3-2008

Standards-Based Assessment System in a Christian Middle School Science Classroom

David J. Mulder
Dordt College

Standards-Based Assessment System in a Christian Middle School Science Classroom

Abstract
Teachers in all sorts of educational settings have a common difficulty: accurately reporting what their students know, understand, and are able to do. Reporting the measurement of these assessments has traditionally been done through the assignment of letter grades. There is a fairly comprehensive and growing body of literature indicating the weakness of traditional letter grades for authentically assessing student understanding. This study examines an alternative to traditional grading practices. After reviewing relevant literature, I designed a standards-based assessment and evaluation system and put it into place in my 8th grade science classroom at a mid-sized Christian school in northwestern Iowa for the first 9-week quarter of the 2007-2008 academic year. I used a mixed-methods design for my study, including case study and survey elements. After the first quarter was completed, I surveyed participating 8th grade science students and their parents for their response to the newly implemented alternative assessment system. The survey results were mixed, likely attributable to the significant cultural shift required of students, parents, and even teachers, to accept an alternative to traditional letter grades for assessing student understanding. Both the student participants and their parents, however, indicated that online progress reports accurately described students' achievement in science. The overall result encourages me to continue the implementation of a standards-based assessment and evaluation system in my science classroom.

Document Type
Thesis

Degree Name
Master of Education (MEd)

Department
Graduate Education

First Advisor
Ed Starkenburg

Keywords
Master of Education, thesis, Christian education, Sioux Center Christian School, science education, middle school, junior high students

Subject Categories
Curriculum and Instruction | Education

Comments
Action Research Report Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Education

A Standards-Based Assessment System in a Christian Middle School Science Classroom

by

David J. Mulder

B.A. Dordt College, 1998

Action Research Report
Submitted in Partial Fulfillment
of the Requirements for the
Degree of Master of Education

Department of Education
Dordt College
Sioux Center, Iowa
March 2008

A Standards-Based Assessment System in a Christian Middle School Science Classroom

by

David J. Mulder

Approved:

_____
Faculty Advisor

_____
Date

Approved:

_____
Director of Graduate Education

_____
Date

Table of Contents

List of Figures

Figure Page

Abstract

Teachers in all sorts of educational settings have a common difficulty: accurately reporting what their students know, understand, and are able to do. Reporting the measurement of these assessments has traditionally been done through the assignment of letter grades. There is a fairly comprehensive and growing body of literature indicating the weakness of traditional letter grades for authentically assessing student understanding. This study examines an alternative to traditional grading practices. After reviewing relevant literature, I designed a standards-based assessment and evaluation system and put it into place in my 8th grade science classroom at a mid-sized Christian school in northwestern Iowa for the first 9-week quarter of the 2007-2008 academic year. I used a mixed-methods design for my study, including case study and survey elements. After the first quarter was completed, I surveyed participating 8th grade science students and their parents for their response to the newly implemented alternative assessment system. The survey results were mixed, likely attributable to the significant cultural shift required of students, parents, and even teachers, to accept an alternative to traditional letter grades for assessing student understanding. Both the student participants and their parents, however, indicated that online progress reports accurately described students' achievement in science. The overall result encourages me to continue the implementation of a standards-based assessment and evaluation system in my science classroom.

Introduction

In schools across North America, teachers assign letter grades to their students' work to inform students (and their parents) of their achievement. Research shows, however, that traditional letter grades are a flawed method for communicating this important information. Letter grades as used in traditional assessment systems are weak measures for at least three reasons: 1) they are ambiguous; 2) they do not advance learning and can, in fact, encourage shallow learning; and 3) they miss the true purpose of assessment. Let us examine each of these critiques.

First, letter grades, as typically assigned, have an ambiguous meaning. Common grading practice allows, and even encourages, teachers to include non-achievement factors into the grades they calculate. Marzano (2000) has found that teachers at all grade levels often include such factors as attendance, effort, and classroom behavior in the calculation of an "academic" grade. Brouwer (2007) argues, "The final mark we assign should reflect current learning achievement *only*. It should be unclouded by other information about the student such as tardiness, effort, participation, or late work…if we include them in the mark, we muddle the communication about current learning achievement" (p. 8). These factors, while important to report, confuse the meaning of a grade, which is primarily a measure of academic achievement.

Additionally, common grading practice "mixes different types of knowledge and skills into single scores on assessments" (Marzano, 2000, p. 13). Teachers often attempt to measure different skills and abilities in a single assessment tool. Imagine a science teacher writing a test with two distinct sections, the first asking students to recall facts, and the second requiring students to think creatively and apply concepts to solve

problems. The problem with this practice is that simply marking the paper with a "B+" does not truly indicate the student's mastery of *either* of these skills. The student's actual achievement of content and skills is lost in the ambiguity of the letter grade.

Secondly, simply assigning letter grades does not advance learning. One of the primary functions of assessment is to give feedback to students about their mastery of content and/or skills (Brouwer, 2007). Merely writing a letter grade at the top of a student's paper does not give constructive feedback. As McTighe & O'Connor (2005) put it, "Pinning a letter (B-) or a number (82%) on a student's work is no more helpful than such comments as 'Nice job' or 'You can do better.' Although good grades and positive remarks may feel good, they do not advance learning" (p. 16).

It can be argued that the practice of assigning letter grades actually encourages shallow learning. Kohn (1993) explores this concept extensively, finding that letter grades actually discourage student risk-taking and diminish students' interest in learning; he argues that the goal for students expecting the "reward" of a letter grade becomes making the best grades they can while expending as little effort as possible. Further, Clymer & Wiliam (2007) report, "In most classrooms, if students forget something that they have previously been assessed on, they get to keep the grade" (p. 38). If students are not encouraged to explore, wrestle with, deeply understand, and even *enjoy* the content under consideration, they will simply memorize what they need to in order to secure a letter grade on a particular test or quiz, and then promptly forget it once the assessment is over. In this case, the letter grade might actually get in the way of real learning!

Thirdly, the practice of assigning letter grades often misses the true purpose of assessment. Wiggins & McTighe (2005) explain that assessment should not be

fundamentally about the grade, but rather, it ought to be aimed at uncovering what students know, understand, and are able to do. In traditional assessment systems, however, the letter grade is often the main emphasis. Stiggins (2005) states:

> Traditionally, schools have used assessment—the pending final exam, the unannounced pop quiz, and the threat of low or failing report card grades—to motivate students. To maximize learning, our teachers believed, maximize anxiety. Assessment has served as the great intimidator. Pressure to get high test scores and good grades, it was believed, would motivate greater effort and thus more learning. (p. 324)

Brouwer (2007) agrees, explaining that many students simply see the assignment of letter grades as something "done to them" (p. 6), either rewarding their efforts or labeling them as underachievers. In either case, the grade does nothing to encourage a student to advance his or her learning.

In the traditional assessment system described above, if students don't perform well on a particular test, assignment, or project, their scores are recorded and remain fixed. There is no option for improvement. We teachers have so much content to "cover" that we have to move on to new topics. There is no other opportunity for students to *learn* what they are expected to learn, and no other opportunity to *show* that they have learned it. A student's grade under this system has an ambiguous meaning at best. In many cases, the grade is really more a measure of how much a student *did not* understand than a measure of what was learned, which makes the mark a weak measure. Using grades this way misses the point of assessment.

Marzano (2000) asserts, “Today’s system of classroom grading is at least 100 years old and has little or no research to support its continuation” (p. 13). Thankfully, in recent years much research has been done into alternatives to traditional grading practices. One such alternative addresses all three of the aforementioned weaknesses of traditional grading systems: a standards-based assessment and evaluation system. A standards-based system may help reduce ambiguity of letter grades, allowing teachers to more accurately measure and report achievement. A standards-based system might be implemented in such a way as to encourage deep learning. And a standards-based system can be an effective way to draw the emphasis back to the real point of assessment: uncovering what students know, understand, and are able to do.

Research Questions

There are two questions I hope this action research project will answer:

1. Will a standards-based assessment system give students and their parents an accurate picture of student achievement in science?
2. Is a standards-based assessment system appropriate and practical for a Christian middle school science program?

Definitions

For the purpose of my action research, the following definitions will be used. These definitions are my own, unless otherwise noted.

Assessment – “all the possible means whereby teachers make judgments about what students have learned” (Hein & Price, 1994, p.1).

Achievement – the student has demonstrated a particular level of evidence to meet a given content standard. Levels of achievement in this study will include:

- "**Beginning**" – the student has just begun to understand a given content standard. The criteria for this content standard have not been met.
- "**Developing**" – the student occasionally meets criteria, but has not yet established proficiency for a given content standard.
- "**Proficient**" – the student regularly meets criteria for a given content standard.
- "**Advanced**" – the student has mastered and often exceeds criteria for a given content standard.

Big Ideas – key understandings students will develop through the course of their studies. Big ideas correspond to content standards in a standards-based assessment system.

Content standard – a statement describing an understanding or skill against which student work can be compared for judgment of achievement. For example, in science education, a common content standard is "Students exhibit abilities necessary to perform scientific inquiry," which is a set of skills a student must master to demonstrate proficiency in science (National Research Council [NRC], 1996, p. 105).

Evaluation – assigning a value to an assessment of achievement.

Feedback – information given to the student to inform him/her of achievement. Not simple comments such as, "Nice job!" but rather, focused information about what must be improved to demonstrate understanding. Feedback may be given via rubric scores or written comments, as is appropriate.

Formative assessment – assessment to aid learning, often embedded in the learning tasks. Formative assessments are primarily desirable to allow the teacher to provide feedback to the student regarding their developing level of achievement.

Grade – the value assigned to a summative assessment; e.g., the traditional A – B – C – D – F grading scheme.

Measurement – the process of comparing a sample of student work for a given assessment to the content standard to judge achievement.

Summative assessment – assessment to measure overall achievement of a given content standard. Final evaluation of achievement occurs in summative assessment; this is where grades are assigned.

