SRI International

Alternate Assessment Design–Mathematics

Technical Report 7: Pilot Task Tryouts

Design and Analysis

March 31, 2011

University of Oregon
Patricia Almond

SRI International
Kavita L. Seeratan
Katherine Nagle
Renée Cameto
Geneva Haertel
Angela DeBarger
Kathryn Morrison
Reina Fujii

SRI Project No.: P18879

This report was supported in part by the Elementary and Secondary Education Act, Enhanced Assessment Grant (ESEA grant No. # S368A08008) through a partnership with the Utah (Lead), Idaho, and Florida State Departments of Education and SRI International as the research entity. Any opinions, findings, or conclusions are those of the authors and do not necessarily reflect the view of ESEA or the U.S. Department of Education.

Contents

Figures

Tables

Alternate Assessment Design–Mathematics Technical Report 7: Pilot Task Tryouts. Design and Analysis

Introduction

This Alternate Assessment Design-Mathematics project supported by an Enhanced Assessment Grant (EAG) is innovative in two respects: it applies evidence centered design (ECD), for the first time, to academic assessments for students with significant cognitive disabilities (SCD) and it integrates ECD and the principles of universal design for learning (UDL) into designing and developing alternate assessments based on alternate achievement standards (AA-AAS). This work extends current knowledge in the field regarding content-centered assessment design by providing a model for future AA-AAS design and development (Cameto, Haertel, DeBarger, & Morrison, 2010). The model begins with grade-level academic content and anticipates possible threats that may arise in the form of construct irrelevant variance (CIV) attributable to a student's disability.

With respect to the teaching, learning, and assessing of academic content, students with SCD challenge conventions of test design and development. We know, however, that assessment can be influential in raising expectations and influencing instruction. The process of designing and developing alternate assessments is essentially new, beginning with the Individuals with Disabilities Education Act in 1997 and has not yet established a record of accomplishment of technical adequacy. Eligible students and assumptions about measuring their achievement plus the variation in design, development, and implementation procedures make traditional assessment design approaches inadequate (Gong & Marion, 2006; Ryan, Quenemoen, & Thurlow, 2004; U.S. Government Accountability Office, 2009). The methods used to date in designing alternate assessments and selecting their targeted content are varied but typically do not match the same technical rigor used for designing general education assessments (Bechard, 2005). This project applied ECD and UDL in the design of tasks for AA-AAS. This work extends current knowledge in the field and provides a model for future alternate assessment development using content-centered assessment design.

Evidence Centered Design

ECD (e.g., Mislevy, Steinberg, & Almond, 2003) uses a rigorous and replicable assessment design process that carefully considers how the content, task, and learner characteristics interact in the creation of assessment tasks. In this approach, co-design teams bring together the expertise of assessment specialists, special educators, and content-area specialists to create design patterns, task templates, task specifications, and exemplar tasks. In this project, the co-design team members were specialists in large-scale and formative assessment, special educators of SWCD, and mathematics educators with experience in instruction and assessment. Expertise was also contributed by representatives from each state department of education, by research staff, and by nationally recognized experts in special education, mathematics, and assessment.

The ECD process involves five layers of activities (Mislevy & Haertel, 2006) on the identification of the substantive domain to be assessed; the assessment argument, the structure of assessment elements such as tasks, rubrics, and psychometric models; the implementation of these elements; and the way they function in an operational assessment, as described below.

1. **Domain Analysis** involves determining the specific content to be included in the assessment. Use of state content standards and the pending common core standards are examples of domain analyses.
2. In **Domain Modeling**, a high-level narrative description of the knowledge, skills, and abilities (KSAs) to be assessed (targeted KSAs), the evidence that needs to be collected, and the features of the tasks that will elicit the evidence are specified in detail. Ancillary or additional KSAs required to respond correctly to a task (not the target) are also specified (e.g., reading skills in a math test). In this way, CIV can be minimized—potential barriers can be removed or their effects reduced by providing appropriate access. The document produced in this layer is called a Design Pattern.
3. The **Conceptual Assessment Framework** is developed from the narrative design patterns and contains technical specifications required for the student, evidence, and task models. In this layer, the "nuts and bolts" of the proposed assessment are specified. For example, evidence to be collected in potential observations, the psychometric measurement submodel and the evaluation submodel that specifies potential rubrics to be used in scoring the tasks are documented.
4. **Implementation involves the authoring of the assessment items or tasks** using Design Patterns and the conceptual assessment framework just described. In this layer the item directives or prompts are created, distracters for multiple choice items are developed, stimulus materials to be used in the item are determined. In addition, the scoring rubrics to be used are created and the scoring process is specified.
5. In **Delivery**, the processes for the assessment administration are specified and reporting and score reports are planned.

