

**NORTH DAKOTA ALTERNATE ASSESSMENT 1
2009-2010**

Technical Manual

North Dakota Department of Public Instruction
April 2010

Copyright © 2010 by the North Dakota Department of Public Instruction, 600 East Boulevard Avenue, Bismarck ND 58505-0440.

Permission to reproduce this material is granted for home, classroom, and workshop use. For all other purposes, please request permission in writing from the North Dakota Department of Public Instruction.

Table of Contents

I. Executive Summary	1
A. Description of Alternate Assessment	4
B. Comparison of the 2007-08 NDAA1 to Prior NDAA1s	9
C. Alignment of NDAA1 to the NDSA	11
D. Standard Setting.....	18
E. Reliability of NDAA1: Improvements	21
F. Reliability of NDAA1: Standard Error of Measurement	23
G. Quality of Submitted Alternate Assessments	24
H. Content Validity: Teacher and Parent Judgments.....	26
I. Construct Validity	29
J. Fidelity of Administration	31
K. Consequential Validity	33
L. Teacher Feedback on the NDAA1.....	34
Appendix A: NDAA1 Forms.....	A-1
Appendix B: Benchmark Maps	B-1
Appendix C: Activity Writing Participants.....	B-1
Appendix D: Activity Writing Training Materials	C-1
Appendix E: Standard Setting Participants.....	D-1
Appendix F: Standard Setting Materials	E-1
Appendix G: Item Statistics	F-1
Appendix H: Executive Summary of NDAA1 Results	G-1

Executive Summary

Various measures of test quality indicate that the ND Alternate Assessment (NDAA1) is a reliable and valid test of student performance on the North Dakota Standards and Benchmarks.

Brief Description of the NDAA1

- The NDAA1 is a selective response-type assessment with situational indicators. Students are assessed on fourteen academic content standards (three reading/language arts, five mathematics, and six science). For each standard, the teacher selects an activity from the NDAA1 list of approved activities that the student will perform. This activity is tied to a particular benchmark within that standard.
- The student performs each activity four times. The teacher indicates the “situation” surrounding each activity by answering eight questions (e.g., in how many settings did the student perform the activity, were the settings integrated, did the student plan and evaluate his/her own performance).
- The student’s score on the NDAA1 is based on how well he/she performed the activity and the situational variables. For example, a student who performs the activity in multiple, integrated settings will receive a higher score than a student who performs the activity in a single, non-integrated setting.
- The NDAA1 is for students with significant cognitive disabilities. The NDAA1 consists of performance activities that are based on alternate achievement standards.

Comparison of 2009-10 NDAA1 to Prior Years

- The 2009-10 NDAA1 is identical to the 2008-09 NDAA1 and is identical to the 2007-08 NDAA1 other than one situational indicator question (did the student monitor his/her performance) being deleted in 2008-09. Prior to 2007-08, various tweaking of the NDAA1 took place to enhance its reliability and validity. However, the essential nature of the NDAA1 has been absolutely consistent over time. It is a selective response-type assessment with consistent components and procedures. While improvements have been made in the various components, these improvements have not invalidated any comparisons of results over time – comparing the results of one year to another is perfectly legitimate.
 - The NDAA1 has always been a test that collects data on student performance on a teacher-selected activity that is tied to the grade-level content benchmarks and collects information on secondary indicators which are based on best practices for students with severe cognitive disabilities.
 - The NDAA1 is submitted via the World Wide Web through a secure website. The assessment has been electronic for three years, but was web-based for the first time in 2007-08. The web version was more successful than previous electronic versions due to being able to have more built-in edit checks and reduce firewall and Macintosh submission issues.

Alignment

- The NDAA1 covers the same content standards and covers varying degrees of complexity as the North Dakota Alternative Assessment 2 (NDAA2) and the North Dakota State Assessment (NDSA) do.
 - The test blueprints indicate the two tests cover the same content standards. Students who take either the NDAA1 or the NDSA receive scores on each of three reading/language arts standards, each of five mathematics standards, and each of six science standards.
 - The complexity ratings indicate that the NDAA1 consists of tasks with a variety of cognitive complexity, just like the NDSA does.

Standard Setting

- A technically-sound process was used in establishing the cut-scores for the NDAA1.
 - This process involved training about 30 teachers and administrators on standard-setting. This standard-setting took place in February 2006 for reading and math and in February 2007, February 2009, and February 2010 for science.

Reliability: Improvements

- The overall reliability of the Alternate Assessments improved in 2005-06 and has remained consistent since that time. The coefficient alphas are above .90 for each of reading, mathematics, and science.
 - Since 2005-06, the NDAA1 consists of more performance tasks, measures more standards, has greater standardization across students, and eliminates teacher error (e.g., designing inappropriate tasks for the student to demonstrate, setting an inappropriate target of performance, creating un-interpretable graphs) to a greater degree.

Reliability: Standard Error of Measurement

- The standard error of measurement of the NDAA1 is reasonably small and ranges from 3.0 percentage points to 4.0 percentage points. (The final scores a student receives are on a scale of 0 to 100 percent.)

Test Quality

- The quality of the submitted Alternate Assessments has increased from 2002 to 2010.
 - Since 2008, less than 1% of tests have had a quality-related problem in terms of teachers not completing information on each of the eight - fourteen performance tasks.

Content Validity

- *Very high percentages of teachers (95-98%) and parents (90-93%) perceive the NDAA1 to be an accurate reflection of student skill level.*
- The percentage of teachers and parents who perceive the NDAA1 to be an accurate reflection of student skill level steadily increased from 2002 to 2005 and has remained fairly steady since then.

Construct Validity

- The relationship between the overall NDAA1 rating and teacher judgment of student skill is moderate and significant. These correlations (around .44-.56) suggest that the assessment is a valid indicator of a student's performance on the assessed benchmarks.

Fidelity of Administration: Number of Unique Activities

- Teachers chose a variety of activities for their students to demonstrate on the NDAA1 which suggests that they are individualizing the NDAA1 to meet the needs of their students. (The purpose of the large number of provided activities is so that teachers can individualize the NDAA1 for each of their student's needs and educational goals. Thus, a variety of activities should have been selected.)

Fidelity of Administration: Accuracy Check

- Because the NDAA1 relies on teachers assigning ratings to certain student behaviors, verifying that the teacher-provided ratings are accurate is an important step in establishing the quality of the NDAA1.
- In general, the detailed information teachers provide on the context for the performance

activity support and confirm the ratings that teachers assigned to the situational questions.

- Compared to teachers who listed fewer settings, teachers who listed more settings tended to assign higher ratings (as should be expected) to the “Number of Setting” dimension.

Teacher Feedback

- When asked if the NDAA1 was easy to administer, 94% said yes.
- When asked if the NDAA1 activities were appropriately linked with the standards, over 90% said yes.
- When asked how the NDAA1 could be improved suggestions included providing more activities and receiving the materials earlier in the fall.
- When asked the best things about the NDAA1 comments included it being easy to administer, the variety of activities, its appropriateness for the students, and that it was aligned to state standards.