A Brief Review of the Relevant Literature

Traditionally, assessment in classrooms has taken the shape of the teacher marking assignments or tests and assigning letter grades based on the number of points the student earned or the percentage of questions the student answered correctly. Over the past number of years, however, classroom assessment practices have begun to shift away from simply marking papers with letter grades. As Vander Ark (2000) puts it, "You can't read a professional journal without tripping over 'authentic assessment,' 'rubrics,' and 'student portfolios.' …Student projects, free writing, and problem-solving are in" (p. 83). Brouwer (2007) would encourage us to consider assessment more broadly than simply assigning grades, thinking of assessment instead as "an ongoing process that not only measures learning, but also supports and encourages it" (p. 6). Along these lines, Hein & Price (1994) assert, "Anything students do can be used for assessment" (p.

13). Clearly, true assessment of student learning is being viewed, by some educators at least, as more than traditional letter grades.

Many authors have commented on the weaknesses of traditional grading practices for assessing student understanding (Clymer & Wiliam, 2007; Glatthorn, 1998; Hein & Price, 1994; Kohn, 1993; Marzano, 2000; National Research Council [NRC], 1996; Tomlinson, 2008; Stiggins, 2005; Van Dyk, 1995). Tomlinson (2008) admonishes teachers to shift their thinking about assessment from a mindset of *judging* students to one of *guiding* students, noting that "Giving students feedback [seems] to be more productive than giving them grades" (p. 10). If teachers really want to understand what their students have learned, they must change from traditional grading practices.

Traditional grading practice is largely comprised of assigning letter grades based on "averaged" scores. In traditional grading systems, when students do an assignment, or take a test, or perform some authentic assessment task, the teacher assigns a number of points to the assigned work. Each assignment is assessed, and students "earn" some number of the points possible for that particular assignment. At the end of the marking period, the teacher calculates the number of points possible, and the number of points each student has earned. This ratio of points earned to points possible comprises the score on which the final grade is based.

From my conversations with other teachers at a variety of grade levels in a variety of schools across North America, I know that calculating an "average" this way is common practice. But what is really being "averaged?" The letter grade assigned by this grading practice does not actually measure student understanding; it simply aggregates how well students performed on each of the tasks assigned to them over the marking

period. On some of these tasks students may have performed well; on others they may have performed poorly. If a student would end up with a "B" average, does that mean *all* his or her work was "B" work? Or was half of it "A" work and half of it "C" work? The letter doesn't actually inform much about the student's *learning*. Simply tallying up points and calculating an average grade does not show students or their parents what they have achieved and how they can continue to improve.

A different assessment system is needed, an alternative to the traditional grading practice of calculating an "average." What might an alternative look like? The two underlying assumptions in Marzano's (2000) work are instructive for creating such an alternative: 1) grades are for providing feedback to students and parents, and 2) a criterion-referenced grading system is best able to provide such feedback.

Science educators should note that these guidelines match the fundamental ideas behind of the *National Science Education Standards* (*NSES*) (NRC, 1996) and *Benchmarks for Science Literacy* (*Benchmarks*) (American Association for the Advancement of Science [AAAS], 1993). Both *NSES* and *Benchmarks* outline specific content standards against which students must be able to demonstrate proficiency. The authors of *Benchmarks* stress the value of criterion-referenced assessment for providing feedback to students and parents (AAAS, 1993). In *NSES*, it is recommended that states and local districts "develop mechanisms to measure students' achievement as specified in the content standards" (NRC, 1996, p.78). This is the primary goal of a move from the traditional letter grading system to a standards-based assessment system.

Such standards-based approaches to assessment benefit teachers and students in a variety of ways. Marzano (2000) suggests that standards provide teachers a framework

for the content to be addressed within a grading period. Stiggins (2005) points out that in standards-based systems, students' goal for achievement shifts from one of competing with classmates for grades to one of personal achievement and competence. Also, while Kluth & Straut (2001) remind teachers that "standards require a wide range of assessment tools" (p. 44), McTighe & O'Connor (2005) point out that using a variety of assessment options can allow for many different means for students to demonstrate their mastery of the content. Finally, Clymer & Wiliam (2007) conclude, "[Standards-based assessment systems] communicate standards for success, helping students see what they need to improve. They reposition the teacher as coach rather than judge, leading to less confrontational classroom environments. Most important, they support the teacher in using assessment to improve learning rather than just to measure it" (p. 42). Given all these strengths, I believe a standards-based assessment and evaluation system will be a superior method for authentically measuring students' achievement.

That is not to say that the shift from traditional grading practices to standards-based assessment and evaluation will be an easy process. Implementing a standards-based system will require some concerted work on the part of schools and teachers. Let us examine some of the ways in which current school and classroom practices may have to be modified to make standards-based assessment a meaningful alternative to traditional methods of calculating letter grades.

First, curricula will need to be restructured. Brouwer (2007) advises that teachers interested in ensuring students understand "key concepts" or "big ideas" begin by clearly identifying these key concepts or big ideas (p. 7). In other words, the content to be taught will have to be restructured from the traditional "topic" structure into content standards.

This process will trim the breadth of content a teacher is able to "cover," but this is in line with the AAAS recommendation in *Benchmarks* that science teachers "radically reduce the sheer amount of material now being covered" (1993, p. XI). This sentiment is echoed by Bybee & Van Scotter (2007), who believe "science curriculums should focus on fundamental scientific concepts and inquiry abilities and develop them in depth" (p. 45).

Secondly, teachers will need to approach assessment differently when using a standards-based assessment and evaluation system. Teachers developing standards-based courses might benefit by planning units using "backward design" as explained by Wiggins & McTighe (2005), by which assessments are actually planned prior to teaching. They encourage teachers to think of assessment more broadly than just tests and instead seek multiple ways of uncovering students' knowledge, understanding, and skills. Offering a variety of assessments to meet individual student needs is in line with recommendations by Kluth & Straut (2001), McTighe & O'Connor (2005) and Tomlinson (2003).

A key to the success of a standards-based system is the embrace of formative assessment by teachers. Many studies (Black, Harrison, Lee, Marshall, & Wiliam, 2004; Clymer & Wiliam, 2007; Sato & Atkin, 2007; Tomlinson, 2008; Wiggins & McTighe, 2005; Wiliam, Lee, Harrison, & Black, 2004) provide evidence for the benefits of teachers using formative assessment practices and real feedback to improve student achievement. Stiggins (2005) recommends teachers use rubrics for providing feedback, and Tomlinson (2008) advises teachers to make liberal use of written and verbal comments to students for the purpose of feedback on their work. I would encourage teachers to consider both of these strategies concurrently!

As students act on feedback and revisit their work, teachers must reassess their students' mastery of the content standards, recording their changing levels of knowledge, understanding, and skill. As Wiggins & McTighe (2005) put it, "The assessment of understanding should be thought of in terms of a collection of evidence over time instead of an 'event'—a single moment-in-time test at the end of instruction—as so often happens in practice" (p.152). Brouwer (2007) too would encourage teachers to think this way, explaining:

> Many [teachers] probably think of assessment primarily as something we do at the end of a lesson, unit or quarter to measure the learning that has or has not taken place. …We need to think of assessment in much broader terms. We need to think in terms of an ongoing process that not only measures learning but also supports and encourages it. (p. 6)

Taking the advice of McTighe & O'Connor (2005), teachers using a standards-based assessment system must allow new evidence to replace old evidence. As students demonstrate their developing achievement of understanding and proficiency of skill, teachers must take note, and give further feedback.

Finally, students—and their parents—will have to be trained to think differently about the purpose of assessment. Students who are familiar with and comfortable with traditional grading practices will need to be educated in the true purpose of assessment: finding out what they know, understand, and are able to do. Students will have to be taught to take a more vested interest in their own educational process. McTighe & O'Connor (2005) advise teachers to give students "opportunities to act on the feedback—to refine, revise, practice, and retry" (p. 16). This will likely be unfamiliar territory for

many students, who are accustomed to traditional grading practices, in which they have only one chance to demonstrate what they know. This is what Kohn (1993) is getting at when he states, "When we are working for a reward [such as a letter grade], we do exactly what is necessary to get it and no more" (p. 63). Students may have to be convinced that the opportunity to rework based on teacher feedback will indeed benefit them.

The shift from traditional grading practices (i.e. the method of "averaging" points to calculate letter grades) to a viable alternative (i.e. a standards-based assessment and evaluation system) will take time and effort on the part of teachers and schools. This shift however, is likely to benefit educators and their students with greater information about the academic achievement of the students.

Methods

Participants

For this action research, I decided to use two intact groups of 8th grade science students I teach for a total of forty-two participants. I teach at a covenantal Christian school located in a small town in rural northwestern Iowa. Our school has approximately 360 students in kindergarten through 8th grade. Students at this school are primarily from white, middle-class families. The parents of a majority of students at this school are college-educated and place a great value on the importance of a high-quality, Bible-based, Christ-centered education. I have observed that many of these parents want to see their children succeed in school. This led me to believe most parents would be amenable to their children participating in a study aimed at increasing student achievement. After

informing parents of my research plan, I received consent from parents of forty-one students to participate in my study.

I decided to use my 8th grade science students as participants in this study because I already know them quite well, having taught them for one school year. As a group, I have found this class of students to be very enthusiastic about trying new things, which, in my mind, made them good candidates for participating in this study. Additionally, I do quite a bit of inventory work in 7th grade (particularly at the beginning of the school year) to learn their strengths and weaknesses, intelligences, and learning styles. As I planned to conduct my research during the first 9-week marking period of the school year, 8th graders seem to be a good choice for subjects in this study.

Materials

A case study design is the most appropriate methodology for this study, as I intend to present a holistic picture of the results. Cresswell (2007) and Merriam & Associates (2002) indicate that case study research is appropriate for a "bounded system," of which my classroom would be an example. The participants of this study (the intact groups of 8th graders I teach) comprise an "accessible case" (i.e., one readily available for the portrayal of the problem), which Cresswell (2007) indicates as an example of a purposeful sample, valid for the subject of a case study (p. 75). The question of generalizability is a legitimate one, but Merriam & Associates (2002) remind us, "What we learn in a particular case can be transferred to similar situations. It is the reader, not the researcher, who determines what can apply to his or her context" (p. 179).

I kept a research journal throughout the term of this study to record my observations and help me develop my case study. This journal was used to keep track of my observations, thoughts, reflections, joys, and frustrations. In it, I noted progress of the research, specific comments and questions from students, observations of students at work, and records of student achievement. Use of such a research journal in case study research is advocated by Yin (2003) in order to collect data and participant-observations.