Universal Design for Learning

Principles of universal design (UD, Mace, 1991) were used to address the challenge of designing and delivering tests that are accessible to and valid for a wide range of students with SCD. Dolan and Hall (2001, 2007) proposed applying UDL so that tests would minimize potential sources of CIV by supporting the ways that students with a diverse set of characteristics interact with the assessment process. Thompson, Johnstone, and Thurlow (2002) adapted Mace's original elements from architecture to derive seven elements of accessible and fair tests: "(1) inclusive assessment population, (2) precisely defined constructs, (3) accessible, nonbiased items, (4) items amendable to accommodations, (5) simple, clear, and intuitive instructions and procedures; (6) maximum readability and comprehensibility, and (7) maximum legibility" (p. 1).

For this study, evidence-centered design (ECD) was used to design and develop tasks for alternate assessments judged against alternate academic achievement standards (AA-AAS) in mathematics. The characteristics of 30 exemplar tasks are briefly described as is the performance of participating students with significant cognitive disabilities. Variation in the performance of participating students on the exemplar tasks produced through the application of ECD is examined, for example, range of communication levels, frequency of task administration across the 30 tasks, and patterns of responding across the 606 task administrations. The results of this analysis will inform the project of lessons learned from the application of ECD to the design and development of exemplar tasks for an AA-AAS in mathematics. This report informs two of the project's major questions:

1. To what extent are the exemplar tasks ($N = 30$) workable as a foundation for demonstrating the application of ECD to AA-AAS in mathematics?
2. Do the 30 exemplar tasks and items address an appropriate range of challenge and accessibility for the varied communication levels of eligible students?

Item Tryouts

The project integrated ECD and UDL and produced 30 AA-AAS exemplar tasks using the ECD process. This study constitutes a proof of concept, a milestone on the way to a fully functioning model that applies ECD to AA-AAS. The Task Tryouts allowed researchers to examine an implementation of the 30 exemplar tasks and corresponding items with a focus on identifying refinements that could be applied to the 30 design patterns and exemplar tasks[1] developed by the project.

Methods

The Task Tryouts study was exploratory and evaluative, focusing on an initial implementation of the exemplar tasks and items. The operational items were tried out with teachers and students with significant cognitive disabilities eligible for AA-AAS in school settings to ensure usefulness and to identify any difficulties that warrant refinement to the exemplar tasks and corresponding items. After refinements to the tasks and corresponding items, a logical next step would be a larger pilot test with a representative sample of AA-AAS eligible students. This study reports the tryouts *only,* and results are primarily descriptive in nature.

Components

Exemplar Tasks. Each exemplar task was displayed within a task template, which described the KSAs to be measured and the relationships among them. It also included descriptions of the stimulus materials presented to students, available tools and technologies, and other aspects of the assessment. Task templates distinguish between targeted and additional KSAs. Potential sources of CIV are isolated and supports are identified to assist students in overcoming barriers presented by the additional KSAs. The exemplar task also presents three versions (items) of the task referred to as item A1 (least complex), item B (more complex), and item C (most complex).The exemplar task specifies an item directive, materials, the correct response, and supports that might mitigate barriers. The 30 tasks cover three grades spans (3 to 5, 6 to 8, and high school) and five math strands (see Table 1 for a distribution of tasks).

Task and Corresponding Items for Implementation. Each exemplar task contains three items with item C most challenging and item A1 least challenging. Figure 2 provides an example of the structure of one exemplar task with three corresponding items. For the item tryouts, the least challenging item (A1) was administered first. If item A1 was answered correctly, the teacher moved on to item B (the next most challenging item within the task). If item B was administered then item C was also administered. Figure 1 displays the decisionmaking flow chart describing the flow of item administration.