Future Activities

- The ND Department of Public Instruction (DPI) will conduct the following activities in 2010-11:
 - Gather fidelity of administration information by observing a representative sample of teachers administer the NDAA1.
 - Collect consequential validity evidence to determine the intended and unintended and desirable and undesirable consequences of the NDAA1.

A. Description of the Alternate Assessment

- *The NDAA1, since its inception, has collected data on student achievement on teacher-selected activities that are tied to grade-level benchmarks and collect information on the secondary indicators in which the student is assessed.*
- *On the 2009-2010 NDAA1, the students demonstrate their skill level on three reading/language art standards, five mathematics standards, and six science standards for grades 4, 8, and 11.*

General Description of the North Dakota Assessment System

The State of North Dakota has established a statewide assessment system that provides for the measurement and reporting of student achievement for all students based on the state's content and achievement standards. The North Dakota State Assessment System consists of the following components.

- *State's standard assessment.* The North Dakota State Assessment (NDSA) is a standard, criterion-referenced assessment for all students, grades 3-8 and 11, in reading/language arts and mathematics, which are aligned to the state's content and achievement standards. The administration of the NDSA transitioned from grades 4, 8, and 12 to grades 3-8 and 11 during the 2004-05 school year. The State initiated science assessments in grades 4, 8, and 11, beginning with the 2006-07 school year. The State has entered into a multi-year contract with CTB/McGraw-Hill, LLC, to provide for the development, administration, scoring, research, technical assistance, and reporting regarding all assessment documents and achievement results.
- *State's standard assessment with accommodations.* The State provides for the appropriate use of accommodations in the administration of the NDSA. The rules for the use of accommodations are prescribed within the Test Coordinator's Manual of the NDSA. The use of the accommodations for students with special needs, when administered within the constructs of allowable rules, provides for equally valid and reliable results and should, therefore, be understood to be an acceptable extension of the use of the State's standard assessment.
- *State's alternate assessment.* The North Dakota Alternate Assessment (NDAA) is for students with *significant* cognitive disabilities and for those with *persistent* learning difficulties. For the purpose of clarity, between the fall of 2005 and 2007, the North Dakota Alternate Assessment (NDAA) included both the 1% and the 2% population in a consistent procedure with differentiated test items and scoring. This NDAA included selected responses to performance-based activities that were either linked (1%) or aligned (2%) to the state achievement standards and benchmarks, and secondary indicators based on best practices for students with severe disabilities. North Dakota no longer uses the language of "linked" or "aligned"; however, in a case where the documentation may still contain that language, the phrase "linked" is synonymous with the 1% and "aligned" is synonymous with the 2%. For the purpose of clarity, the NDAA 1 is based on alternate achievement standards and is for students with severe cognitive disabilities (the "1% test"). The NDAA 2 is based on modified achievement standards and is for students with persistent cognitive disabilities (the "2% test").
- Like the NDSA, the NDAA1 assesses students with significant cognitive disabilities in grades 3-8 and 11 in reading/language arts and mathematics and grades 4, 8, and 11 in science. North Dakota state guidelines and training guides local educators in the proper

administration of the appropriate alternate assessment level. The determination of which alternate assessment level is most appropriate for a student with either a significant or a persistent cognitive disability rests with the student's individualized education program team, as provided for by federal and state law.

- *State's alternate assessment for students with significant cognitive disabilities.*
The North Dakota Alternate Assessment -1 (NDAA1) is for students with *significant* cognitive disabilities. Federal law provides for the administration of an appropriate alternate assessment for students with *significant* cognitive disabilities that is aligned to the state's content standards and *linked* to the state's challenging achievement standards. The use of an alternate assessment for students with *significant* cognitive disabilities is applicable for use under the 1% alternate assessment accountability rule provided for within federal regulation. Within this narrative, the terms "state's alternate assessment for students with significant cognitive disabilities," "state alternate assessment linked to state's alternate achievement standards," "state's linked alternate assessment," and "state assessment for the 1% reporting requirement" are used interchangeably and are to be understood as equivalent terms. This document provides a description of the NDAA1 and its test quality characteristics.

- *State alternate assessment for students with persistent learning difficulties.*
Federal guidance provides for the administration of an appropriate alternate assessment for students with *persistent* learning difficulties that is aligned to the state's content standards and *aligned* to the state's challenging alternate (i.e., modified) achievement standards. The use of an alternate assessment for students with persistent learning difficulties is applicable for use under the 2% alternate assessment accountability rule provided for within federal guidance. Within this narrative, the terms "state's alternate assessment for students with persistent learning difficulties," "state's alternate assessment *aligned* to the state's *modified* achievement standards," "state alternate assessment *aligned* to state's alternate achievement standards," and "state assessment for the 2% reporting requirement" are used interchangeably and are to be understood as equivalent terms. The description and test quality characteristics of the NDAA2 are described in the NDAA2 Technical Manual and are not covered here. The NDDPI has provided an IEP Flowchart of questions and answers to aid IEP teams in making annual decisions on which assessment option a student should participate and NDAA1 and NDAA2 Side by Side Comparison.

The Components of the NDAA1

The NDAA1 has the following components.

- Students are assessed on 8-14 standards.
 - Three Reading Standards (grades 3-8 and 11)
 - Five Mathematics Standards (grades 3-8 and 11)
 - Six Science Standards (grades 4, 8, and 11)

- Students complete one activity for each standard.
 - For each standard, the teacher selects an activity from the NDAA1 list of approved activities that the student will perform. This activity is tied to a particular benchmark within that standard.
 - A student performs three reading/language arts activities, one for each of the three reading/language arts standards; five mathematics activities, one for each

of the five mathematics standards; and six science activities, one for each of the six science standards.

- All activities are linked to the standards/benchmarks and are for students with the most severe cognitive disabilities (the “1%” population for AYP purposes).
- Students perform each activity four times. (The “**4 Performance Points**”)

 - The teacher records student performance on each of the four trials. Performance is measured by the percentage of correct responses. (See Appendix A for a copy of the data sheet.)

Sample Activities for Grade 8 Math

Math Benchmark 2.1: Make observations about relationships between two- and three- dimensional figures e.g., a cube is made with six squares.

1. Given five, three- dimensional shapes and pictures of a variety of three-dimensional shapes, the student will link the correct shape to the appropriate picture (e.g., triangle shape to picture of a pyramid).
2. Given five basic shapes, the student will correctly name each shape.
3. Given five sets of a concrete object and two pictures (one that matches the item and one that does not), the student will indicate which picture goes with the object.
4. Given sets of five solid shapes, the student will sort the shapes into like groups.
5. The student will locate five square items in the school environment.

