficient time for implementation in the 2004-2005 school year.

(e) On or before August 1, 2003, the statewide task force shall report its findings and recommendations developed under this section to the joint education interim committee. Following its review and approval, the committee shall present recommendations including any necessary enabling legislation, to the legislature at the 2004 budget session.

(f) For school year 2003-2004 and prior to implementation of the recommendations developed under paragraph (a)(ii) of this section for implementation during school year 2004-2005, and as necessary to comply with the federal NCLB Act, the state department of education shall, in consultation with the task force established by this section, establish a transitional plan to provide a temporary system of rewards and sanctions for all schools and school districts. Prior to implementation, the transitional plan shall be presented to the joint education interim committee on or before October 1, 2003. The transitional plan shall terminate upon expiration of the 2003-2004 school year.

Section 202. Rewards and sanctions imposed under the federal No Child Left Behind Act of 2001 shall apply only to those schools receiving federal Title I funds until the processes and systems developed under section 201 of this act are implemented and operational.