Sample—Tasks and Students. For the pilot study, an assessment protocol included directions for the teacher or test administrator, a script for the teacher clarifying what to say and

[1] The use of design patterns and exemplar tasks in this project is described extensively in other reports from this project and can be found at http://alternateassessmentdesign.sri.com.

what materials to present during testing, how many attempts to perform the task a student is afforded, how evidence or student responses are to be recorded and gathered. Each state identified a pilot study coordinator had participated in the task design and who was trained by the SRI project staff to oversee the pilot study in their state. The pilot study coordinators then trained the teachers identified to conduct the pilot study in their state.

Unique Tasks. Each state was assigned seven unique tasks to administer independently. The state produced the assessment protocol according to project instructions to ensure that the core tasks are administered consistently across the three states. Teachers administered each task to six students. A total of 21 tasks were administered.

Core Tasks. Each state also administered each of nine core tasks to six students. The SRI project staff developed and produced the assessment protocol to ensure that the core tasks are administered in a consistently across the three states.

Results

Student Sample

Three states recruited students (*N* = 186) to participate in the task tryouts: State A (*n* = 70), State B (*n* = 98), State C (*n* = 18). Nearly half (48%) of participating students were classified as having an intellectual disability, 20% having autism, 20% having multiple disabilities, and the remaining 10% were distributed across the remaining federal disability categories. Participating students were enrolled across the range of assessment and accountability grades (3 through 8 and high school): elementary (39%), middle school (28%), and high school (29%). Participating students were classified at three communication levels: pre-symbolic (23%), emerging symbolic (29%), and symbolic (48%) (Browder, Flowers, & Wakeman, 2008). Tables 1, 2, and 3 below provide more detail about disability, grade, and communication levels of participating students.

Task Sample

There were 30 exemplar tasks administered during the AAD-Math task tryouts. Across all tasks, there were 631 tasks administered. Each participating state administered a set of nine common "core" tasks plus seven additional tryout tasks per state, that is, each state conducted tryouts on 16 tasks. For the purposes of this study, 606 task administrations were included in the analysis. There were 29 records excluded from the analyses due to incomplete data. Four data records had no valid response for task A-1, 7 had no valid student identifier, and 18 had no age or grade designation. Across the five mathematics strands and the three grade levels, there were 342 administrations of core tasks and 264 administrations of noncore tasks included in these analyses. Table 5 provides details for the breakdown of administrations by strand, grade, and type. Additional detail about each task is provided in tables 5a and 5b.

Findings

Response Patterns

The 606 task administrations were examined based on patterns of responding (refer to Table 6). There were 133 instances in which the initial task (A-1) was responded to correctly and students went on to respond to a more complex version of the task, for example, task B, task C, or both tasks B and C correctly. This group of task administrations was labeled A1 Plus and represented the higher-performing task administrations. Another group of task administrations was classified as A1 Correct Only (*N* = 150) and included task administrations in which students responded correctly to the less complex task A1 but did not respond correctly to either of the more complex items correctly. A third group of task administrations was classified as A1 Incorrect or No Response or Refusal and included tasks in which students did not respond correctly to the initial task A1. The response patterns were coded based on response codes within the task response data set where "1" represented a correct response, "2" represented an incorrect response, "3" represented no response, and "4" represented refusal. Missing data were recoded to a value of "9" and the values for the three items were concatenated to produce a three character pattern label in the format: "111" when responses to A1, B, and C were all correct, "122" when the response to item A1 was correct and the responses to items B and C were incorrect. The unique patterns are portrayed within the broader groupings in Table 7.

There were 133 task administrations in which participating students responded to items B and/or C. Further examination showed that these task administrations represented some students who took more than one task (*N* = 73) across all three grade ranges, for example, Gr 3 to 5 (*n* = 76), Gr 6 to 8 (*n* = 21), Gr 9 to 12 (*N* = 36), and that all three states contributed students in this group. There were 23 exemplar tasks in which participating students performed correctly on the more complex items and the 23 items were from all five strands of mathematics.

Task Performance on Complex Tasks

In order to get a better understanding about how students performed on the most complex task, a cross-tabulation was produced showing student responses to item C (the most complex item) for every task administered. For this analysis, core and noncore tasks are presented in separate tables (Tables 8a and 8b). There were 36 (11%) core task administrations in which participating students responded correctly to item C. There were 15 (6%) noncore task administrations in which participating students responded correctly to item C. Further investigation into performance patterns and student characteristics for students performing at different levels of complexity are warranted.