The use of multiple data sources in case study research is strongly advocated by Cresswell (2007), so, in order to supplement my own observations, I developed two brief surveys to collect data regarding students' and parents' views of this standards-based assessment and evaluation system. The surveys are a modified version of the survey used by Xue, Meisels, Bickel, Nicholson, & Atkins-Burnett (2000). The survey they developed was carefully analyzed for reliability and validity: it was divided into four subscales, after which Cronbach alphas and inter-subscale correlations were computed to test internal reliability of each subscale. Further, descriptive statistics were calculated for each subscale, in order to establish a consistent validity of the measures (Xue, et al., 2000, p. 12). Both the student survey and the parent survey I used are comprised of modified Likert-scale items, as well as a few constructed response items. The student and parent surveys contain similar items; the student version is available as Appendix A and the parent version can be found in Appendix B. To get a sense of the readability of the Likert-scale items, I piloted the student survey with a recent graduate of SCCS, and the parent survey with her mother. Both daughter and mother judged the instruments easily readable and both expressed intrigued curiosity about the new system, which encouraged me.

Procedure

I developed the standards-based assessment and evaluation plan described above during the summer of 2007. I began by listing all of the topics and concepts I have traditionally taught in 8th grade science during the first quarter of the school year. Referring to the guidelines outlined in the *National Science Education Standards* (NRC, 1996) and *Benchmarks for Science Literacy* (AAAS, 1993), I spent about eight hours analyzing this four-page listing, eventually distilling ten content standards. These ten were the "big ideas" I would teach my students during the first quarter. (The content standards are available in Appendix C.)

Using these ten content standards, I began planning out how I would teach the content they define during the first nine-week quarter of the year. I employed backward design as described by Wiggins & McTighe (2003): mapped out goals, prepared assessments of these goals, and planned teaching strategies to ensure student mastery of the stated goals. I used these ten content standards to guide my planning of the assessment vehicles to be used throughout the quarter to assess students' understandings of the "big ideas." I developed a variety of formal and informal assessments to be used throughout the quarter, from small-group research reports to differentiated writing assignments to content quizzes to informal journal responses. I also designed a final exam of sorts to give students a final opportunity to demonstrate what they had learned throughout the quarter. (The exam was comprised of nine short essay questions addressing the first nine content standards. Standard #10 is about writing in science, which was assessed by my reading their responses to the nine questions. Students had several class periods at the end of the quarter as well as study hall time if needed to

complete the exam. The exam questions can be found in Appendix D.) After designing an assessment plan for the content standards, I began planning actual learning activities, including inquiry projects, lectures, readings, discussions, and lab activities.

School started in late August. The night before school started, we had an open house, during which I met with the parents of my 8th grade students to explain my rationale for moving to a standards-based assessment system, and how their children would now be assessed. I also used this meeting as an opportunity to solicit permission for student participation in my study. On the first day of school, I spent the entire science class period explaining the new assessment and evaluation plan to my students. Over the days and weeks that followed I continued to answer questions and reassure my students, as some had very real concerns about how their effort would be rewarded, similar to the reaction predicted by Kohn (1993) when and if a teacher would stop assigning traditional letter grades. While I haven't written letter grades on student papers for several years, the questions and comments I heard from some of my students indicate it was a surprisingly radical shift for some students to not see percentages (e.g., 90%) or even raw scores (e.g., 18/20) on their papers.

When students submitted work, I would assess it according to the standards I had developed using the four-step scale I described earlier (beginning—developing—proficient—advanced) and give written comments as to the strengths and weaknesses of their work. Often times I would assess students' mastery of multiple standards on a single assessment vehicle. Instead of a traditional "score" at the top of their paper, they might see something more like Figure 1.

Figure 1 – A sample student assessment score.

Body Systems Project ___*Student's Name Here*___

Standard #3: Proficient
Standard #7: Developing
Standard #10: Proficient

I also had to develop a new sort of gradebook. Traditional gradebooks are simply not designed to show growth over time. I designed a gradebook using spreadsheet software. My gradebook design was heavily based on recommendations and samples from Marzano (2000) and Clymer & Wiliam (2007). Each student's understandings of the ten standards were recorded on a separate page of the spreadsheet. Each assignment was recorded in a separate row and assessed for one or more standards, which were listed across as columns. To make the recording easier, I used numbers to represent the different achievement levels: 1 = beginning, 2 = developing, and so on. I decided that "half-scores" were reasonable; if students were *very* close to achieving a new level, I did allow a ".5" score. For instance, if a student was close to approaching proficiency, I might assess their achievement at 2.5 rather than 2. At the top of each column, I included the student's current assessed achievement for that standard. As the quarter went on, these levels would change to reflect growth and change over time. I included averages for each standard as well, just for the sake of comparison. (I found it interesting that the "average" often did *not* match the current assessed achievement; it is important to remember that "averaging" grades does not inform us much about a student's developing level of achievement.) Figure 2 is a sample page from my gradebook.

In order to share students' developing understandings of the standards with them and their parents, I developed an online progress report. First, I assigned each student a

Figure 2 – A sample gradebook spreadsheet page

Student's Name										
Assessed Level:	4	4	3.5	4	4	3	4	3	3	3
Standard #	1	2	3	4	5	6	7	8	9	10
Average:	3.7	3.4	3.2	3.3	3.4	3	3.2	3.2	3	2.8
Body Systems Project			3				3			2
Quiz - Introduction to Body Systems			2.5				3			
Image-Bearer Summary	3									3
Nutshell - Energy and Raw Materials				3	3					
Food Labs								3		
Image-Bearer Summary, take 2	4									3.5
Image-Bearer Summary, take 3										
Nutrient R.A.F.T.				4	4					3
Quiz - Nutrients					4			3	3	
Nutshell - Most imp. digestive organ		3					3			
Quiz - Digestion		2	3				3			
Requiz - Nutrients (or) Digestion										
Counting Grass Lab									3	2.5
Quiz - Bone Names			4							
Nutshell - Bones - Living?		4								
Quiz - Muscle Names			4							
Bones & Muscles Summary		4	3				4			
Nutshell - Kidneys						3				

random identification number using a random number generator at www.random.org. I added a summary page to my gradebook spreadsheet to facilitate creating the online report. The report was a simple webpage with a listing of the ten content standards, an explanation of the achievement levels (1 = beginning, 2 = developing, etc.), and a table listing students' achievement of the standards. I used students' randomly assigned identification numbers to help protect their privacy. Figure 3 is a partial progress report.

After developing this online report format, I emailed parents their child's identification number and some instruction on how to read and use this report for information about their child's achievement in science. I updated the report at least every two weeks, for a total of six reports during the first nine-week quarter of the school year.

Figure 3 – A partial online progress report

The "Big Ideas" for SCCS 8th Grade Science, 1st Quarter

1) Human beings are created in the Image of God!
2) Living things (including human beings) are designed with order.
3) Living things have a particular structure with associated functions.
4) Human beings (like all organisms) need energy, and have structures to get energy.
5) Human beings (like all organisms) need raw materials, and have structures to get them.
6) Human beings (like all organisms) produce wastes, and have structures to remove them.
7) Human body systems perform particular functions working together for the good of the whole person.
8) Human beings have a responsibility to care for their bodies–including proper nutrition and exercise.
9) "Science" is an organized, but flexible, way of exploring God's world, including our own bodies.
10) Part of "doing science" is clearly communicating what you've discovered to others.

In the table below, the numbers represent students' achievement of the "big ideas" listed above.
1 = "beginning"
2 = "developing"
3 = "proficient"
4 = "advanced"

ID #	1	2	3	4	5	6	7	8	9	10
1273	3	2	3	3	2.5	3	3	3	3	2.5
1292	3.5	3.5	3.5	3	3.5	3	3	3	3	3
1517	4	3	3	3	3	2.5	3	2	3	3
1806	4	3.5	3.5	3.5	3.5	3.5	3	4	4	4
2971	3	3	4	3.5	3.5	3	3	3	3	3
5261	4	4	4	4	4	4	4	4	4	4
5284	3.5	3.5	3.5	3	3	3	3	2.5	3.5	3

Finally, at the end of the first quarter, it was time to compose report cards. Students' report cards for science included a listing of the ten standards with their level of achievement noted for each and a few comments about their work habits and classroom demeanor. I would have preferred to simply have this listing of students' mastery of the standards represent their achievement in science. However, though it seemed less than authentic after not using letter grades at all for nine weeks, I did assign my students a final grade for the quarter. As Marzano (2000) eloquently explains, "Given that overall letter grades…are so ingrained in our society, it is probably best not to do away with

them at this time. That is not to say that they have merit but, rather, that a school…will probably meet a great deal of resistance if it attempts to suddenly [stop using letter grades]" (p.109).

In the system I've developed, each student's grade in science is an aggregate of his or her final content standard evaluations at the end of the quarter. In essence, their grade is a score out of 40 possible points (10 standards x 4 achievement levels) for the quarter. While assigning a percentage grade still reflects traditional grading practices, in this assessment and evaluation system, grades I assign will take on specific meaning and purpose that will inform students—and their parents—of their achievement in science.

At the outset of this new adventure in assessment, I assured my students that if they achieved "proficient" understanding of all ten standards, they would not score less than a B+. Thus, I used the benchmark of 30/40 being assigned "B+" as an anchor in developing a grading scale that seemed appropriate for this system. Figure 4 is the grading scale I used to assign grades.

Figure 4 – Grading scale used in this project

A	36 to 40	C	20 to 22.4
A-	33 to 35.9	C-	17.5 to 19.9
B+	30 to 32.9	D+	15 to 17.4
B	27.5 to 29.9	D	12.5 to 14.9
B-	25 to 27.4	D-	10 to 12.4
C+	22.5 to 24.4	F	10

With the planning, teaching, and assessment tasks of this project completed, the time had come for surveying students and parents to get their feedback. I used Zoomerang (www.zoomerang.com), a web-based survey service, to administer the

surveys I had developed. I gave my 8th grade participants some class time to complete their survey. I emailed a link to the parent survey to the parents who agreed to participate. I followed this initial email with two reminder emails over the next week to encourage their participation. Finally, it was time to look at what my students and their parents thought of this new assessment and evaluation system.

