Diagnostic Online Math Assessment

&

Adaptive Diagnostic Assessment of Mathematics

Technical Document

Published by Let’s Go Learn, Inc.
# Table of Contents

Adaptive Diagnostic Assessment of Mathematics & Diagnostic Online Math .............................................. 3  
Assessment Specifications ......................................................................................................................... 3  
  ADAM: K-7 ................................................................................................................................................... 3  
  DOMA: Pre-Algebra .................................................................................................................................. 3  
  DOMA: Algebra ......................................................................................................................................... 4  
ADAM & DOMA Content Validity ............................................................................................................. 6  
DOMA Pre-Algebra Construct Validity: Correlation Study (2007) .............................................................. 8  
DOMA: BMS Internal-Consistency Reliability Study ................................................................................ 9  
  Introduction .............................................................................................................................................. 9  
  Methods .................................................................................................................................................. 9  
  Analysis ............................................................................................................................................... 10  
  Conclusions ........................................................................................................................................ 10  
ADAM: K-7 Internal-Consistency Reliability Study ................................................................................ 11  
  Introduction .............................................................................................................................................. 11  
  Methods .................................................................................................................................................. 11  
  Analysis ............................................................................................................................................... 12  
  Conclusions ........................................................................................................................................ 12  
ADAM K-7 Test-Retest Reliability Study (2012) ..................................................................................... 16  
DOMA Pre-Algebra Test-Retest Reliability Study (2007) .................................................................... 17  
DOMA: Algebra Internal-Consistency Reliability Study ........................................................................ 18  
  Introduction .............................................................................................................................................. 18  
  Methods .................................................................................................................................................. 18  
  Analysis ............................................................................................................................................... 19  
  Conclusions ........................................................................................................................................ 20  
References ................................................................................................................................................... 22
Adaptive Diagnostic Assessment of Mathematics & Diagnostic Online Math Assessment Specifications

**ADAM: K-7**

*ADAM: K-7* assesses across 5 major math strands which span 44 sub-tests of K-7 mathematics. Grade range for all strands is K to 7

- Numbers and Operations: 14 sub-test; 661 criterion-referenced test items, in 105 constructs
- Measurement: 7 sub-tests; 133 criterion-referenced test items, in 34 constructs
- Geometry: 11 sub-test; 203 criterion-referenced test items, in 53 constructs
- Data Analysis: 8 sub-tests; 106 criterion-referenced test items, in 36 constructs
- Algebraic Thinking: 4 sub-test; 305 criterion-referenced test items, in 43 constructs

**DOMA: Pre-Algebra**

*DOMA: Pre-Algebra* consists of 14 sub-tests, as well as a pre-screening, math facts, and reading comprehension sections. The pre-screening is administered first as part of the OAASIS platform of adaptive test-taking, while the math facts and reading comprehension sections may be triggered by performance on the assessment. While other LGL assessments are criterion-referenced to specific grade-level expectations, *DOMA: Pre-Algebra* uses test items criterion-referenced to pre-requisite knowledge expectations.

- Pre-Screening: 14 criterion-referenced test items, one from each sub-test of the full assessment.
• Integer Operations: 11 criterion-referenced test items

• Fraction Operations: 12 criterion-referenced test items

• Decimal Operations: 9 criterion-referenced test items

• Comparing and Converting: 10 criterion-referenced test items

• Estimating and Rounding: 6 criterion-referenced test items

• Evaluating Exponents: 6 criterion-referenced test items

• Ratios and Proportions: 5 criterion-referenced test items

• Simplifying Expressions: 6 criterion-referenced test items

• Coordinate Graphing 8 criterion-referenced test items

• Linear Functions and Extending Patterns: 8 criterion-referenced test items

• Simple Equations: 6 criterion-referenced test items

• Geometry: 11 criterion-referenced test items

• Interpreting Data: 10 criterion-referenced test items

• Simple Probability: 7 criterion-referenced test items

DOMA: Algebra

DOMA: Algebra, a course-specific diagnostic assessment, consists of 11 Algebra I-specific constructs, as well as a pre-screening section, much like the DOMA: Pre-Algebra assessment.
Let's Go Learn

- Pre-Screening: 22 criterion-referenced test items, representing two questions from each sub-test