Four Performance Points				
	Day/Week 1	Day/Week 2	Day/Week 3	Day/Week 4
Date				
# correct				
# possible				
% correct	%	%	%	%

Sample Situational Indicators

1. In how many settings did the student perform this task? (Up to 2 settings per environment.)
0. None 1. One 2. Two 3. Three 4. Four or more
2. Of these settings, how many were integrated with age-level peers?
0. None 1. One 2. Two 3. Three 4. Four or more

- Teachers indicate the “situation” surrounding each activity by answering eight questions. (The “**8 Situational Indicators**”)

 - **Settings:** Number of settings in which student performs skill
 - **Integrated Settings:** Number of integrated settings in which student performs skill
 - **Independence:** Level of choice/independence student exhibits (this is measured by four individual dimensions: does student select, plan, evaluate, and adjust his/her performance)
 - **Supports:** Level of supports the student needs
 - **Assistive Technology:** Level of assistive technology the student needs

- The teacher provides information related to his/her expert judgment on the student’s achievement.
 - For each of the fourteen activities a student performs, the teacher is asked (1) whether he/she thinks student achievement as indicated on the NDAA1 for that activity accurately reflects student achievement on the targeted skill and (2) his/her estimate of student achievement on the benchmark to which the activity is related. The specific questions are:

- How accurately do you (the teacher) think the results on this activity reflect the student's typical performance on this particular skill? (*This item is used for the content validity analysis – see Section H.*)
 - What would you, as the teacher, estimate the student's performance on this benchmark to be (based on the student's performance on this particular skill being assessed **and** the student's performance on other skills related to this benchmark)? (*This item is used for the construct-related validity analysis – see Section I.*)
- Parents complete the Parent/Guardian Validation Survey. For each of the fourteen activities, parents are asked: (1) if they have reviewed their child's NDAA1 for that standard, (2) whether the NDAA1 accurately reflects their child's performance on the identified skills, and (3) in which environments (home, school, community, other) they have seen their child perform the targeted skills. (*These items are used for the content validity analysis – see Section H.*)

After the student has completed the NDAA1, the teacher submits the NDAA1 assessment via the web and mails the Parent/Guardian Validation survey to the North Dakota Department of Public Instruction.

How the NDAA1 is Scored

On each of the activities (three reading/language arts, five mathematics, and six science), a student earns points on both the Performance Points and the Situational Indicators. The 8 Situational Indicators are worth up to 17 possible points:

- **Settings:** Number of settings in which the student performs skill -- *worth up to 3 points*
- **Integrated Settings:** Number of integrated settings in which the student performs skill - *worth up to 3 points*
- **Independence:** Level of choice/independence student exhibits (this is measured by four individual dimensions: does student select, plan, evaluate, and adjust his/her performance) -- *worth up to 4 points, 1 point for each dimension*
- **Supports:** Level of supports the student needs -- *worth up to 4 points*
- **Assistive Technology:** Level of assistive technology the student needs -- *worth up to 3 points*

A student also earns "points" for the Performance Points that are based on their average percentage correct score for the activity. The Performance Points are weighted against the total possible points of the Situational Indicators. The Performance Points receive 56% of the weight, and the Situational Indicators receive 44% of the weight.

The total percentage of points a student earns across the three Reading Standards, across the five Mathematics Standards, and across the six science standards is calculated. This percentage of points for each of reading, mathematics, and science is translated to proficiency levels. Display A-1 indicates how the percentage translates to proficiency level.

**Display A-1: Proficiency Levels for the NDAA1
Percentage of Total Points a Student Must Earn for a Given Proficiency
Level**

	Reading	Mathematics	Science Grade 4	Science Grade 8	Science Grade 11
Novice	0-49%	0-49%	0-48%	0-49%	0-44%
Partially Proficient	50-65%	50-65%	49-67%	50-74%	45-69%
Proficient	66-89%	66-89%	68-91%	75-98%	70-97%
Advanced	90-100%	90-100%	92-100%	99-100%	98-100%

B. Comparison of the 2009-10 NDAA1 with Prior NDAA1s

- *The essential nature of the NDAA1 has been consistent over time. It is a selective response-type assessment with the same components. While refinements have been made in the various components, these have not invalidated any comparisons of results from one year to another.*
 - *The NDAA1 continues to be a test that collects data on student achievement on a teacher-selected activity that is tied to the grade-level content standards and benchmarks, and records information on the environment in which the student performs the selected activities.*

Each year since 2001, the NDAA1 has been refined based on an analysis of the submitted NDAA1 test booklets and input from the scoring team and teachers.

Minor revisions have been made to the test in recent years. For example, an analysis of the spring 2002 NDAA1 test booklets indicated that many teachers had problems in (a) designing activities that were aligned with the standard, and (b) writing performance objectives for their students. Thus, a section on how to write activities and performance objectives was added to the 2002-03 administration manual. After the spring 2003 NDAA1, an Excel template for creating line graphs was made available to teachers to reduce submission of unreadable graphs. Furthermore, the scoring rubric for measuring growth in performance was improved to increase scorer accuracy and inter-rater agreement. None of these modifications changed the essential nature of the NDAA1. The NDAA1 has always collected data on student achievement on a teacher-selected activity that was tied to the benchmarks. For 2002, 2003, 2004, and 2005, the test consisted of teachers writing and collecting data on one performance activity for each of reading/language arts and mathematics. Furthermore, data on the Situational Indicators were collected each of these years.

In summer 2005 the following refinements were made to the NDAA1:

- The number of activities a student was required to demonstrate on the assessment was increased from two to eight to more closely align the NDAA1 with the North Dakota State Assessment (NDSA).
- Instead of requiring teachers to write their own activities and set their own performance targets for the NDAA1 students, teachers select activities with established performance targets from an approved list of activities. The activities and corresponding performance objectives were written by a team of expert teachers using the process described in Section C. The activities were designed to be linked to the achievement standards for those students with significant cognitive disabilities.

Other changes have included:

- In 2007-08, the question that pertained to social interaction with peers was removed.
- In 2008-09, the question that pertained to the student monitoring his/her own performance was removed.
- In 2009-10, based on the May 2009 alignment study, those activities that were not aligned with the benchmark were deleted as well as items that were not considered to be academic in nature..

Despite these refinements, the essential nature of the NDAA1 has not changed. The NDAA1 collects data on student achievement on a teacher-selected activity that is tied to the grade-level benchmarks. Further, the Situational Indicators remain a main component of the NDAA1. Teachers continue to have flexibility in choosing the activities on which their students will perform. (Given the unique needs of the population, teacher flexibility is critical in order to have

the test accurately reflect a meaningful task for the student relating to the standards.) The main differences in the NDAA1 since 2005-2006 and that of prior years' is that (1) students now have to demonstrate their skill level on eight instead of two reading and math performance tasks and (2) the performance tasks are aligned to the grade-level content standards and benchmarks and are linked to the state's challenging achievement standards.

C. Alignment of the NDAA1 to the NDSA

The alignment of the NDAA1 to the NDSA constitutes content validity evidence.