Results

Design of the Analysis

I had two primary research questions for this action research project:

1. Will a standards-based assessment system give students and their parents an accurate picture of student achievement in science?
2. Is a standards-based assessment system appropriate and practical for a Christian middle school science program?

As I indicated previously, I made notes of observations and recorded personal reflections in a research journal throughout the project to help answer these questions, but as a participant-observer in this research, I recognize that I might not provide the most objective perspective. I used the student and parent surveys to give a fuller picture of the results of this study.

In order to help answer the first question, I considered items #1 – 10 on both the student and parent surveys. Items #1 – 9 are modified Likert-scale items in which participants are asked to mark their level of agreement or disagreement with given statements. These statements were phrased in such a way as to make a neutral response

the null hypothesis for each item. Item #10 is a constructed response question regarding the benefits and drawbacks of the progress reports.

In order to help answer the second question, I considered items #11 – 18 on both the student and parent surveys. Similar to the first section, items #11 – 17 are modified Likert-scale items with a neutral response as the null hypothesis. Item #18 is a constructed response question meant to elicit feedback about a standards-based assessment and evaluation system in general.

All 41 of my 8th grade participants completed the student survey. 27 parents completed the parent survey. I used simple descriptive statistics to analyze the data from the modified Likert-scale items. (Graphs of these data are available in Appendix E.)

I developed a coding scheme to help analyze the results of the constructed response items. I used the following categories to code participant responses:

1 – A generally positive response

2 – A generally negative response

3 – A neutral response (ambivalent, or no strong preference)

4 – A mixed response (some positives, some negatives; or gave specific comments regarding strengths and weaknesses)

I had originally intended only the first three categories. However, as I reviewed the responses to these open-ended items, it seemed clear that a fourth category was necessary because there were a significant number of responses that fell into this category.

Data Analysis

Let us then consider how the data collected from these surveys might help answer these research questions. My first question is, "Will a standards-based assessment system give students and their parents an accurate picture of student achievement in science?"

Figure 5 displays the statistical description of the first nine items on the student survey. Considering the student survey responses, the progress reports were an effective means of communicating achievement of the standards. In each case, the mean was greater than 3, indicating overall agreement to each item. Further, for every item, the positive responses ("agree" or "strongly agree") outnumber the negative responses ("disagree" or "strongly disagree").

Figure 5 – Student survey – Reponses to items #1 through 9

Survey Item:	#1	#2	#3	#4	#5	#6	#7	#8	#9
1 (Strongly Disagree)	1	1	1	2	1	2	3	3	1
2 (Disagree)	5	7	4	6	7	8	6	6	3
3 (Neutral)	19	17	21	12	17	17	10	18	15
4 (Agree)	11	11	13	17	13	13	11	8	15
5 (Strongly Agree)	5	5	2	4	1	1	11	6	7
n =	41	41	41	41	41	41	41	41	41
mean =	3.341	3.293	3.268	3.244	3.390	3.073	3.512	3.195	3.585

Likewise, the parent survey responses indicate a strong acceptance of the online progress reports for feedback regarding student achievement of the standards. Figure 6 contains the statistical descriptions of items #1 through 9 on the parent survey. Several items had no negative responses at all, and in each case the positive responses far outnumbered the negative. Again, in each case the mean was greater than 3 indicating an overall positive response; the mean for some items was 4, or even greater!

Figure 6 – Parent survey – Responses to items #1 through 9

Survey Item:	#1	#2	#3	#4	#5	#6	#7	#8	#9
1 (Strongly Disagree)	0	0	0	0	0	0	0	1	1
2 (Disagree)	0	2	1	0	1	2	0	3	0
3 (Neutral)	7	5	6	2	2	6	6	7	14
4 (Agree)	18	18	18	17	20	17	17	13	10
5 (Strongly Agree)	2	2	2	8	4	2	4	3	2
n =	27	27	27	27	27	27	27	27	27
mean =	3.815	3.741	3.778	4.222	4.000	3.704	3.926	3.519	3.444

Regarding item #10, the constructed response item related to the feedback given through progress reports, the response was a bit more mixed, but still very positive. All forty-one student participants responded to this item, while thirteen of the twenty-seven parent participants responded. Again, parents had a clearly positive response. Student response was also more positive than negative, but to a weaker degree compared to the parents. Figure 7 displays the coded response rates for item #10.

Figure 7 – Coded student and parent responses to item #10

Student Response to Item #10			Parent Response to Item #10	
Generally Positive Response	14		Generally Positive Response	7
Generally Negative Response	12		Generally Negative Response	4
Neutral Response	9		Neutral Response	1
Mixed Response	4		Mixed Response	1

My second research question, “Is a standards-based assessment system appropriate and practical for a Christian middle school science program?” is by nature a complex question. Based on the student and parent responses to my surveys, the answer is a bit complicated as well.

Figure 8 displays the statistical data from student survey items #11 through 17. Based on their responses to the questions asked in the second half of the survey, it would seem that students did not have a very positive experience with this new assessment and

evaluation system. While neutral responses were still common, negative responses outnumbered positive responses on five of the seven questions in this section. In fact, only two items had a mean greater than 3. The students seem to believe they know what this new assessment system is "all about;" at the same time, they don't seem to like it. I attribute this to the fact that my students have never experienced such a dramatic alternative to the traditional grading practices with which they have grown up.

Figure 8 – Student survey – Reponses to items #11 through 17

Survey Item:	#11	#12	#13	#14	#15	#16	#17
1 (Strongly Disagree)	12	6	3	10	7	4	13
2 (Disagree)	10	7	6	10	15	5	9
3 (Neutral)	10	16	21	12	10	18	7
4 (Agree)	5	7	8	4	6	12	6
5 (Strongly Agree)	4	5	3	5	3	2	6
n =	41	41	41	41	41	41	41
mean =	2.488	2.951	3.049	2.610	2.585	3.073	2.585

Figure 9 displays the statistical data from the second section of the parent survey. Overall, parents responded quite positively to the standards-based assessment system. Item #14 asked parents how their children liked using the standards-based system, and their response to this item mirrors that of their children. Other than this item, the means were greater than 3, indicating a positive response. Also, in each case—excepting item #14—positive responses outweigh negative responses.

Forty of the forty-one students participating in this study responded to item #18. I coded their responses using the same four categories as before. As only six parents responded to item #18, I did not code their responses; there was simply too little data to draw strong conclusions. Based on the coded results to this item, students seem to have mixed feelings about the standards-based assessment system. In general, however, their

Figure 9 – Parent survey – Reponses to items #11 through 17

Survey Item:	#11	#12	#13	#14	#15	#16	#17
1 (Strongly Disagree)	0	0	0	3	0	0	1
2 (Disagree)	4	5	1	5	4	1	4
3 (Neutral)	14	9	10	14	15	7	9
4 (Agree)	7	11	14	3	6	16	11
5 (Strongly Agree)	2	2	2	2	2	3	2
n =	27	27	27	27	27	27	27
mean =	3.259	3.370	3.630	2.852	3.222	3.778	3.333

response is only slightly more negative than positive. Given the fairly negative response given by students in items #11 through 17, I would have expected these comments to be even more negative than they actually are. The coded student responses are displayed in Figure 10.

Figure 10 – Coded student responses to item #18

Student Response to Item #18	
Generally Positive Response	13
Generally Negative Response	15
Neutral Response	8
Mixed Response	4

Discussion

Summary

I believe that my first research question—regarding using a standards-based assessment system to give an accurate picture of student achievement in science—has been answered affirmatively. Both students and parents indicated overall that the progress reports integrated into this standards-based assessment and evaluation system were beneficial and useful for feedback regarding achievement in my science class.

I acknowledge that there were some critical comments about the progress reports from students. That being said, 10 out of the 12 generally negative comments were phrased something like this 8th grader's complaint: "I did not like this at all, I like letter grades a lot more because I can better understand and just know how I am doing." Note that this criticism isn't really about the progress reports themselves, but about the lack of *a letter grade* on the report. This indicates that students believe they understand the meanings of letter grades in a traditional assessment system, and are hesitant to accept something unfamiliar in the place of the letter grade they believe they understand.

Many students and parents had very encouraging comments about the progress reports in response to item #10 on the survey. One student wrote, "I really liked the online progress report. It is so nice to be able to see where I am at, what I can improve on, etc." Another 8th grader commented, "I like the progress reports because it tells you what you need to study more about." A number of parents echoed these students' sentiments with their comments: "I like the more frequent feedback!" and "It's easier to know what the student knows or is struggling with." Overall, a greater number of both students and parents indicate that the progress reports—even without a letter grade—were acceptable and useful for informing them about student achievement in science.

In summary, I acknowledge that there was some strongly critical feedback about the lack of a letter grade on the progress reports, indicating that *some* students and a *few* parents did not fully appreciate the standards-based assessment system as an alternative to traditional letter grades. However, the generally positive response by both students and parents to items #1 – 9 on the survey, as well as the high number of positive comments given to item #10, demonstrate that a majority of students and parents *did*

ascertain an accurate picture of student achievement in science through this standards-based assessment system.

The answer to my second research question—about the practicality and propriety of a standards-based assessment and evaluation system in a Christian middle school science classroom—is not as immediately evident. As I mentioned before, this is a complicated question, and there are several different elements to consider in attempting an answer. I note three elements worth discussing here: 1) the system must not be too cumbersome for teachers to implement, 2) it must promote deep understanding of key curricular concepts, and 3) it must acknowledge the unique gifts and needs of each learner while encouraging growth. These three elements will actually require a change in school culture on the whole to be implemented *completely*, and so I fear that the results of the brief, nine-week scope of my study will merely be instructive for teachers, rather than authoritative. In any case, let us examine how these three elements were exhibited in my implementation of a standards-based assessment system.

First, is a standards-based system too cumbersome for teachers to implement realistically? This is a difficult question for me to answer with certainty. It will surely take some work for teachers to adapt their classroom practices to this system, especially at the outset. It took me a concentrated day to modify one nine-week quarter of my 8th grade science curriculum, boiling it down to just ten content standards. The eight hours or so I worked on this "concentrated day," however, does not account for the time over the past few years I spent thinking about how I might restructure my curriculum to allow students greater understanding of key concepts. Additionally, some teachers might balk at not teaching "for the details," especially if they have been doing so for a long time.

However, the benefit of this process of streamlining the content has paid off for me. Using the standards-based system, my students could know clearly from the first day of school exactly what content they had to master by the end of the first quarter!