- Evaluating Advanced Exponents: 7 criterion-referenced test items

- Solving Linear Equations: 6 criterion-referenced test items

- Graphing and Analyzing Linear Equations: 9 criterion-referenced test items

- Relations and Functions: 7 criterion-referenced test items

- Solving and Graphing Inequalities: 5 criterion-referenced test items

- Solving and Graphing Systems: 8 criterion-referenced test items

- Polynomial Operations: 8 criterion-referenced test items

- Factoring Polynomials: 7 criterion-referenced test items

- Radical Expressions and Equations: 7 criterion-referenced test items

- Quadratic Equations: 7 criterion-referenced test items

- Rational Expressions and Equations: 8 criterion-referenced test items
ADAM & DOMA Content Validity

The validity of an assessment instrument refers to its ability to support valid assessment inferences. That is, do test results support a valid conclusion about a student's level of knowledge or skill? Building a valid test begins with accurate definitions of the constructs (i.e., the knowledge domains and skills) to be assessed. If the assessment activities in a test (i.e., the test items) tap into the constructs that the test is designed to assess, then the test has content validity. Although additional factors affect overall test validity, content validity is the basic logical bedrock of any test.

Let's Go Learn's math assessments' content validity comes from best-practices in math curriculum. The development of these cutting-edge math products has been spearheaded by math specialist and teacher-trainer, Paul Giganti of UC Berkeley and Cal State Hayward. Prior to his work in professional development, Paul Giganti taught math in public schools for over 15 years. He has directed federally-funded professional development programs in California under the auspices of the California Post-Secondary Educational Commission. Currently he is the coordinator of the California Mathematics Council Festivals Programs and Professional Development. In addition to his classroom teaching and professional development career, Paul Giganti has published several children's picture books about mathematics.

In addition to the expertise of Paul Giganti, LGL derives construct validity for the ADAM & DOMA Series of tests by its alignment to both Common Core State Standards (CCSS) and state standards. DOMA: Basic Math Skills test was originally aligned to the California state mathematics standards in the Number and Measurement strands, as well as NCTM National Standards for Mathematics. ADAM K-7, the sequel to the DOMA
Basic Math Skills assessment, was redesigned fundamentally and expanded to cover all 5 NCTM major math strands and nearly all of the CCSS. ADAM is aligned to CCSS and state standards in all 50 states. Further, DOMA: Pre-Algebra and DOMA: Algebra are aligned to NCTM standards, CCSS, and all 50 state standards.
DOMA Pre-Algebra Construct Validity: Correlation Study (2007)

The DOMA: Pre-Algebra assessment was administered to 233 middle school students. Students performed the DOMA: Pre-Algebra assessment online. During the same week students were also give the Orleans-Hannah Algebra Prognosis Test published by Harcourt. This assessment is administered in a group setting. It is a paper and pencil assessment.

The results were as follows when comparing the “raw score” on the DOMA assessment and the total number of questions correct on the Orleans-Hannah Algebra Prognosis Test:

Correlation: 0.754

Delta: 0.066

This represents a high correlation with low variance. This means that the DOMA: Pre-Algebra assessment and the Orleans-Hannah Algebra Prognosis Test are highly correlated, and demonstrates further construct validity for the DOMA: Pre-Algebra assessment.
DOMA: BMS Internal-Consistency Reliability Study

Introduction

According to Allen & Yen (1979), reliability can be defined as the consistency between the observed scores on an assessment and the true scores. There are multiple methods of assessing the reliability of an assessment. One way is the test-retest method, comparing the consistency between one administration and a subsequent second administration of the same assessment. Another method of determining reliability is by using internal consistency, a measure of the consistency of results of items within one test. This report will present the internal consistency results for Diagnostic Online Math Assessment: Basic Math Skills (DOMA BMS) (Let’s Go Learn, Inc.).

Methods

Let’s Go Learn, Inc., a developer of online diagnostic reading and math assessments, has created an online diagnostic assessment to evaluate students’ math abilities, Diagnostic Online Math Assessment: Basic Math Skills (DOMA BMS). DOMA BMS assesses students across three different subtests of math: Number Skills, Fractions, and Measurement. Further, the assessment is built on an adaptive-logic platform, limiting the number of items seen by each student by both determined start point and by early termination. For each sub-test, there is one version of the assessment.