- *The NDAA1 and North Dakota State Assessment (NDSA) cover the same content standards and cover varying degrees of complexity.*
 - *The test blueprints indicate the two tests cover the same content standards. Students who take either the NDAA1 or the NDSA receive scores on each of three reading/language arts standards, each of five mathematics standards, and each of the six science standards.*
 - *The complexity ratings indicate that the NDAA1 consists of tasks with a variety of cognitive complexity, just like the NDSA does.*

In July 2005, 20 teachers met in Bismarck, ND to generate a standard set of reading and mathematics activities for the NDAA1. These teachers represented districts and schools of various geography, size, grade level (grades 3-12), and for both content areas (English Language Arts and mathematics). These particular teachers were chosen based on their expertise in the classroom, their prior experience with students with disabilities, and their familiarity with the state standards.

The teachers were presented a test blueprint of the North Dakota State Assessment (NDSA) which indicates the reading/language arts and mathematics standards and benchmarks that are assessed by the NDSA. They were instructed to write linked performance activities at three different levels for each of the benchmarks covered by the NDSA. This provided a blueprint of the NDAA1 that corresponds to the item blueprint for the NDSA. For example, if three benchmarks under Reading Standard 2 are assessed on the NDSA, then linked activities were written for each of these three benchmarks for the NDAA1. Appendix B shows the benchmark maps.

The teachers' task was to create linked activities tied to the reading/language arts and mathematics grade-level standards and benchmarks. Teachers were given instruction on how to write tasks that match the content of grade-level standards and benchmarks. They were instructed on how to determine the "essence" of the benchmark and its level of cognitive complexity. The essence reflects the intent of the standard, i.e., it indicates the key knowledge and/or skills covered by the benchmark. For example, this benchmark "6.1 Use conventions of grammar related to sentence structure i.e., sentence reduction, parallel structure, elliptical clauses, conjunctions, clausal and phrasal patterns" has as its essence: "Apply rules of grammar to writing sentences."

After the essence was determined, the teachers determined the lowest level of complexity at which the benchmark and its corresponding essence could be assessed and wrote an activity at this lowest complexity level based on their best professional judgment of students with significant cognitive disabilities. (These are the "linked" activities as described on page 8.) Because these activities serve as "entry points" to the grade-level standards and benchmarks, these activities are written based on grade-level academic content.

After the teachers wrote the three levels of activities for each of the benchmarks on the test blueprint, they rated each activity on Bloom's Taxonomy. Bloom's Taxonomy consists of six levels of cognitive complexity and ranges from the lowest level named "Knowledge" (e.g., recall and recognition of facts) to the highest level named "Evaluation" (e.g., predict, support, judge). The mathematics group then evaluated the reading/language arts group's work and vice versa on (a) whether the activities were understandable, were linked, measured the targeted content

area (mathematics or reading/language arts), and represented meaningful activities and (b) whether the complexity ratings were appropriate. If any activities were found to have problems with understandability, level of linkage, what they were measuring, their meaningfulness, or their complexity level, they were flagged and returned to the original writers for editing. This continued until all members agreed on the outcomes.

Because this activity-writing process was approved in the peer review for the reading and math NDAA1, a similar process of activity writing took place in July 2006 for science. (See Appendix C for a list and description of the teachers who participated and Appendix D for a copy of the training materials.)

Displays C-1, C-2, and C-3 show the number and percentage of tasks that have a given cognitive complexity rating. The Bloom Taxonomy ratings indicate that the NDAA1 consists of tasks with a variety of cognitive complexity.

A comparison of the NDAA1 to the NDSA shows that both cover the same content standards and have varying degrees of complexity. Students who take either the NDAA1 or the NDSA receive scores on each of three reading/language arts standards, each of five mathematics standards, and each of six science standards. The complexity ratings indicate that the NDAA1 consists of tasks with a variety of cognitive complexity, just as the NDSA does. The NDAA1 provides coverage of depth of knowledge that is appropriate for an alternate assessment and that is comparable to that of NDSA. The documents *Alignment Analysis of Mathematics Standards and Assessments 2006* and *Alignment Analysis of Reading Standards and Assessments 2006* may be referenced for details on the cognitive complexity of the NDSA.

A notable difference between the NDAA1 and NDSA is the number of “items” a student answers per standard. On the NDAA1, the student performs one activity four times per standard; on the NDSA a student typically answers several multiple-choice items per standard. However, completing an activity four times on the NDAA1 represents four “items”. For example, if a teacher had the student perform the activity, “Given a magazine or newspaper article, the student will identify five persuasive statements from the article,” the student would be given a different magazine or article each of the four times. Furthermore, the student is expected to perform this item in multiple, integrated settings and to select, plan, and monitor his/her performance (as measured by the Situational Indicators described in section A). Thus, the student is not performing the exact same task under the exact same circumstances for each of the four activity administrations. The ability to generalize a skill to various settings under various circumstances is a very critical skill for students with significant cognitive disabilities because generalization is often very difficult for these students. Thus, the requirement for a student to demonstrate a selected activity four times is reasonable for this group of students and is one that allows for breadth and depth of complexity in the activities performed.

Display C-1: Cognitive Complexity of Benchmarks
Percentage of activities at a given cognitive complexity level on the NDAA1
Reading

Bloom's Taxonomy Rating	Grade tested:						
	3	4	5	6	7	8	11
1 Knowledge	61%	23%	14%	44%	58%	31%	20%
2 Understanding	26%	27%	41%	26%	26%	50%	30%
3 Application	5%	27%	26%	26%	11%	13%	5%
4 Analysis	5%	19%	19%	42%	5%	6%	45%
5 Synthesis	3%	40%	0%	0%	0	0%	0%
6 Evaluation	0%	0%	0%	0%	0%	0%	0%

Display C-2: Cognitive Complexity of Benchmarks
Percentage of activities at a given cognitive complexity level on the NDAA1
Mathematics

Bloom's Taxonomy Rating	Grade tested:						
	3	4	5	6	7	8	11
1 Knowledge	91%	68%	41%	58%	50%	29%	48%
2 Understanding	9%	22%	24%	30%	39%	45%	41%
3 Application	0%	10%	21%	9%	0%	26%	10%
4 Analysis	0%	0%	15%	2%	7%	0%	2%
5 Synthesis	0%	0%	0%	0%	4%	0%	0%
6 Evaluation	0%	0%	0%	0%	0%	0%	0%

Display C-3: Cognitive Complexity of Benchmarks
Percentage of activities at a given cognitive complexity level on the NDAA1
Science

Bloom's Taxonomy Rating	Grade tested:		
	4	8	11
1 Knowledge	71%	68%	47%
2 Understanding	28%	32%	51%
3 Application	1%	0%	2%
4 Analysis	0%	0%	0%
5 Synthesis	0%	0%	0%
6 Evaluation	0%	0%	0%

D. Standard Setting

- *A technically-sound process was used in establishing the cut-scores for the NDAA1.*

In February 2006, a group of 21 teachers and administrators met to determine the cut scores for each of the four achievement levels of the NDAA1 for reading and mathematics. These teachers and administrators were from a representative sample of districts and schools based on geography, size, and organizational type (e.g., elementary, middle, high, combination). These individuals were chosen based on their expertise in the classroom, their prior experience with students with disabilities, and their familiarity with the state content and achievement standards (see Appendix D for a list of the participants). The standard-setting process was facilitated by Dr. Susan Wagner of Data Driven Enterprises with assistance from Wayne Ball of Mountain Plains Regional Resource Center (MPRRC).