Conclusions and Discussion

For the AA-AAS task tryouts, preliminary results show that for a portion of tasks administered during the task tryouts (fall 2010), there were 133 out of 606 task administrations in which tasks were responded to correctly at the more complex levels of challenge, that is, the challenging items were accessible and were responded to correctly by a portion of the eligible students, the students with significant cognitive disabilities.

These analyses are preliminary. More detailed examination is being conducted in order to draw inferences and to address the following questions.

- Had the students received instruction on the mathematics tasks that were administered?
- Were the directions for teachers easy to understand?
- What were the teachers' perceptions of the feasibility of administration of the tasks and materials?

Data were gathered during the task tryouts in the form of teacher surveys to supplement the student data. The survey data are being examined to expand our understanding about how teachers viewed their experience, the task and item administration, and factors that influenced their selection of participating students.

Tables

Table 1. Primary Disability (*n* = 186)

Primary Disability	Percent
Intellectual disability (1)	48%
Autism (2)	20%
Low incidence (3,4,9,10)*	5%
Other (5,6,7)*	2%
Multiple disabilities (11)	20%
Other health impairment (12)	5%
Total	100

* Low incidence: Hearing/deaf, traumatic brain injury, deaf-blindness, orthopedic impairment.
* Other: Emotional disturbance, speech/language, specific learning disabilities.

Table 2. Grade of Enrollment

Level	Grade	Frequency	Percent	Frequency	Percent
Elementary	3	15	8%	73	39%
	4	25	13%		
	5	33	18%		
Middle School	6	13	7%	53	28%
	7	24	13%		
	8	16	9%		
High School	9	15	8%	54	29%
	10	16	9%		
	11	8	4%		
	12	15	8%		
Unknown	NA	6	3%	6	3%
	Total	186	100%		

NA = Not available.

Table 3. Grade Level by Communication Level

Grade Group		Pre-Symbolic	Early Symbolic	Symbolic	Total
Elementary	Count	13	25	35	73
	Row %	18%	34%	48%	39%
Middle School	Count	10	11	32	53
	Row %	19%	21%	60%	29%
High School	Count	16	17	21	54
	Row %	30%	32%	39%	29%
Unknown	Count	3	1	2	6
	Row %	50%	17%	33%	3%
Total	Count	42	54	90	186
	Row %	23%	29%	48%	100%

Table 4. Number of Tasks Administered— Item Which Obtained a Valid Score

Item A1 Response	Frequency	Percent
Correct (1)	283	46.7
Incorrect (2)	243	40.1
No Response (3)	57	9.4
Refused (4)	23	3.8
Total	606	100

Table 5. Task Administrations by Strand, Grade Level, and Group (*N* = 606)

		Strand					
Tasks	Grade Level	Algebra	Data Analysis	Geometry	Measurement	Numbers Operations	Total
Core	3-5	29		46	43	28	146
	6-8	30		30		29	89
	9-12	54		53			107
	Subtotal	113	0	129	43	57	342
Unique	3-5	31	21	6	10	77	145
	6-8		16		19	23	58
	9-12		13		31	17	61
	Subtotal	31	50	6	60	117	264
	Total	144	50	135	103	174	606

Table 5a. Frequencies for Each Core Task (*N* = 342)

Task Level	Task Description	Strand: Algebra	Data Analysis	Geometry	Measure-ment	Numbers Operations	Total
Gr 3-5	ALG B2: Represent variables using letters or symbols (2)	29					29
	GEO A1: Identify, compare, analyze 2 & 3 dimensional shapes (6)			46			46
	MSR B2: Select & apply appropriate tools to measure (8)				43		43
	NUM A3: Understand fractions (12)					28	28
Total		29		46	43	28	146
Gr 6-8	ALG B1: Develop concept different uses of variables (17)	30					30
	GEO A1: Describe, classify-2 & 3 dimensional shapes (19)			30			30
	NUM A1: Fractions, decimals, percents (22)					29	29
Total		30		30		29	89
Gr 9-12	ALG B3: Use symbolic algebra in math relationships (25)	54					54
	GEO A1: Analyze-attributes 2 & 3 dimensional shapes (27)			53			53
Total		54		53			107