In order to evaluate the reliability of DOMA BMS, the Rasch model was used to evaluate overall model fit for each sub-test. A sample of 4,675 students in grades K-12 taken from 9 school districts nationwide was used.
Analysis

Analysis showed overall unidimensionality of each sub-test as well as good overall model fit. Reliability levels were strong (see Table 1).

Table 1

*DOMA BMS Sub-test Reliability (α)*

<table>
<thead>
<tr>
<th>Sub-test</th>
<th>Reliability (α)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>.88</td>
</tr>
<tr>
<td>Fractions</td>
<td>.88</td>
</tr>
<tr>
<td>Measurement</td>
<td>.84</td>
</tr>
</tbody>
</table>

\( n = 4675 \)

Conclusions

For most high-stakes assessments, a reliability of .7 or higher is usually considered a reasonable level for strong reliability. Since *DOMA BMS* is a formative assessment, the levels indicated for all sub-tests more than demonstrate a high level of reliability.
Introduction

According to Allen & Yen (1979), reliability can be defined as the consistency between the observed scores on an assessment and the true scores. There are multiple methods of assessing the reliability of an assessment. One way is the test-retest method, comparing the consistency between one administration and a subsequent second administration of the same assessment. Another method of determining reliability is by using internal consistency, a measure of the consistency of results of items within one test. This report will present the internal consistency results for Diagnostic Online Math Assessment: Basic Math Skills (DOMA BMS) (Let’s Go Learn, Inc.).

Methods

Let’s Go Learn, Inc., a developer of online diagnostic reading and math assessments, has created an online diagnostic assessment to evaluate students’ math abilities. Adaptive Diagnostic Assessment of Mathematics (ADAM) K-7 assesses students across five major strands of mathematics: Number & Operations, Measurement, Data Analysis, Geometry, and Algebraic Thinking. Furthermore, the assessment is built upon a highly adaptive-logic platform, limiting the number of items seen by each student by both determining start points, variable jump sizes, and early termination. A total of 44 sub-tests make up the five strands of mathematics. For each sub-test, there is one version of the assessment. Sub-tests are organized into a linear scope and sequence of teachable math skills that match how students learn mathematics in the classroom.
In order to evaluate the reliability of *ADAM K-7*, the Rasch model was used to evaluate overall model fit for each strand. A sample of 28,226 students in grades K-12 taken from 20 school districts nationwide was used.

**Analysis**

Analysis showed overall unidimensionality of each sub-test as well as good overall model fit. Reliability levels were strong (see Table 1).

**Table 1**

*ADAM K-7 Sub-test Reliability (a)*

<table>
<thead>
<tr>
<th>Strand</th>
<th>Reliability (a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numbers &amp; Operations</td>
<td>.70</td>
</tr>
<tr>
<td>Measurement</td>
<td>.61</td>
</tr>
<tr>
<td>Algebraic Thinking</td>
<td>.52</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>.37</td>
</tr>
<tr>
<td>Geometry</td>
<td>.78</td>
</tr>
</tbody>
</table>

\( n = 28,226 \)

**Conclusions**

For most high-stakes assessments, a reliability of .7 or higher is usually considered a reasonable level for strong reliability. Since *ADAM K-7* is a formative assessment, the levels indicated for all strands more than demonstrate a high level of reliability.
DOMA: Pre-Algebra Internal-Consistency Reliability Study

Introduction

According to Allen & Yen (1979), reliability can be defined as the consistency between the observed scores on an assessment and the true scores. There are multiple methods of assessing the reliability of an assessment. One way is the test-retest method, comparing the consistency between one administration and a subsequent second administration of the same assessment. Another method of determining reliability is by using internal consistency, a measure of the consistency of results of items within one test. This report will present the internal consistency results for Diagnostic Online Math Assessment: Pre-Algebra (DOMA: Pre-Algebra) (Let’s Go Learn, Inc.).

Methods

Let’s Go Learn, Inc., a developer of online diagnostic reading and math assessments, has created an online diagnostic assessment to evaluate students’ math abilities, Diagnostic Online Math Assessment: Pre-Algebra (DOMA: Pre-Algebra). DOMA Pre-Algebra assesses students across 14 areas of math knowledge: integer operations, fraction operations, decimal operation, comparing and converting, estimating and rounding, evaluating exponents, ratios and proportions, simplifying expressions, coordinate graphing, linear functions, simple equations, geometry, interpreting data, and simple probability. The pre-screening is administered first as part of the OAASIS platform of adaptive test-taking, while the math facts and reading comprehension sections may be triggered by performance on the assessment. Further, the assessment is built on an adaptive-logic platform, limiting the number of items seen by each student by both
determined start point and by early termination. For each sub-test, there is one version of the assessment.