The purpose of this standard-setting group was to provide recommendations on how the NDAA1 should be scored and how achievement levels should be assigned. They were given information on the NDAA1 – its content and structure (e.g., the aligned and linked activities, the 4 Performance Points, the 10 Situational Indicators) and its preliminary results. The standard-setting process involved intense review of the NDAA1 and data, small group deliberations, multiple votes, and consensus building.

The group reviewed prior years' results, fall 2005 preliminary results, content of the activities, and the Situational Indicators. The large group then divided into three smaller groups to do the various standard-setting tasks that included determining:

- Which of the 10 Situational Indicators should be counted?
- How much weight should the 4 Performance Points get?
- How much weight should the 10 Situational Indicators get?
- Once the 10 Situational Indicators and 4 Performance Points are integrated, what cut-scores would determine the four achievement levels (i.e., novice, partially proficient, proficient, advanced) for reading/language arts and for mathematics?

These issues were discussed first by the three small groups; then as a large group. Differences were resolved and consensus was reached on the issues.

A similar process was followed for science in February 2007, in February 2009, and again in February 2010. Appendix E lists the participants in the standards setting process, and Appendix F includes all the materials used in this process. In February 2010, about 30 participants were involved in the standards-setting process. Participants broke into grade level groups: an elementary group that reviewed the grade 4 science test; a middle school group that reviewed the grade 8 science test; and a high school group that reviewed the grade 11 science test.

Display D-1 indicates the percentage of total points required for placement in each assessment level. (Students receive a final score that is based on total percent of points earned.) Display D-2 indicates the percentage of students at each achievement level based on these cut scores. These were the data that were available at the time of the cut-score setting and that were used by the group.

**Display D-1: Proficiency Levels for the NDAA1
Percentage of Total Points a Student Must Earn
for a Given Proficiency Level**

	Reading	Mathematics	Science Grade 4	Science Grade 8	Science Grade 11
Novice	0-49%	0-49%	0-48%	0-49%	0-44%
Partially Proficient	50-65%	50-65%	49-67%	50-74%	45-69%
Proficient	66-89%	66-89%	68-91%	75-98%	70-97%
Advanced	90-100%	90-100%	92-100%	99-100%	98-100%

**Display D-2: Preliminary Results of the NDAA1
At the Time of the Science Standard-Setting
Percent Who Scored at A Given Proficiency Level in Reading**

	Number	Novice	Partially Proficient	Proficient	Advanced	Proficient/ Advanced
Overall	677	3%	6%	53%	38%	91%
Grade 3	101	1%	7%	57%	35%	92%
Grade 4	84	0%	4%	57%	39%	96%
Grade 5	98	4%	12%	58%	26%	84%
Grade 6	94	1%	7%	59%	33%	91%
Grade 7	105	4%	4%	50%	43%	92%
Grade 8	99	7%	5%	48%	39%	88%
Grade11	96	4%	2%	42%	52%	94%

**Display D-3: Preliminary Results of the NDAA1
At the Time of the Science Standard-Setting
Percent Who Scored at A Given Proficiency Level in Mathematics**

	Number	Novice	Partially Proficient	Proficient	Advanced	Proficient/ Advanced
Overall	688	2%	6%	51%	41%	92%
Grade 3	100	2%	10%	46%	42%	88%
Grade 4	79	3%	3%	57%	38%	95%
Grade 5	96	3%	2%	53%	42%	95%
Grade 6	92	1%	9%	58%	33%	90%
Grade 7	116	2%	5%	50%	43%	93%
Grade 8	102	2%	11%	56%	31%	87%
Grade11	103	2%	5%	39%	54%	93%

**Display D-4: Preliminary Results of the NDAA1
At the Time of the Science Standard-Setting
Percent Who Scored at A Given Proficiency Level in Science**

	Number	Novice	Partially Proficient	Proficient	Advanced	Proficient/ Advanced
Overall	268	2%	16%	66%	15%	82%
Grade 4	77	1%	10%	62%	26%	88%
Grade 8	94	3%	22%	67%	7%	74%
Grade 11	97	1%	15%	69%	14%	84%

E. Reliability of NDAA1: Improvements

Several steps were taken to increase the reliability of the North Dakota Alternate Assessment.

- *The overall reliability of the NDAA1 has increased.*
 - *Compared with prior years, since 2005-06, the NDAA1 consists of more performance tasks, measures more content standards, has greater standardization across tasks and students, and eliminates teacher error to a greater degree.*

Increase in Number of Assessed Tasks

Prior to the 2005-06 school year, a student's reading/language arts and mathematics score on the NDAA1 was based on one performance task each. Since 2005-06, a student's reading/language arts score is based on three performance tasks; a student's mathematics score is based on five performance tasks. A student's science score is based on six performance tasks. The increase in the number of tasks increases the reliability of the test.

Increase in Number of Content Standards Assessed

Related to an increase in the number of tasks, the test measures more content standards since 2005-06. Each of the three reading/language arts tasks measures a different reading/language arts content standard. Each of the five mathematics tasks measures a different mathematics content standard. Each of the six science tasks measures a different science content standard. The NDAA1 covers more content standards, and thus has more breadth than assessments from prior years. As such, the NDAA1 is more closely aligned to the North Dakota State Assessment which also measures three reading/language arts content standards, five mathematics content standards, and six science content standards.

Greater Standardization

Starting in 2005-06, teachers were given a list of approved tasks for each of the Reading and mathematics standards. Starting in 2006-07, teachers were given a list of approved tasks for each science standards. The reasons for providing teachers with a list of tasks from which to choose were to eliminate the:

1. Great variability in the quality of tasks teachers created.
2. Possibility of teachers designing inappropriate (e.g., not age-appropriate, not aligned with the content standard) tasks for the student to demonstrate on the NDAA1.
3. Possibility of teachers setting an inappropriate target of performance (i.e., either too easy or too difficult) for a student to demonstrate on the chosen task. In years past, some teachers would select a level of performance that was too difficult for the student to demonstrate; as such, the student was penalized on the NDAA1 for not meeting the target performance. Or, some teachers would select a level of performance that was too easy for the student to demonstrate; as such, the student was penalized on the NDAA1 for not showing growth in performance on the chosen task.

Because teachers must select from an approved list of tasks and because for each task students are expected to demonstrate performance at an 80% correct level, students are not penalized for teachers selecting inappropriate tasks or tasks with too easy or too difficult performance targets. This has resulted in all students needing to demonstrate the same level of performance on the same task; performance expectations are no longer teacher-determined.