Table 5b. Frequencies for Each Unique Task (*N* = 264)

		Strand					
Task Level	Task Description	Algebra	Data Analy-sis	Geo-metry	Measure-ment	Numbers Operations	Total
Gr 3-5							
1	ALG A1: Geometric & numeric patterns (1)	15					15
3	ALG C1: Model problems using objects/representations (3)	16					16
4	DAT A3: Represent data in tables, graphs (4)		6				6
5	DAT B1: Describe & compare datasets (5)		15				15
7	GEO A4: Explore congruence & similarity (7)			6			6
9	MSR B3: Use benchmarks to estimate measurements (9)				10		10
10	NUM A1: Understand place value base ten number systems (10)					15	15
11	NUM A2: Recognize/generate equiv representations (11)					11	11
13	NUM A4: Use models, benchmarks judge size of fractions (13)					6	6
14	NUM B1: Understand multiplication & division (14)					6	6
15	NUM B3: Identify/use relationships between operations (15)					21	21
16	NUM C2: Develop fluency add, subtract, multiply, divide (16)					18	18
Total		31	21	6	10	77	145
Gr 6-8							
18	DAT A2: Select-use graphic data representations (18)		16				16
20	MSR A2: Understand & convert units (20)				5		5
21	MSR B2: Select-tools/techniques to measure w/precision (21)				14		14
23	NUM A7: Integers to represent & compare quantities (23)					17	17
24	NUM C4: Solve problems involving proportions (24)					6	6
Total			16		19	23	58

		Strand					
Task Level	Task Description	Algebra	Data Analysis	Geometry	Measurement	Numbers Operations	Total
Gr 9-12							
26	DAT B1: Analyze univariate measurement data (26)		13				13
28	MSR A1: Make decisions-units/scales to solve problems (28)				31		31
29	NUM A1: Understand very large/very small numbers (29)					6	6
30	NUM C1: Fluency w/real numbers, vectors, matrices (30)					11	11
Total			13		31	17	61

Table 6. Performance Estimates (High, Mid, Low) Based on Response Patterns

Description	Frequency	Percent
A1 Plus B &/or C (High Performance)	133	22%
A1 Correct Only (Mid Performance)	150	25%
A1 Incorrect/No Response/Refusal (Low)	283	47%
Total	606	100%

Table 7. Frequencies for Patterns of Responding Across Three Items A1, B, and C

Description	Pattern	Frequency	Percent
• All 3 correct	111	31	5.1
• A & B correct	112	78	12.9
• A & B correct	113	2	0.3
• A & B correct	114	1	0.2
• A & B correct	119	1	0.2
• A Correct C Correct	191	2	0.3
• A Correct C Correct	121	18	3
A1 Plus B &/or C		133	22%
• A Correct B Incorrect	122	109	18
• A Correct B Incorrect	123	10	1.7
• A Correct B Incorrect	129	14	2.3
• A Correct B No Response	132	2	0.3
• A Correct B No Response	133	8	1.3
• A Correct B No Response	134	1	0.2
• A Correct B No Response	139	1	0.2
• A Correct B Refused	142	1	0.2
• A Correct B No Data	192	1	0.2
• A Correct B No Data	193	1	0.2
• A Correct B No Data	199	2	0.3
A1 Correct Only		150	25%
• A Incorrect B & C No Data	299	243	40.1
• A No Response B & C No Data	399	57	9.4
• A Refusal B & C No Data	499	23	3.8
A1 Incorrect or No Response/Refusal		323	53%
Total		606	100%