My process of marking papers had to change greatly to make this system a reality. At first, it did take about twice as long to mark a stack of papers, mostly because I was not accustomed to writing many comments on students' papers; I had previously just written raw scores or perhaps percents at the top with a generically positive comment such as, "Nice job!" if their work merited it. It took some time to learn how to help students understand what would improve their achievement. As time went on, however, this process became easier—and thus faster—for me. Also, I had worried at first that the categories of "beginning," "developing," "proficient," and "advanced" might prove difficult to discern. I found, to my delight, that using these categories was actually *easier* for marking papers than using percents. As I read a student's work, I could quite easily judge how well they met the standard. If their understanding was accurate, and explained clearly, their work was "proficient." If it was not yet "proficient," it was usually quite clear whether their level of understanding was "developing," or just "beginning." And if their understanding was clearly beyond my expectation for a typical 8th grader, it was easy to mark the work "advanced." While it does take a few more minutes to get through a stack of student work marking papers this way, I found that writing the number at the top of the paper (1 = "beginning," 2 = "developing," etc.) *along with written comments* gave students much more formative feedback than they had ever had with my previous system, and it gave them much more information about how they can improve. I confess, I am still a little concerned with how students will *apply* this

feedback. Students need to be educated on how to use the feedback and continue to deepen their understandings of the standards. This is something I began to realize toward the end of my study. I would encourage teachers implementing such an assessment system take *plenty* of time to explicitly teach students how to use feedback to revisit their work, as well as providing them with multiple opportunities to do so.

Second, to be a truly appropriate assessment system, it must promote deep learning. In my design of the system, this was one of the key ideas I drew from *NSES* (NRC, 1996) and *Benchmarks* (AAAS, 1993): to limit content and really focus on mastery of a smaller number of key concepts rather than simply "covering" a wide variety of minutia. Comments from parents and students indicate that this was effective and appreciated. One parent wrote, "I like the concept of teaching things that will be retained and not just 'remembered'. Memorization in many ways is a temporary thing and not always a good way to judge a child's learning." Another parent commented likewise, "I am encouraged to see you emphasize the complete understanding of their work, not just memorizing the facts for a test." A student also noted, "I really appreciate that this system focuses on what we KNOW instead of how many points we missed on tests or assignments." Judging from these comments, it seems at least some parents and students understand the real strength of this sort of assessment system; it is designed to measure their current achievement, give feedback on where and how to improve, and—in the long run—encourage deep understanding of fewer topics rather than shallow memorization of many facts, only to forget them as soon as the quiz is over.

Thirdly, to be an appropriate assessment system, it must support the learning needs of all students while encouraging their growth. Many authors advocate

individualizing assessment practices in order to encourage the learning of all their students (Brouwer, 2007; Callahan, 1999; Kluth & Straut, 2001; McTighe & O'Connor, 2005; Tomlinson, 2008; Tomlinson, 2003; Van Dyk, 1995). The written feedback I provided on every assignment students submitted was intended to give them specifically tailored information about the strengths and weaknesses of their work, and how to improve what needed improving.

Allowing "new evidence" to replace "old evidence"—advice explicitly given by Clymer & Wiliam (2007) and echoed by McTighe & O'Connor (2005)—is a key to this. One student commented, "I like it that you get a second chance to prove what you know." He or she clearly picked up on this element of the standards-based system.

A number of students, however, failed to appreciate this. One student wrote, "I didn't really like this grading system because you're not really pressured to learn anything. You can flunk every quiz and get 4's for all the big ideas on the last quiz/test and get an A." Another student wrote, "I do not like this system at all, I need pressure or else I do not do well or do my work. The old system was better because you really had to try to keep your grade up." These comments seem to reflect students' conditioning for "earning" a grade, rather than working for understanding.

These critical comments actually bring up a key point. In light of the mixed response to this standards-based assessment and evaluation system demonstrated by many students and some parents, I find it interesting to note that students performed *very* well during the first quarter of the year. Most students' final quarter grades in science were comparable to their achievement last year, and eleven of the 41 participating students saw *higher* grades during this first quarter than their grade average in science

last year. Bearing in mind that I had set a benchmark of 30/40 (an average achievement of "proficient" for all ten standards) as a "B+", not a single student had a score less than a "B-" on the grading scale I had set. Figure 11 shows a grade distribution for the first quarter of the year in science.

Figure 11 – Final grade distribution for first quarter

Grade	Point Range (out of 40 possible)	Number of Students
A	36 to 40	11
A-	33 to 35.9	10
B+	30 to 32.9	14
B	27.5 to 29.9	5
B-	25 to 27.4	2

As I read through observations recorded in my research journal about conversations with students throughout this study and read the comments written by students and parents, I began to notice a pattern. As I alluded to earlier, most of the complaints with this new assessment system seem to be with the fact that I *wasn't* giving letter grades. Perhaps a few samples will illustrate. A student wrote, "I don't like all the numbers and things you give us like a 3.5 or 4; I just want a grade." Another wrote, "I don't like that it shows a number and not an actual grade," and a third agreed, "I personally like getting a letter grade better." One 8th grade student, obviously disgruntled and impassioned, wrote

> I don't care for the grading system because I like letter grades and they are going to get me into college, not how well I understand it. This system might help some people who don't get that great of grades, but I work hard for my grades and other

> people don't, and I think it isn't fair that people who don't give a [expletive] about grades get the same grades as people who work hard.

It would seem that this student really understands the traditional letter-grade system so prevalent in our schools, and probably knows how to work that system. It is sad that he or she is so much more concerned with the grade, rather than caring whether or not he or she is learning.

Some parents shared these students' sentiments, simply desiring letter grades. One parent wrote, "We live in the letter grade world. I understand where you are going. However, assigning a grade is a part of our system." Another parent echoed this, writing, "I appreciate the time and effort you put into exploring new ways to evaluate. It does seem, however, that they do appreciate the letter grade more. Perhaps because it is what they grew up with—who knows?" This last comment hits at the heart of the matter. Students and parents—and many teachers too—have been conditioned to think that letter grades are the best way to assess and evaluate student learning, because it is all most of us know. Most of the frustration and dislike expressed regarding this standards-based assessment and evaluation system on the part of both students and parents seems to stem from a belief that letter grades are the only way to describe and put a label on what has been learned. A full acceptance by students and parents—and even some teachers—of a standards-based assessment and evaluation will require a distinct shift in school culture away from a competitive, teacher-judging, grade-based assessment mindset to a more collaborative, teacher-guiding, understanding-seeking frame of mind. This will require more education of students, parents, teachers, administrators, board members, and community members regarding the true meaning and purpose of assessment.

My frustration with some students' and parents' love of the letters aside, overall, I would say that a standards-based assessment and evaluation system is indeed practical and appropriate for a Christian middle-school science classroom. One parent's comment summed up my thoughts very well, saying, "I think this process of analyzation [sic] takes time for all of us to figure out. One quarter is not sufficient time to make a complete judgment. I encourage you to keep implementing the system for a time. I really appreciate your effort of having the kids knowing the details AND the big picture of science." I felt the same way at the end of the 9-week marking period; it seemed that I was just starting to understand how this assessment system really works when the planned timeline for this action research came to an end. As such, I have decided to continue experimenting with this assessment system. The nine week scope of this study was simply not long enough for me to completely implement this system, and affect the sort of "cultural shift" required for students and parents—not to mention myself—to fully understand and appreciate this new system.

For instance, one of the ways this assessment system is designed to support students' learning is that it is intended to consistently give students the option and opportunity to demonstrate their advancing understanding. This concept was not fully understood, or at least misinterpreted, as evidenced by this parent's comment:

> A grading system that weights later assignment more heavily than early assignments does not promote learning as you go. Our child made the comment that it made no difference how you did until the last assignment/test. That was the only thing that counted. A stronger emphasis on learning as you go and mastering

> the material along the way, as well as giving credit for end of term knowledge, is needed.

This criticism is constructive, and it actually causes me to wonder if I put too much emphasis on what students had learned by the *end* of the quarter and not enough emphasis on what they were doing along the way. This was surely not my intent, but I can understand how students and parents, who are used to a more traditional grading system, might feel that way. In the future, I will likely give students a separate "effort" score on their progress reports and end-of-quarter report cards to help address this weakness.

As I've continued implementing this standards-based assessment system, I have observed that the climate in my classroom has changed somewhat. My students are growing more comfortable with not seeing "grades" on their papers. Several students who initially resisted the change to a standards-based system have commented to me recently that they have really begun to see the benefit for their understanding in science. I continue to deepen my understanding of how to specifically give students feedback for improvement, and I have found that many students are more likely to accept that they do not totally understand a concept *yet*—and take that information as a challenge to deepen their understanding. I will likely continue to modify aspects to streamline the system a bit as time goes on, but a standards-based assessment and evaluation system will likely be a feature in my science classroom in years to come.

Conclusions

There are surely limitations to the application of this research. This study is the story of my classroom and my experience in implementing a very different assessment system from the one I had used previously. I, as the primary researcher, had a large personal investment in this project. While I have tried to remain as objective as possible, I recognize my sinful human nature and admit that there is some bias. As I am a participant-observer in this action research, it will have limited generalizability, as I have attempted to take a holistic approach to describing my endeavors. Additionally, the classroom atmosphere, student and parent demographic, and school setting will limit the application of this research. Also of note is the fact that I only applied this new system to 8th grade science students; different results might be obtained at different grade levels or in different content areas.

That being said, standards-based assessment and evaluation systems may have wider implications for Christian educators at all grade levels and across content areas, because of the emphasis on meeting the individual needs of the unique image-bearers we teach, and on growth over time. If Christian educators are convinced that the learners in their classrooms are unique in their strengths and weaknesses and gifts and talents and needs, we ought to teach them—and assess them—that way. This venture into standards-based assessment has helped me to better understand where individual students were both struggling and succeeding, and to what degree. Using this system, it was easier to give specific feedback to students regarding what they had mastered and what still needed more attention, giving specific feedback and encouragement. When a student finally mastered a concept with which he or she had been struggling for a time, it was clear to

me that proficiency had been attained, and I was able to celebrate specific successes with the student. Implementing this new assessment system has helped me to better practice what I preach regarding my view of the students I teach.