Elimination of Teacher Error

In prior years, several types of teacher error were present. Two types have already been mentioned: teachers designing inappropriate tasks for the student to demonstrate and teachers

setting an inappropriate target of performance. Other types of teacher error included creating uninterpretable graphs and not including enough data points on student performance. Teachers were required to plot student performance on the selected task in a line graph or bar chart. Some teachers created graphs that were uninterpretable; as such, the student's NDAA1 could not be scored. In addition, some teachers did not provide enough data points to show a student's performance on the task over time. Once again, the student's NDAA1 could not be scored.

Starting in 2005-06, the test is designed to clearly show that four data points are needed. The teacher enters these four data points in a spreadsheet, and the teacher does not have to create a graph of performance. Further, as mentioned above, the greater standardization of the tasks and performance targets has eliminated this type of teacher error.

In 2006-07, the ND Department of Public Instruction (DPI) established "anchor items" on the NDAA1. At each grade level, a reading anchor activity and a mathematics anchor activity was determined. All students needed to complete these anchor activities. The purpose of these anchor activities was to allow for better equating of scores over time, greater reliability, and the opportunity to collect additional validity evidence. These anchor activities served as a way to increase the standardization of the tasks across students.

In 2008-09, these anchor activities were standardized. For example, in the past, the anchor activity might be: the student will identify five community signs. It was up to the teachers to determine what five signs the student would identify and to get pictures of the five signs. In 2008-09, the specific signs the student should identify were determined, and the teacher was provided with the necessary stimulus materials to administer the task. This will allow for greater standardization across students and greater equating across test administrations than before. In 2009-10, anchor activities for science were developed and implemented.

In summary, all of the above factors have increased the overall reliability of the NDAA1. Compared to prior years, starting in 2005-06, NDAA1:

- has more performance tasks,
- measures more standards,
- has greater standardization across students, and
- eliminates teacher error to a greater degree, resulting in selection of appropriate and aligned performance tasks and appropriate targets of student performance.

F. Reliability of NDAA1: Standard Error of Measurement

Reliability of the NDAA1 is evidenced in its standard error of measurement (SEM).

- *The SEM of the NDAA1 is reasonably small.*
 - *The SEM ranges from 3.0 percentage points to 4.1 percentage points.*
- *The SEM in 2009-10 is similar to those of previous years.*

The SEM is an estimate of the amount of error present in an individual student's test score. The SEM represents the range in which a student's true test score most likely falls. The SEM was calculated for each of the three NDAA1 tests: reading/language arts, mathematics, and science (see Display F-1). As indicated in the display, there is a 95% chance that any given student's true test score is within 6-8 percentage points of the student's calculated test score. (Test scores represented the percentage of points a student earned and thus could range from 0-100%.) For example, the average student on the mathematics NDAA1 received a score of 83.9%. There is a 95% chance that this average student's true score lies between 78.1% - 90.3%.

Display F-1: Standard Error of Measurement on the NDAA1

	Standard Deviation	Alpha Coefficient	SEM	SEM 2	Mean Score	95% Confidence Interval	
						Low	High
Reading	.133	.906	.041	.082	83.2%	75.0%	91.4%
Mathematics	.126	.942	.030	.061	84.2%	78.1%	90.3%
Science	.139	.949	.031	.063	82.9%	76.6%	89.2%

Interpretation Example: The average score on the Reading test was 83.2%. The SEM (calculated from the standard deviation and alpha coefficient) on the Reading is 0.043. Thus, there's a 95% chance that the true score of a student with a calculated score of 83.2% is between 75.0% and 91.4%.

The SEMs were highest in 2005-06. They have been fairly consistent since 2006-07.

Display F-2: Standard Error of Measurement and Mean Scores on the NDAA1 Over Time

	2005-06		2006-07		2007-08		2008-09		2009-10	
	SEM	Mean Score	SEM	Mean Score	SEM	Mean Score	SEM	Mean Score	SEM	Mean Score
Reading	0.045	74.90%	0.039	85.60%	0.042	82.70%	0.043	82.8%	.041	83.2%
Mathematics	0.035	76.20%	0.027	86.70%	0.027	84.10%	0.030	83.9%	.030	84.2%
Science			0.024	86.00%	0.026	84.20%	0.021	85.7%	.031	82.9%

G. Quality of Submitted Alternate Assessments

A total of 719 NDAA1s from grades 3-8 and 11 were submitted for scoring in February 2010. Of these, less than one percent received an invalid score due to missing information.

- *The quality of the submitted NDAA1s has dramatically increased from 2002 to 2010.*
 - *Since 2008, less than 1% of tests have had a quality-related problem in terms of teachers not completing information on each of the eight - fourteen performance tasks (see Display G-1).*

Number of Tests Submitted for Scoring:

2002=153 2003=133 2004=241 2005=870 2006=1,183 2007=1,471 2008 = 674 2009=673 2010=719

Note: In 2005 students in grades 3-8 and 11 were assessed; prior to 2005 students in only grades 4, 8, and 11 were assessed. In 2005-06 and 2006-07, the “2%” students were included in these figures. In 2008, only the “1%” students in grades 3-8 and 11 were included.

Display G-1: Assessment Quality Indicators Over Time

Percent of NDAA1s that:	2002	2003	2004	2005	2006	2007	2008	2009	2010
Had NO problems	62%	86%	90%	98%	99%	99%	99%	99%	99%
Received invalid scores on all content areas	22%	3%	2%	<1%	<1%	<1%	<1%	<1%	<1%
Had at least one of the following problems:	38%	14%	10%	2%	1%	1%	<.1%	<.1%	<.1%
Had at least one inadequate graph (uninterpretable, not enough data points)	25%	11%	5%	2%	N/A	N/A	N/A	N/A	N/A
Had at least one unaligned goal	14%	2%	4%	<1%	N/A	N/A	N/A	N/A	N/A
Had at least one goal with unclear performance criteria	14%	1%	1%	<1%	N/A	N/A	N/A	N/A	N/A
Had missing data	N/A	N/A	N/A	N/A	1%	1%	<.1%	<.1%	<.1%

H. Content Validity: Teacher and Parent Judgments

Validity of the NDAA1 is evidenced by perception of teachers and parents regarding the extent to which the NDAA1 accurately reflects the student’s level of achievement.

The following tables indicate what percent of tests the teachers and parents perceived as accurate.