Table 8a. Core Tasks—Common to Three States

		Response to Item C					
		Correct	Incorrect	No Response	Refused	No Data	Total
ALG B2: Represent variables using letters or symbols (2)	Count		12	1		16	29
	% in Task		41%	3%		55%	100%
GEO A1: Identify, compare, analyze 2 & 3 dimensional shapes (6)	Count	1	23	4		18	46
	% in Task	2%	50%	9%		39%	100%
MSR B2: Select & apply appropriate tools to measure (8)	Count	11	14	2		16	43
	% in Task	26%	33%	5%		37%	100%
NUM A3: Understand fractions (12)	Count	11	1			16	28
	% in Task	39%	4%			57%	100%
ALG B1: Develop concept different uses of variables (17)	Count		14	1		15	30
	% in Task		47%	3%		50%	100%
GEO A1: Describe, classify-2 & 3 dimensional shapes (19)	Count	3	9	2		16	30
	% in Task	10%	30%	7%		53%	100%
NUM A1: Fractions, decimals, percents (22)	Count		10			19	29
	% in Task		35%			66%	100%
ALG B3: Use symbolic algebra in math relationships (25)	Count	6	13			35	54
	% in Task	11%	24%			65%	100%
GEO A1: Analyze-attributes 2 & 3 dimensional shapes (27)	Count	4	17	3	1	28	53
	% in Task	8%	32%	6%	2%	53%	100%
Total	Count	36	113	13	1	179	342
	% in Task	11%	33%	4%	0%	52%	100%

Table 8b. Noncore Tasks—Individual State Tasks

		Response to Item C					
		Correct	Incorrect	No Response	Refused	No Data	Total
ALG A1: Geometric & numeric patterns (1)	Count		11	2		2	15
	% in Task		73%	13%		13%	100%
ALG C1: Model problems using objects & representations (3)	Count	2	4	1		9	16
	% in Task	13%	25%	6%		56%	100%
DAT A3: Represent data in tables, graphs (4)	Count					6	6
	% in Task					100%	100%
DAT B1: Describe & compare data sets (5)	Count	1	9			5	15
	% in Task	7%	60%			33%	100%
GEO A4: Explore congruence & similarity (7)	Count					6	6
	% in Task					100%	100%
MSR B3: Use benchmarks to estimate measurements (9)	Count		5			5	10
	% in Task		50%			50%	100%
NUM A1: Understand place value base ten number systems (10)	Count		1			14	15
	% in Task		7%			93%	100%
NUM A2: Recognize/generate equivalent representations (11)	Count	2				9	11
	% in Task	18%				82%	100%
NUM A4: Use models, benchmarks judge size of fractions (13)	Count					6	6
	% in Task					100%	100%
NUM B1: Understand multiplication & division (14)	Count					6	6
	% in Task					100%	100%
NUM B3: Identify/use relationships between operations (15)	Count	6	2	1		12	21
	% in Task	29%	10%	5%		57%	100%
NUM C2: Develop fluency add, subtract, multiply, divide (16)	Count		10	1		7	18
	% in Task		56%	6%		39%	100%
DAT A2: Select-use graphic data representations (18)	Count		2			14	16
	% in Task		13%			88%	100%
MSR A2: Understand & convert units	Count		3		1	1	5

		Response to Item C					
		Correct	Incorrect	No Response	Refused	No Data	Total
(20)	% in Task		60%		20%	20%	100%
MSR B2: Select-tools/techs measure w/precision (21)	Count	1	4			9	14
	% in Task	7%	29%			64%	100%
NUM A7: Integers to represent & compare quantities (23)	Count		3			14	17
	% in Task		18%			82%	100%
NUM C4: Solve problems involving proportions (24)	Count		1			5	6
	% in Task		17%			83%	100%
DAT B1: Analyze univariate measurement data (26)	Count	1	6	1		5	13
	% in Task	8%	46%	8%		39%	100%
MSR A1: Make decisions-units/scales to solve problems (28)	Count	1	13	2		15	31
	% in Task	3%	42%	7%		48%	100%
NUM A1: Understand very large/very small numbers (29)	Count	1	1			4	6
	% in Task	17%	17%			67%	100%
NUM C1: Fluency w/real numbers, vectors, matrices (30)	Count		3			8	11
	% in Task		27%			73%	100%
Total	Count	15	78	8	1	162	264
	% in Task	6%	30%	3%	0%	61%	100%