Science educators in particular—at all grade-levels—may find this study informative. Much research has been done in recent years in the realm of formative assessment and standards-based assessment in science (AAAS, 1993; Clymer & Wiliam, 2007; NSTA, 1998; NRC, 1996; Sato & Atkin, 2007), and this study adds to the research base. I, for one, intend to apply this standards-based assessment and evaluation system to my whole 7th and 8th grade science program, as I have become convinced that this is a superior method for assessing students' developing understandings and measuring their achievement than the more traditional grading methods I had used for so long before this.

A constant challenge for every teacher is assessing what his or her students really know, understand, and are able to do. However, by arranging content by standards, incorporating more formative feedback, and focusing on "meaning over memorization" (Wiggins & McTighe, 2005), teachers can help their students grow academically and attain deeper understanding of key content. Helping our students succeed is, after all, our calling as Christian educators.

References

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*, 86(1), 8-21.

Brouwer, E. (2007). Assessment for learning: A blessing for our students. *Christian Educators Journal*. 47(2), 6-9.

Bybee, R. & Van Scotter, P. (2007). Reinventing the science curriculum. *Educational Leadership*, 64(4), 43-47.

Callahan, C. (1999). Classrooms for learners, not winners and losers. *High School Magazine*, 7(1), 22-26.

Cresswell, J. W. (2007). *Qualitative inquiry & research design: Choosing among five approaches* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Clymer, J. & Wiliam, D. (2007). Improving the way we grade science. *Educational Leadership*, 64(4), 36-42.

Glatthorn, A. (1998). *Performance assessment and standards-based curricula: The achievement cycle.* Larchmont, NY: Eye on Education.

Hein, G. & Price, S. (1994). *Active assessment for active science: A guide for elementary school teachers.* Portsmouth, NH: Heinemann.

Kluth, P. & Straut, D. (2001). Standards for diverse learners. *Educational Leadership.* 59(1), 43-46.

Kohn, A. (1993). *Punished by rewards: The trouble with gold stars, incentive plans, A's, praise, and other bribes.* Boston, MA: Houghton Mifflin.

Marzano, R. (2000). *Transforming classroom grading.* Alexandria, VA: Association for Supervision and Curriculum Development.

Merriam, S. B. & Associates. (2002). *Qualitative research in practice: Examples for discussion and analysis.* San Francisco, CA: Jossey-Bass.

McTighe, J. & O'Connor, K. (2005). Seven practices for effective learning. *Educational Leadership*, 63(3), 10-17.

National Research Council. (1996). *National science education standards*. Washington, DC: National Academy Press.

National Science Teachers Association. (1998). *NSTA pathways to the science standards* (Middle school ed.). Arlington, VA: NSTA Press.

Sato, M. & Atkin, J.M. (2007). Supporting change in classroom assessment. *Educational Leadership*, 64(4), 76-79.

Stiggins, R. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan*, 87(4), 324-328.

Tomlinson, C.A. (2008). Learning to love assessment. *Educational Leadership*, 65(4), 8-13.

Tomlinson, C.A. (2003). *Fulfilling the promise of the differentiated classroom: Strategies and tools for responsive teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.

Vander Ark, D. (2000). *From mission to measurement*. Grand Rapids, MI: Christian Schools International.

Van Dyk, J. (1995). The kids in the middle: Winners or losers? Part II. *Christian Educators Journal*, 34(4), 11-12.

Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education,* 11(1), 49-65. Retrieved July 8, 2007 from EBSCOhost Academic Search Premier database.

Wiggins, G., & McTighe, J. (2005). *Understanding by design* (Expanded 2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Xue, Y., Meisels, S., Bickel, D., Nicholson, J., & Atkins-Burnett, S. (2000). *An analysis of parents' attitudes towards authentic performance assessment.* Retrieved August 1, 2007, from EBSCOhost ERIC database (ED443868).

Yin, R. K. (2003). *Case study research: Design and method* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Appendix A

Student Survey – Attitudes About Mr. Mulder's Grading System

The following statements are about *your* feelings about the progress reports we used in science during the first quarter of this school year. There are no "right" or "wrong" answers. Please read each statement. Circle the number that matches your feeling of agreement or disagreement. All questions must be answered. Thank you.

My progress reports help me:

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1. Understand my strengths.	5	4	3	2	1
2. Understand where I need help.	5	4	3	2	1
3. Understand how well my achievement compares with expectations in science.	5	4	3	2	1
4. Understand how well I am meeting Mr. Mulder's expectations for learning in science.	5	4	3	2	1
5. Understand the progress I am making in science.	5	4	3	2	1
6. Understand how Mr. Mulder is helping me learn.	5	4	3	2	1
7. Understand how well I am doing overall in science.	5	4	3	2	1
8. Think about improving my work.	5	4	3	2	1
9. Take pride in my work.	5	4	3	2	1

10. Comments regarding the *progress reports*:

The following statements are about *your* feelings about the standards-based grading system we used in science during the first quarter of this school year. There are no "right" or "wrong" answers. Please read each statement. Circle the number that matches your feeling of agreement or disagreement. All questions must be answered. Thank you.

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

11. Compared to typical report card letter grades, I like the standards-based system better.

5 4 3 2 1

12. The standards-based grading system helps me know more about my schoolwork than Mr. Mulder's previous system.

5 4 3 2 1

13. I think the standards-based grading system helps me understand what I'm learning.

5 4 3 2 1

14. I like using the standards-based grading system.

5 4 3 2 1

15. I would recommend a standards-based grading system to other teachers or other schools.

5 4 3 2 1

16. I feel that I understand what Mr. Mulder's new grading system is all about.

5 4 3 2 1

17. If given the choice, I want Mr. Mulder to continue using this standards-based system instead of Mr. Mulder's previous system.

5 4 3 2 1

18. Comments regarding Mr. Mulder's new *assessment plan* in general:

Appendix B

Parent Survey – Attitudes About Mr. Mulder's Grading System

The following statements are about *your* feelings about the progress reports we used in science during the first quarter of this school year. There are no "right" or "wrong" answers. Please read each statement. Circle the number that matches your feeling of agreement or disagreement. All questions must be answered. Thank you.

My child's progress reports help me:

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1. Understand my child's strengths.	5	4	3	2	1
2. Understand where my child needs help.	5	4	3	2	1
3. Understand how well my child's achievement compares with expectations in science.	5	4	3	2	1
4. Understand how well my child is meeting Mr. Mulder's expectations for learning in science.	5	4	3	2	1
5. Understand my child's progress in science.	5	4	3	2	1
6. Understand how Mr. Mulder is helping my child learn.	5	4	3	2	1
7. Understand how well my child is doing overall in science.	5	4	3	2	1

My child's progress reports help my child:

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
8. Think about improving his/her work.	5	4	3	2	1
9. Take pride in his/her work.	5	4	3	2	1

10: Comments regarding the *progress reports*:

The following statements are about *your* feelings about the standards-based grading system we used in science during the first quarter of this school year. There are no "right" or "wrong" answers. Please read each statement. Circle the number that matches your feeling of agreement or disagreement. All questions must be answered. Thank you.

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

11. Compared to typical report card letter grades, I like the standards-based system better.

5 4 3 2 1

12. The standards-based grading system helps me know more about my child's schoolwork than Mr. Mulder's previous system.

5 4 3 2 1

13. I think the standards-based grading system helps my child understand what he/she learning.

5 4 3 2 1

14. My child likes using the standards-based grading system.

5 4 3 2 1

15. I would recommend a standards-based grading system to other teachers or other schools.

5 4 3 2 1

16. I feel that I understand what Mr. Mulder's new grading system is all about.

5 4 3 2 1

17. If given the choice, I want Mr. Mulder to continue using this standards-based system instead of Mr. Mulder's previous system.

5 4 3 2 1

18. Comments regarding Mr. Mulder's new *assessment plan* in general:

Appendix C

Content Standards Developed for Use in this Study

The "Big Ideas" for SCCS 8th Grade Science, 1st Quarter

1) Human beings are created in the Image of God!

2) Living things (including human beings) are designed with *order*:

    - Cell – tissue – organ – system – organism – population – ecosystem

3) Living things have a particular *structure* with associated *functions.*

4) Human beings (like all organisms) need *energy,* and have structures to get energy.

5) Human beings (like all organisms) need *raw materials,* and have structures to get them.

6) Human beings (like all organisms) produce *wastes,* and have structures to remove them.

7) Human body systems perform particular functions working together for the good of the whole person:

    - Skeletal/muscular: support and movement
    - Digestive: gaining nutrients for energy and raw materials
    - Circulatory: transportation
    - Respiratory: gaining oxygen and removing carbon dioxide
    - Excretory: removing wastes

8) Human beings have a responsibility to care for their bodies—including proper nutrition and exercise.

9) "Science" is an organized, but flexible, way of exploring God's world, including our own bodies.

10) Part of "doing science" is clearly communicating what you've discovered to others.