- *Very high percentages of teachers (95-98%) and parents (90-93%) perceive the NDAA1 to be an accurate reflection of student skill level (see Displays H-1 and H-5).*
- *The percentage of teachers who perceive the NDAA1 to be an accurate reflection of student skill level steadily increased from 2002 to 2005 and has remained fairly steady since then (see Display H-2).*
- *The percentage of parents who perceive the NDAA1 to be an accurate reflection of student skill level steadily increased from 2002 to 2006 and has remained fairly steady since then (see Display H-4).*

Display H-1: Content Validity – Teacher Attitudes

Teacher Views on Accuracy of NDAA1

Percent of Teachers Who Selected a Given Response

Across Grade Levels, Number of Students: Reading = 681; Math = 691; Science = 268

	NDAA1 accurately reflects the student’s skill level.	Student performs better than NDAA1 indicates.	Student performs worse than NDAA1 indicates.
Reading	97%	2%	1%
Mathematics	96%	3%	1%
Science	97%	2%	1%

Students on Which Teachers Made as Accuracy Judgments:

2002=115
 2003=133
 2004=241
 2005=813
 2006=1,107
 2007=1,471
 2008=655
 2009=672
 2010=719

Display H-2: Content Validity – Teacher Attitudes Over Time

Teacher Views on Accuracy of NDAA1

Percent of Teachers who Said, “NDAA1 accurately reflects the student’s skill level”

Across Grade Levels

	2002	2003	2004	2005	2006	2007	2008	2009	2010
Reading	87%	82%	92%	98%	97%	97%	97%	95%	97%
Mathematics	90%	84%	90%	97%	97%	97%	97%	97%	96%
Science						98%	99%	98%	97%

Note: The reason for the increase in the number of judgments over time is due to more grade levels being assessed. In 2005 and 2006 students in grades 3-8 and 11 were assessed; prior to 2005 students in only grades 4, 8, and high school were assessed. In 2006 and 2007, the NDAA2 data are included. In 2008, 2009, and 2010, and only NDAA1 data are included.

Display H-3: Content Validity – Parent Attitudes

Parent Views on Accuracy of NDAA1
 Percent of Parents Who Selected a Given Response
 Across Grade Levels, Number of Students =.555

	NDAA1 accurately reflects the student’s skill level.	Student performs better than NDAA1 indicates.	Student performs worse than NDAA1 indicates.	Not enough info to judge
Reading	93%	5%	1%	2%
Mathematics	92%	4%	1%	3%
Science	90%	5%	1%	4%

Display H-4: Content Validity – Parent Attitudes Over Time

Parent Views on Accuracy of NDAA1s
 Percent of Parents Who Said, “NDAA1 accurately reflects the student’s skill level”
 Across Grade Levels

	2002	2003	2004	2005	2006	2007	2008	2009	2010
Reading	84%	83%	90%	91%	93%	93%	92%	93%	93%
Mathematics	83%	86%	90%	90%	92%	94%	91%	94%	92%
Science						94%	90%	94%	90%

Display H-5: Content Validity – Parent Confirmation of Skill

Percent of parents who observed their child demonstrate the targeted skill in 2010

	At Home	In Community	At School	Other
Reading	84%	55%	71%	11%
Mathematics	87%	58%	70%	12%
Science	76%	54%	64%	11%

Display H-6: Content Validity – Parent Confirmation Over Time

Percent of Parents Who Said That They Observed Their Child Demonstrate the Targeted Skill
 Across Grade Levels

	2002	2003	2004	2005	2006	2007	2008	2009	2010
Reading	91%	94%	85%	89%	94%	97%	85%	94%	90%
Mathematics	87%	91%	86%	88%	94%	98%	88%	94%	91%
Science						97%	98%	90%	83%

Number of Parents Making Accuracy Judgments:

2002=109 2006=1,048
 2003=125 2007=1,376
 2004=230 2008= 602
 2005=776 2009=603
 2010= 555

Note: The reason for the increase in the number of parents is due to more grade levels being assessed. In 2005 students in grades 3-8 and 11 were assessed; prior to 2005 students in only grades 4, 8, and 11 were assessed. In 2006 and 2007, the NDAA2 data are included. In 2008 2009, and 2010, only NDAA1 data are included.

I. Construct Validity

Validity is evidenced in the correlation between two separate judgments of performance. If the score the student receives on the NDAA1 correlates with a separate **and valid** score from another source, this implies that the NDAA1 is valid. Teachers were asked to provide a judgment of the student's performance level on the benchmark. The NDAA1 rating was then correlated with the teacher rating.

- *The relationship between the overall NDAA1 rating and teacher judgment of student skill is moderate and significant. These correlations suggest that the assessment is a valid indicator of a student's performance on the assessed benchmarks.*
 - *The correlations in 2010 are lower than that obtained in 2006. This downward trend began in 2007 and continued in 2009 but have stabilized or increased in 2010.*

Display I-1: Construct Validity - Correlations

Correlation of Teacher-Assigned Performance Rating with NDAA1 Overall Rating

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Reading	.11	.27	.37	.51	.40	.66	.59	.51	.42	.44
Mathematics	.11	.20	.28	.37	.45	.64	.61	.54	.44	.44
Science							.47	.51	.38	.56

Interpretation Example: Correlations below .20 = marginal to nonexistent; .20 to .50 = small to moderate; .50 to .70 = moderate to high; above .70 = very high.

To better illustrate the magnitude of the correlation between the teacher rating and the overall score, the following table shows the average total assessment score at each level of the teacher rating. If the total test score increases with increases in teacher rating, then the two scores are correlated.

Display I-2: Construct Validity – Teacher Judgment

Average Total Test Score (Percent of Total Points Earned) by Teacher Overall Performance Rating

	Teacher Rating of Student Skill Level:			
	1 Novice	2 Partially Proficient	3 Proficient	4 Advanced
Reading	67%	78%	85%	90%
Mathematics	69%	80%	85%	91%
Science	64%	75%	85%	92%

Interpretation Example: Of all the students who received a teacher-assigned performance rating of 1 for reading/language arts, on average, they earned 67% of the total points on the reading/language arts test; of the students who received a teacher-assigned performance rating of 2 for Reading, they earned an average of 78% of total points.

The above evidence of construct validity is called convergent validity which occurs when two measures of similar constructs are related. Divergent validity is another form of construct validity and is indicated by a **weak** relationship between two **different** constructs. The overall assessment rating for reading/language arts was correlated with teacher judgment for mathematics. The table below indicates that the convergent validity correlations are generally higher than the divergent correlations. These correlations provide further evidence of the construct validity of the assessment.

Display I-3: Convergent and Divergent Validity - Correlations

Correlation of Teacher-Assigned Performance Rating with NDAA1 Overall Rating

Student NDAA1 Rating on:	Teacher Judgment of:		
	Reading	Mathematics	Science
Reading	0.42	0.39	0.46
Mathematics	0.39	0.44	0.46
Science	0.48	0.52	0.56

Internal consistency is a measure of the extent to which items on a test measure a single scale. If the items that make up a given construct on a test are highly correlated with one another, then this is evidence of the validity of the test. Coefficient Alpha, a typical internal consistency statistic, was calculated for the three reading/language arts content standards, the five mathematics content standards, and the six science content standards. Results indicate that the three activities that make up the reading/language arts test, the five activities that make up the mathematics test, and the six activities that make up the science test represent a single trait.