Figures

Figure 1. Flowchart of Task Administration

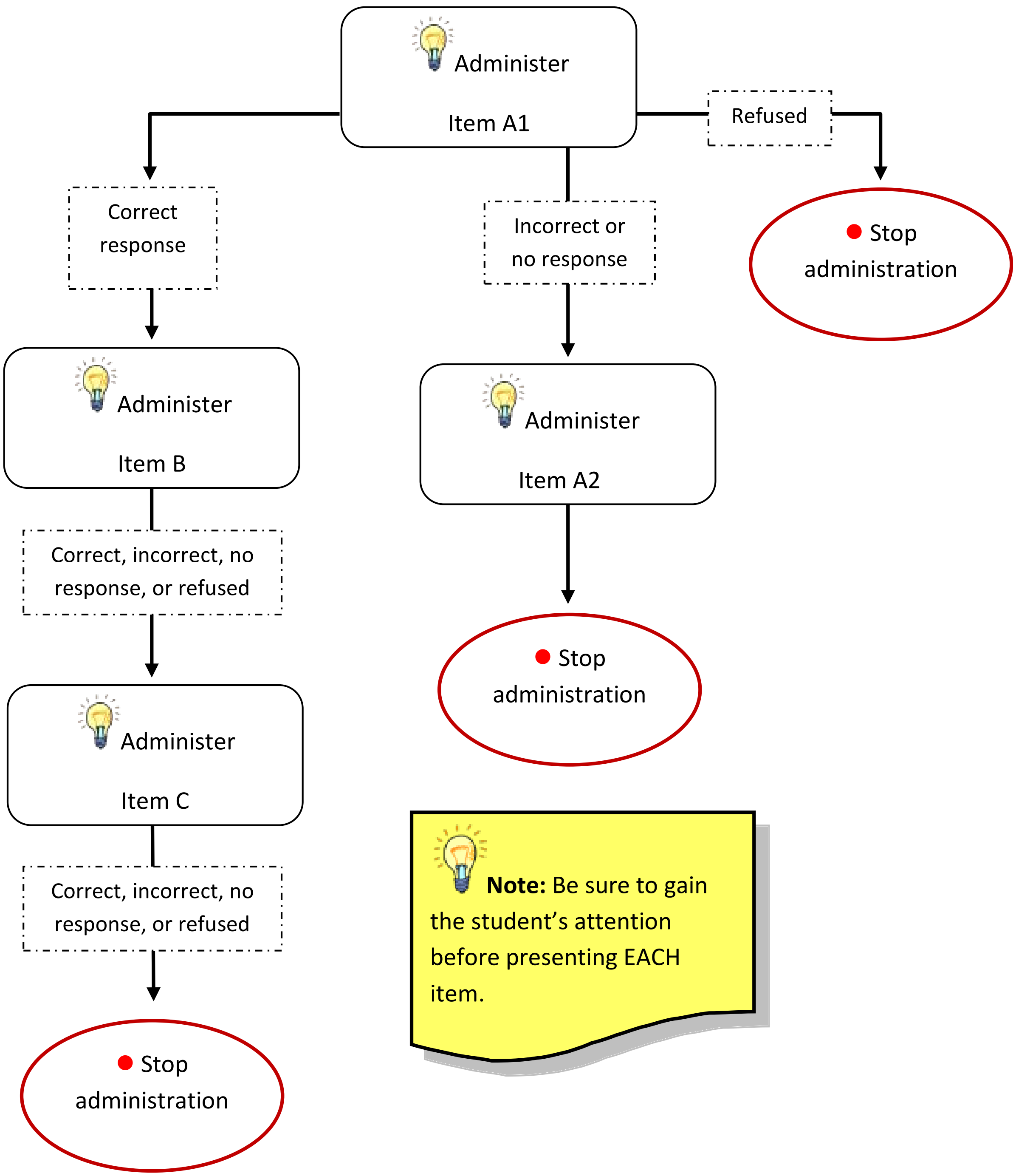


Figure 2. Administration Directions for One AAD-Math Exemplar Task *(Training PowerPoint)*

1

Anatomy of a Task: Geometry

Attributes	General Information
Grade Band	Grades 6 to 8
Task Code	GeoA1 g6-8
Assessment Target	Precisely describe, classify, and understand relationships among types of two- and three-dimensional objects using their defining properties
Rationale: Why is this important ?	Analyzing defining properties of geometric shapes leads directly to high school geometry and other kinds of deductive reasoning

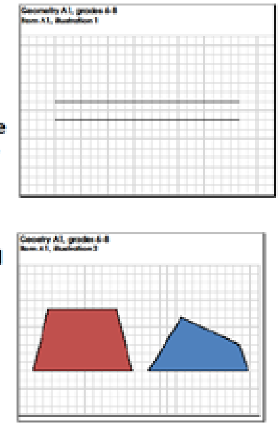
2

Item A1

Directions When the student is attending, teacher/ administrator says, "**These lines are parallel** (teacher indicates drawing of two parallel lines on a grid). **Which of these shapes has sides that are parallel** (teacher indicates handout with two shapes)?"