Appendix D

Final "Exam" Questions

1. We've found frogs are quite similar to human beings. How similar are they? Are frogs created in God's Image? Why or why not? (Standard #1)

2. You dissected several organs as you dissected your frog. Describe the *order* designed into your frog (from cells to tissues to organs to systems to the whole organism) as specifically as you can. (Standard #2)

3. Choose one organ or organ system from the frog's body. Describe how it is designed (its *structure*). Then describe how it works (its *function*). (Standard #3)

4. Your body is designed with specific structures to get energy, including your respiratory system (to get oxygen you need in order to do cell respiration). Compare the frog's respiratory system to your own. How are they similar? How are they different? (Standard #4)

5. Your body is designed with specific structures to get the raw materials you need, including your digestive system. Compare the frog's digestive system to your own. How are they similar? How are they different? (Standard #5)

6. Your body is designed with specific structures to get rid of wastes, including your excretory system. Describe how the frog gets rid of wastes. How is that like your excretory system? How is it different from your excretory system? (Standard #6)

7. We discussed several body systems this quarter, including the skeletal, muscular, digestive, circulatory, respiratory, and excretory systems. Describe how *<u>at least</u>* two of these systems work together for the good of the whole person. (Standard #7)

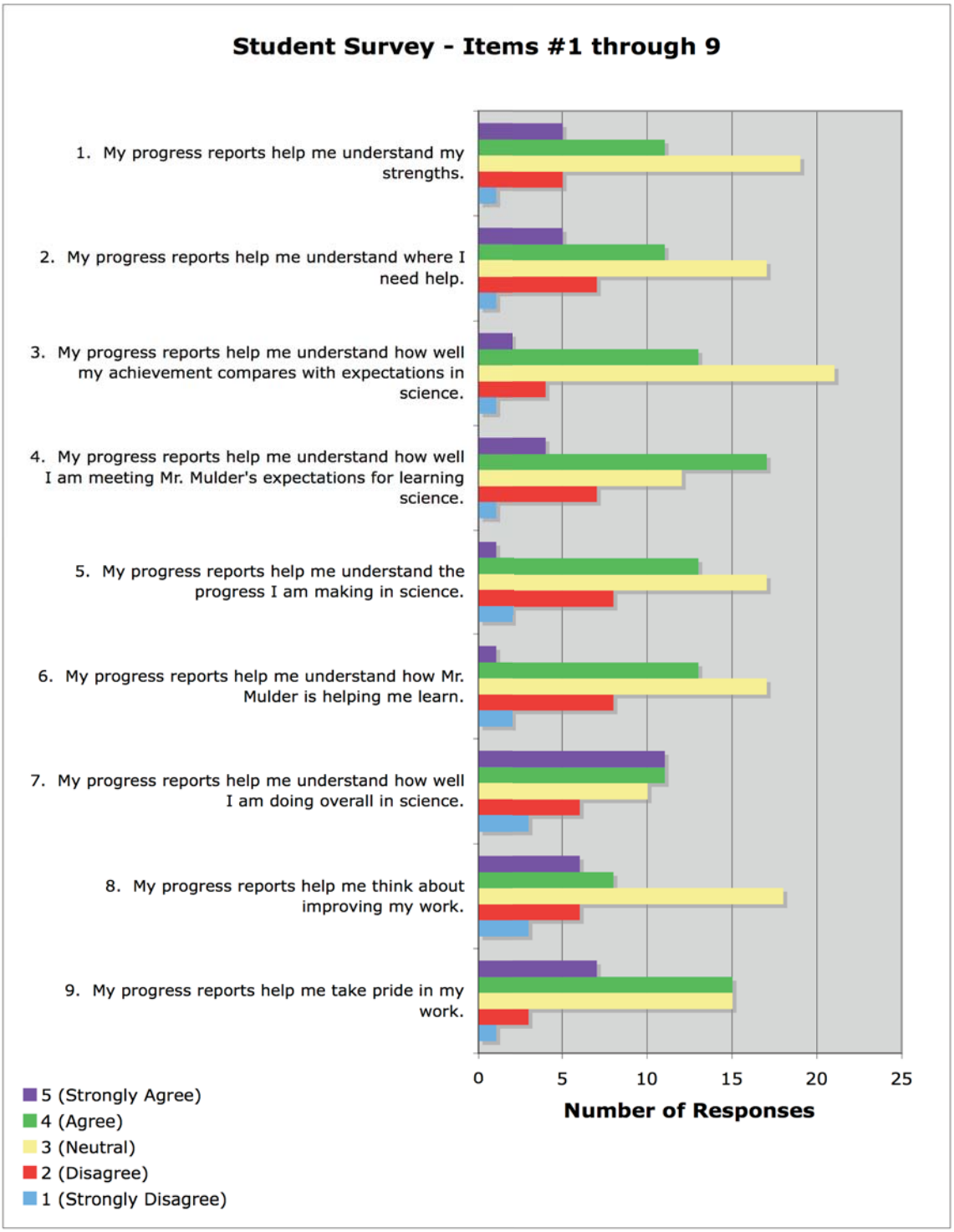
8. What responsibility do you have to take care of your body? What can (and *should*) you do to keep healthy? (Standard #8)

9. How was dissecting your frog an example of "doing science"? (Standard #9)

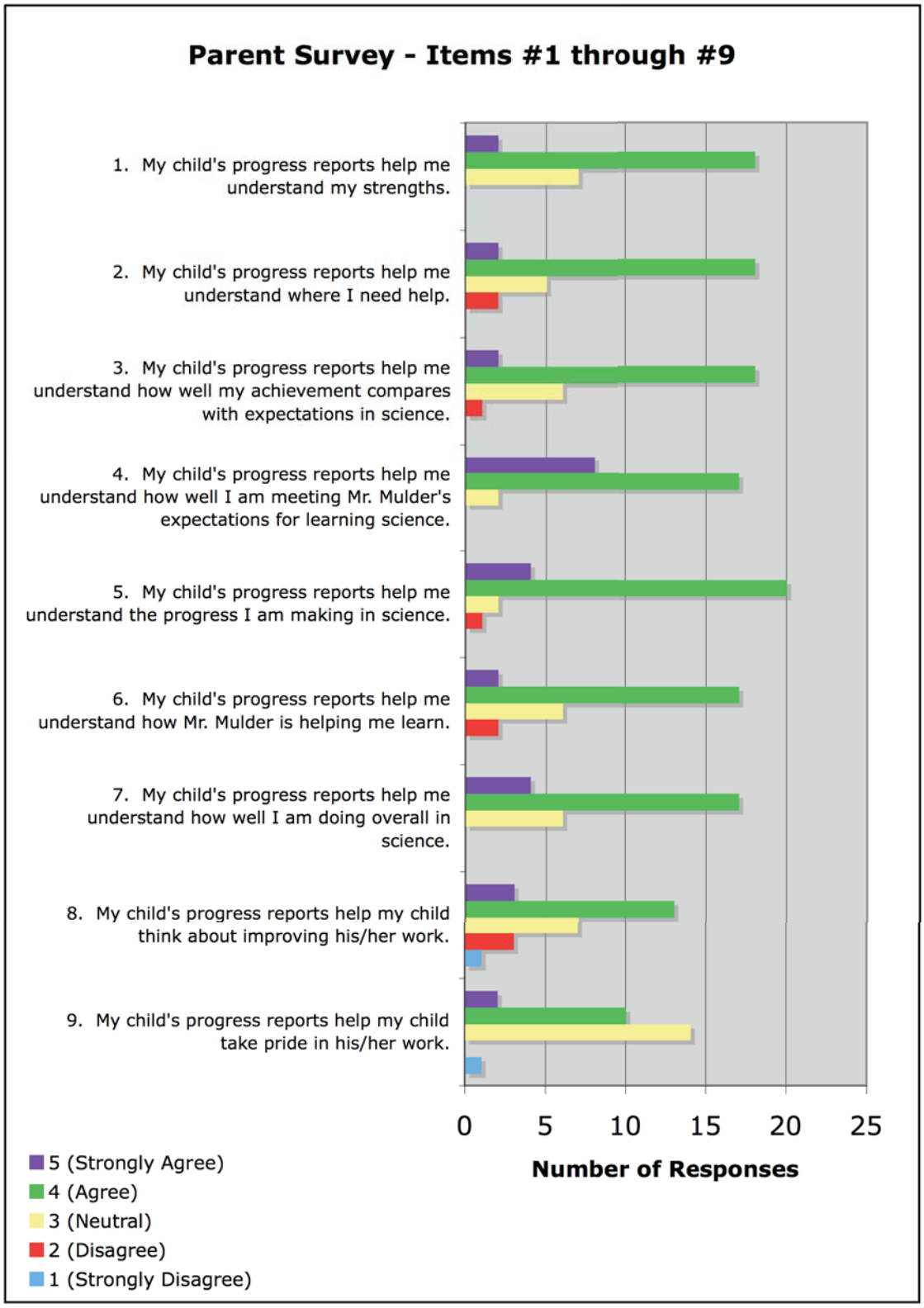
Appendix E

Graphs of Student and Parent Survey Responses

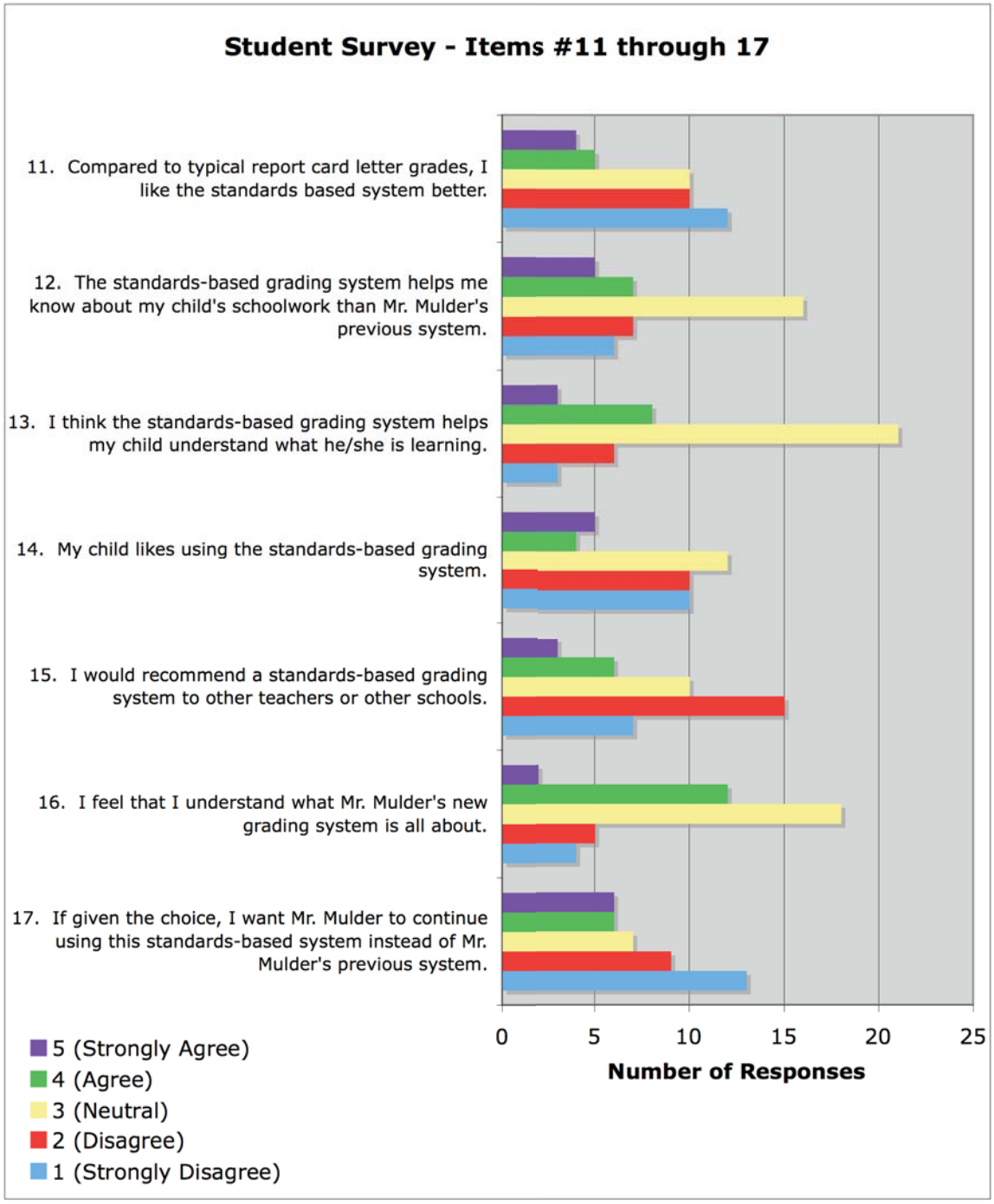
Results of Student Survey, Items #1 through 9:



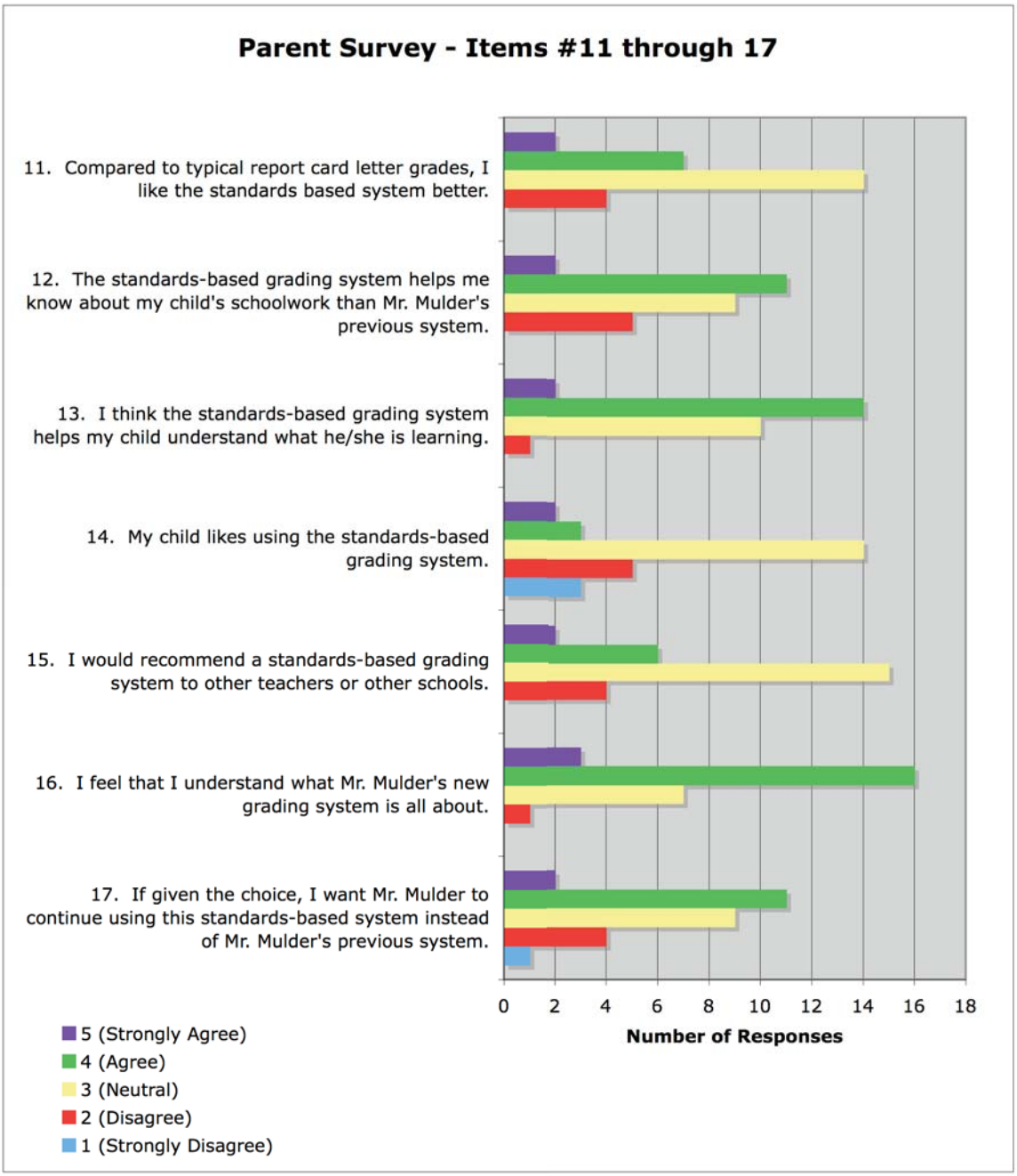
Results of Parent Survey, Items #1 through 9



Results of Student Survey, Items #11 through 17:



Results of Parent Survey, Items #11 through 17:



Acknowledgments

Many people have had a hand in the completion of this project, and their contributions must be acknowledged.

First of all, thanks to my advisor for this project, Dr. Ed Starkenburg, who helped me shape the research questions investigated in this study. Also, thanks to the Graduate Program Director, Dr. Pat Kornelis, whose gentle encouragement and advice about the research helped keep my work focused and on track. Their assistance in reflecting on the outcomes and shaping the resulting report has also been greatly appreciated.

I am extremely grateful to Tim Van Soelen for the time he spent coaching me on research design. His specific advice on case study methodology and data collection tools, as well as the analysis of resulting data was incredibly valuable.

I am deeply indebted to Dr. John Van Dyk, whose influence on my thinking about teaching, learning, and assessment has been both profound and pervasive. The spark of this project began years before the research commenced as he encouraged me to approach my classroom practice seeking to become, more and more, an "authentically Christian teacher."

I am most appreciative of my wife, Melissa, who has lovingly—and patiently—supported and encouraged me through the duration of this project, proofreading the manuscript several times, offering advice, and generally keeping me sane. I love you.

Finally, ultimate thanks to my Lord and Savior, Jesus Christ, to whose honor and glory my work must be done!

Department of Education
Dordt College
Sioux Center, Iowa

VITA

Name: David J. Mulder Date of Birth: May 24, 1976

Home Address: 525 2nd Ave. NE, Sioux Center, IA 51250

Education

Dordt College, Fall 1994 – Spring 1998.
B.A. Elementary Education

Academic Experience

Adjunct Instructor of Education, Dordt College, Sioux Center, Iowa.
Fall 2006 – Present

- Courses taught:
  - Methods for Teaching Science in Elementary and Middle Schools (Fall 2007 – Present)
  - Methods for Teaching Social Studies in Elementary and Middle Schools (Fall 2006)

Teacher, Sioux Center Christian School, Sioux Center, Iowa.
August 2001 – Present

- Teach two sections of 8th grade science, two sections of 7th grade science, and one section of 7th grade Bible in a junior high setting of approximately 90 students.
- Design, implement, and assess units of study.
- Collaborate with colleagues in planning and presenting multidisciplinary units.
- Lead staff development in curriculum and instruction.
- Co-founded and continue to lead a voluntary professional reading and discussion group.
- Chairperson of the Science Curriculum Committee, Fall 2007 – present.
- Chairperson of the Chapel Committee, Fall 2003 – Spring 2007.

Teacher, Calvin Christian School, Escondido, California.
September 1998 – June 2001

- Taught two sections each of 6th, 7th, and 8th grade mathematics in a junior high setting of approximately 160 students.

Publications

Mulder, D. J. (In Press). *Technology Connects: Chemistry (Level I).* Albuquerque, NM: Gravitas Publications.

- A supplemental workbook for the Chemistry (Level I) curriculum. Offers explanations of technological applications of chemistry topics studied in the general chemistry curriculum.

Presentations

Mulder, D. J. (October 2007). "Don't Study Science".

- An examination of current thinking of "best practices" in science education.
- Presented at Heartland Christian Teachers' Convention at Dordt College

Mulder, D. J. (October 2007). "Ahab vs. Elijah: A Backward-Designed, Differentiated Unit Plan".

- Aimed at encouraging teachers to employ backward design and differentiation strategies in their classrooms, I shared a Bible unit I teach to my grade 7 students from planning to assessment.
- Presented at Heartland Christian Teachers' Convention at Dordt College

Mulder, D. J. (August 2007). "No Boredom Allowed".

- An exposition of current research on sources of boredom in classrooms, and strategies teachers can employ to reduce student boredom.
- Presented at a joint inservice meeting for teachers from Sheldon, Sanborn, and Ocheyeden Christian Schools.

Mulder, D. J. (August 2007). "Meeting the Needs of Diverse Learners".

- A full-day workshop designed to help teachers explore and understand the unique learning profiles of their students, and tailor instruction to those profiles through differentiated instruction.
- Presented at Cono Christian School, Walker, IA.

Mulder, D. J. (October 2006). "No Boredom Allowed".

- An exposition of current research on sources of boredom in classrooms, and strategies teachers can employ to reduce student boredom.
- Presented at Heartland Christian Teachers' Convention at Dordt College.

Crull, S. and Mulder, D. J. (October 2005). "Keep It Cool: an Interdisciplinary Math and Science Unit".

- A summary of a multidisciplinary integrated unit prepared and taught by junior high teachers at Sioux Center Christian School during the Spring of 2005.
- Presented at Heartland Christian Teachers' Convention at Dordt College.

Mulder, D. J. (March 2005). "A Response to Arthur Jones' Keynote: Problems and Challenges in Science Education".

- Presented at B.J. Haan Educators' Conference, Dordt College.

Mulder, D. J. (March 2005). "Food Labs".

- An action lab presented to fellow science teachers.
- Presented at B.J. Haan Educators' Conference, Dordt College.

Professional Memberships

National Science Teacher's Association, Fall 2004 – Present