In order to evaluate the reliability of DOMA Pre-Algebra, the Rasch model was used to evaluate overall model fit for each sub-test. A sample of 62,631 students in grades K-12 taken from school districts nationwide was used.

Analysis

Analysis showed overall unidimensionality of each sub-test as well as good overall model fit. The overall internal-consistency reliability level for DOMA: Pre-Algebra was .81, which demonstrates strong levels of reliability. Reliability levels were generally strong (see Table 1), although some individual sub-test reliability levels were questionable.

Table 1

DOMA Pre-Algebra Sub-test Reliability (α)

<table>
<thead>
<tr>
<th>Sub-test</th>
<th>Reliability (α)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integer Operations</td>
<td>.80</td>
</tr>
<tr>
<td>Fraction Operations</td>
<td>.82</td>
</tr>
<tr>
<td>Decimal Operation</td>
<td>.79</td>
</tr>
<tr>
<td>Comparing and Converting</td>
<td>.61</td>
</tr>
<tr>
<td>Estimating and Rounding</td>
<td>.57</td>
</tr>
<tr>
<td>Evaluating Exponents</td>
<td>.68</td>
</tr>
<tr>
<td>Ratios and Proportions</td>
<td>.05</td>
</tr>
<tr>
<td>Simplifying Expressions</td>
<td>.37</td>
</tr>
</tbody>
</table>
For most high-stakes assessments, a reliability of .7 or higher is usually considered a reasonable level for strong reliability. Since DOMA Pre-Algebra is a formative assessment, the levels indicated for most sub-tests more than demonstrate a high level of reliability.
ADAM K-7 Test-Retest Reliability Study (2012)

Test-Retest is the ability of a test to be taken once and then immediately again and have similar results. Let’s Go Learn undertook a test-retest analysis in the Fall of 2012 with the ADAM K-7 assessment. Students were given the assessment two times over a one week period. The results were again excellent. Variability was low, meaning that the ADAM K-7 assessment is very precise and can be re-administered with low bias.

Sample size: n=50

The reliability coefficient based on the above test-retest reliability study is shown in the table below. These values are statistically significant and represent a high level of reliability.

<table>
<thead>
<tr>
<th>Subtest</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.764</td>
</tr>
<tr>
<td>2</td>
<td>.738</td>
</tr>
<tr>
<td>3</td>
<td>.482</td>
</tr>
<tr>
<td>4</td>
<td>.669</td>
</tr>
<tr>
<td>5</td>
<td>.687</td>
</tr>
<tr>
<td>6</td>
<td>.741</td>
</tr>
<tr>
<td>7</td>
<td>.756</td>
</tr>
<tr>
<td>8</td>
<td>.718</td>
</tr>
<tr>
<td>9</td>
<td>.813</td>
</tr>
<tr>
<td>10</td>
<td>.599</td>
</tr>
</tbody>
</table>
DOMA Pre-Algebra Test-Retest Reliability Study (2007)

Test-Retest is the ability of a test to be taken once and then immediately again and have similar results. Let’s Go Learn undertook a test-retest analysis in Q3 2007 with the DOMA: Pre-Algebra assessment. Students were given the assessment two times over a one week period. The results were again excellent. Variability was low, meaning that the DOMA: Pre-Algebra assessment is very precise and can be re-administered with low bias. Sample size: n=230

DOMA: Pre-Algebra assessment is composed of 14 constructs that define pre-algebra. In the individual student detailed report and in the classroom summary reports, teachers look at how students perform across these 14 constructs. Summarily, they can see mastery, partial, or non-mastery in each of these constructs for each student. For the analysis of this study, mastery of a construct was denoted by a 1, partial mastery with a 0.5, and non-mastery with a 0. So the highest raw score possible was a 14 and the lowest raw score was a 0. Translating this raw score to a percent mastery, the range of scores was 0% to 100%.

Analyzing the change in scores, the variance of this sample was 2.28. This translates to an average score variance as a percentage of 16.3%. This represents low variance and supports the statement that DOMA: Pre-Algebra can be use as a pre-post measure without significant bias.