Display I-4: Construct Validity – Internal Consistency

	Coefficient Alpha
Reading	0.91
Mathematics	0.94
Science	0.95

Collecting Additional Construct Validity Evidence

In 2010-11, the ND Department of Public Instruction (DPI) will gather for each NDAA1 student, additional data from teachers who administer the NDAA1 on data (e.g., report card grades, classroom assignments, tests) they already have on student achievement on the ND academic content standards. The quality of these data sources will then be examined to determine if these data could be correlated with scores on the NDAA1 as further evidence of the construct validity of the NDAA1.

J. Fidelity of Administration: Use of Activities and Accuracy of Teacher-Provided Ratings

“Fidelity of administration” refers to whether the teachers are administering the NDAA1 in the way that it was designed to be administered. One way of examining the fidelity of administration is to examine if teachers used a variety of activities from the list of provided activities. The purpose of the large number of activities is so that teachers can individualize the NDAA1 for each of their students, considering needs and educational goals. Thus, a variety of activities should be selected.

Because the NDAA1 relies on teachers assigning ratings to certain student behaviors, verifying that the teacher-provided ratings are accurate is a second method for determining the fidelity of administration and, in general, is an important step in establishing the quality of the NDAA1. In addition to assigning ratings to student behaviors on the Situational Indicators, teachers are also asked to provide specific details of student performance. To determine if the detailed information provided by teachers supports the numerical ratings, one of the detailed questions was compared with three one of the Situational Indicator numerical ratings.

Ensuring Fidelity of Administration

In 2009-10, the ND Department of Public Instruction (DPI) will design and implement a monitoring program to ensure the fidelity of administration using a survey/training model. DPI will send trained staff members to observe teachers administering the NDAA1. A representative sample of teachers test administration performance will be reviewed

- *Teachers chose a variety of activities for their students to demonstrate on the NDAA1 which suggests that they are individualizing the NDAA1 for each student (see Display J-1).*
- *In general, the detailed information provided by teachers support and confirm the numerical ratings that teachers assigned to the Situation Indicators.*
 - *Compared with teachers who listed fewer settings, teachers who listed more settings on the detailed question tended to assign higher numerical ratings (as should be expected) to the “Number of Setting” Situational Indicator (see Display J-2).*
- *For Reading Standard 2, Mathematics Standard 1, Science Standard 2 (grades 4 and 8), and Science Standard 4 (grade 11), anchor items were used. Thus, teachers at each grade level were required to administer the same activity.*

Display J-1: Number of Unique Activities Chosen to be Assessed on the NDAA1
 For Each Standard
 By Grade Level

	Grade						
	3	4	5	6	7	8	11
Reading							
Standard 2	1	1	1	1	1	1	1
Standard 3	18	19	16	11	7	2	14
Standard 6	20	24	19	2	3	3	2
Mathematics							
Standard 1	1	1	1	1	1	1	1
Standard 2	20	12	12	16	13	12	12
Standard 3	13	4	15	18	9	15	14
Standard 4	24	19	16	15	10	13	11
Standard 5	14	10	6	14	17	11	12
Science							
Standard 1	NA	3	NA	NA	NA	9	5
Standard 2	NA	1	NA	NA	NA	1	8
Standard 3	NA	13	NA	NA	NA	3	17
Standard 4	NA	12	NA	NA	NA	9	1
Standard 5	NA	8	NA	NA	NA	6	9
Standard 6	NA	2	NA	NA	NA	3	8

Display J-2: Accuracy Check – Number of Settings Dimension

Average Number of Settings Listed in Detail Question

By Number of Settings on Situational Indicator

Detail Question: For each of school settings, community settings, home settings, and other settings, indicate the settings in which the activity was performed:

	Q1: In how many settings did student perform skill		
	1	2	3+
Reading	1.1	2.1	4.2
Mathematics	1.1	2.2	4.3
Science	1.0	2.2	4.1

Note: Number of Settings dimension and the number of settings listed on the detail question correlated 0.75 for reading/language arts, 0.75 for mathematics, and 0.80 for science.

Interpretation Example: For those students who received a teacher-assigned rating of 1 on the Number of Settings Situational Indicator for Reading, teachers on average listed 1.1 settings on the detailed question; for those students who received a teacher-assigned rating of 2 on the Number of Settings Situational Indicator, teachers on average listed 2.1 settings on the detailed question. Notice the step pattern of 1.1 → 2.1 → 4.2.

See Appendix G for item statistics on all the individual activities. See Appendix H for a report on the results of the fall 2009 NDAA1.

K. Consequential Validity

Consequential validity refers to the consequences that result from using a particular test. The consequences of the NDAA1 can be intended or unintended and positive or negative. For example, an intended positive consequence may be that students with disabilities have increased access to the general curriculum. An unintended positive consequence may be increased collaboration between special education and general education teachers. An unintended negative consequence may be decreased instruction on functional skills.

The consequences of a given test are determined over a long period of time. Thus, data on consequential validity will be conducted over a multi-year period. However, any preliminary data will be reported in each year's Technical Manual.

The DPI will integrate any consequential validity regarding the NDAA1 into the state's comprehensive consequential validity study of the statewide assessment system.

L. Teacher Feedback on the NDAA1

All teachers who completed one or more NDAA1s were asked to complete a survey on the NDAA1. Completed surveys were received from 157 teachers. 293 teachers submitted an alternate assessment. Thus, 57% of teachers completed a survey.

- *When asked if the NDAA1 was easy to administer, 94% said yes.*
- *When asked if the NDAA1 activities were appropriately linked with the standards, over 98% said yes.*
- *When asked how the NDAA1 could be improved suggestions included providing more activities and receiving the materials earlier in the fall.*
- *When asked the best things about the NDAA1 comments included it being easy to administer, the variety of activities, its appropriateness for the students, and that it was aligned to state standards.*

Teachers rated the NDAA1 very positively. Over 70% agreed that:

- The Math Activities were appropriately linked with the State Math Standards (98% agreed).
- The Reading Activities were appropriately linked with the State Reading Standards (98%).
- The Science Activities were appropriately linked with the State Science Standards (98%).
- The process for completing the NDAA1 online was easy (96% agreed).
- The directions on how to administer the NDAA1 were clear (94% agreed).
- The NDAA1 has positively impacted my use of the state standards (82% agreed).
- The method for scoring the NDAA 1 represents an appropriate way to evaluate student performance on the State Standards (75% agreed).

Over 70% of teachers agreed that the NDAA1 has positively impacted student learning (72% agreed).

When asked to describe (in an open-ended item) how the assessment could improve, 135 respondents (71%) answered this question.

- 30% mentioned more activity variety
- 19% mentioned more activities for low functioning students
- 15% mentioned removing situational indicators
- 15% mentioned receiving materials to prepare earlier.

When asked to describe (in an open-ended item) the positive features about the assessment, 144 respondents (75%) answered this question:

- 44% mentioned the variety of activities
- 38% mentioned that it was easy to use
- 21% mentioned that the measures were appropriate for special education students
- 21% mentioned that the assessment aligned to the state standards