Materials Illustrations 1 & 2

Correct Response Student indicates the trapezoid



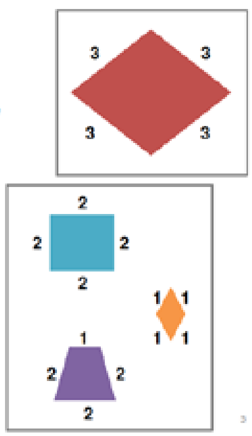
3

Item B

Directions When the student is attending, teacher says, "**All the sides of this shape are the same length. Each side is three inches long** (teacher indicates four sides of rhombus). **Here are three more shapes. Show me which of these shapes also have sides that are all the same length.**"

Materials Illustrations 1 & 2

Correct Response Student identifies the square and the rhombus



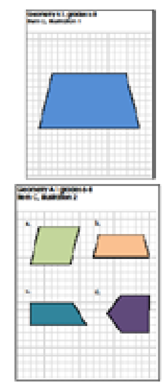
4

Item C

Directions When the student is attending, teacher says, "**Here is a shape with four sides. Only one pair of the sides is parallel** (teacher indicates the four sides and the parallel sides on a trapezoid, then puts the trapezoid aside). **Here are four more shapes** (teacher indicates the handout). **Show me the shapes that also have four sides and only one pair of parallel sides.**"

Materials Illustrations 1 & 2

Correct Response Student identifies the two trapezoids



References

Bechard, S. (2005). *Developing alternate assessments using expanded benchmarks from a nine-state consensus framework in reading, writing, mathematics, and science*. Paper presented at the 2005 National Council on Measurement in Education (NCME), Annual Conference, Montreal, Canada, Montreal, Canada

Browder, D. M., Flowers, C., & Wakeman, S. Y. (2008). Facillitating participation in assessments and the general curriculum: Level of symbolic communication classification for students with significant cognitive disabilities. *Assessment in Education: Principles, Policy, and Practice, 15*(2), 137-151.

Cameto, R., Haertel, G., DeBarger, A., & Morrison, K. (2010). *Alternate assessment design-mathematics technical report 1: Project overview applying evidence-centered design to alternate assessments in mathematics for students with significant cognitive disabilities*. Menlo Park, CA: SRI International. Retrieved from http://alternateassessmentdesign.sri.com/documents/AAD_M_TechRpt1.pdf

Dolan, R. P., & Hall, T. E. (2001). Universal Design for Learning: Implications for large-scale assessment. *IDA Perspectives, 27*(4), 22-25.

Dolan, R. P., & Hall, T. E. (2007). Developing accessible tests with universal design and digital technologies: Ensuring we standardize the right things. In L. L. Cook & C. C. Cahalan (Eds.), *Large-scale assessment and accommodations: What works* (pp. 95-111). Arlington, VA: Council for Exceptional Children.

Gong, B., & Marion, S. (2006). *Dealing with flexibility in assessments for students with significant cognitive disabilities*. Minneapolis: MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from http://cehd.umn.edu/NCEO/ OnlinePubs/Synthesis60.html

Mace, R. L. (1991). *Definitions: Accessible, adaptable, and universal design* (Fact Sheet). Raleigh, NC: Center for Universal Design, NCSU.

Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice, 25*(4), 6-20.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives, 1*, 3-67.

Ryan, J. M., Quenemoen, R. F., & Thurlow, M. (2004). *I say potato, you say potato: The assessment-speak gao between general and alternative assessment experts*. Minneapolis, MN: National Center on Educational Outcomes.

Thompson, S., & Thurlow, M. (2002). *Universally designed assessments: Better tests for everyone!* . Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from http://education.umn.edu/NCEO/OnlinePubs/Policy14.htm

U.S. Government Accountability Office. (2009). *No Child Left Behind Act: Enhancements in the Department of Education's review process could improve state academic assessments*. Washington, DC: Author. Retrieved from http://www.gao.gov/new.items/d09911.pdf