The reliability coefficient based on the above test-retest reliability study is .844, which is statistically significant, and represents a high level of reliability.
DOMA: Algebra Internal-Consistency Reliability Study

Introduction

According to Allen & Yen (1979), reliability can be defined as the consistency between the observed scores on an assessment and the true scores. There are multiple methods of assessing the reliability of an assessment. One way is the test-retest method, comparing the consistency between one administration and a subsequent second administration of the same assessment. Another method of determining reliability is by using internal consistency, a measure of the consistency of results of items within one test. This report will present the internal consistency results for Diagnostic Online Math Assessment: Algebra (DOMA: Algebra) (Let’s Go Learn, Inc.).

Methods

Let’s Go Learn, Inc., a developer of online diagnostic reading and math assessments, has created an online diagnostic assessment to evaluate students’ math abilities, Diagnostic Online Math Assessment: Algebra (DOMA: Algebra). DOMA Algebra assesses students across 11 areas of Algebra I knowledge: Evaluating Advanced Exponents; Solving Linear Equations; Graphing and Analyzing Linear Equations; Relations and Functions; Solving and Graphing Inequalities; Solving and Graphing Systems; Polynomial Operations; Factoring Polynomials; Radical Expressions and Equations; Quadratic Equations; and Rational Expressions and Equations. The pre-screening is administered first as part of the OAASIS platform of adaptive test-taking, while the math facts and reading comprehension sections may be triggered by performance on the assessment. Further, the assessment is built on an adaptive-logic
platform, limiting the number of items seen by each student by both determined start point and by early termination. For each sub-test, there is one version of the assessment.

In order to evaluate the reliability of *DOMA Algebra*, the Rasch model was used to evaluate overall model fit for each sub-test. A sample of 62,631 students in grades K-12 taken from school districts nationwide was used.

**Analysis**

Analysis showed overall unidimensionality of each sub-test as well as good overall model fit. The overall internal-consistancy reliability level for *DOMA: Algebra* was .64, which demonstrates moderate levels of reliability. Reliability levels were generally moderate (see Table 1), although some individual sub-test reliability levels were questionable.

**Table 1**

*DOMA Algebra Sub-test Reliability (α)*

<table>
<thead>
<tr>
<th>Sub-test</th>
<th>Reliability (α)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluating Advanced Exponents</td>
<td>.46</td>
</tr>
<tr>
<td>Solving Linear Equations</td>
<td>.42</td>
</tr>
<tr>
<td>Graphing and Analyzing Linear Equations</td>
<td>.42</td>
</tr>
<tr>
<td>Relations and Functions</td>
<td>.41</td>
</tr>
<tr>
<td>Solving and Graphing Inequalities</td>
<td>.34</td>
</tr>
<tr>
<td>Solving and Graphing Systems</td>
<td>.51</td>
</tr>
<tr>
<td>Polynomial Operations</td>
<td>.44</td>
</tr>
<tr>
<td>Factoring Polynomials</td>
<td>.27</td>
</tr>
</tbody>
</table>
Conclusions

For most high-stakes assessments, a reliability of .7 or higher is usually considered a reasonable level for strong reliability. The overall reliability coefficient for DOMA Algebra is .64, and since DOMA Algebra is a formative, low-stakes assessment, that level demonstrates a high level of reliability. In addition, the Pre-Screener, which is used to drive the overall adaptive logic of DOMA Algebra is .70, which demonstrates a high level of reliability. The levels indicated for most sub-tests demonstrate a moderate level of reliability as well; however, there are some subtests that demonstrate a notably lower level of reliability. For diagnostic purposes, and in creating a profile of student strengths and weaknesses, the reliability levels demonstrated by the individual subskills are still high enough to indicate a confident level of test reliability. It should further be considered, however, that these subskills are not individual testing sections, but only reporting sections.

There are a number of explanations as to why the reliability of the subskill sets decreased when examined separately. First of all, it is not unexpected to see reliability levels decrease as the number of items assessed decreases. When the subskill sets are small, there is less variability in the student responses. Further, there is the issue of missing data, which will effect the reliability levels. Since no one student ever takes all test items, there is always missing data. In the reliability of the overall assessment, the
missing data is less of a factor than in the reliability of the individual subskills, due again
to the smaller number of items being measured. The missing data there serves the restrict
variability even further, in turn increasing the standard error and increasing the mean
square error as well.
References