J·T·L·A

The Journal of Technology, Learning, and Assessment

Volume 7, Number 2 · December 2008

Developing a Taxonomy of Item Model Types to Promote Assessment Engineering

Mark J. Gierl, Jiawen Zhou, & Cecilia Alves

www.jtla.org

A publication of the Technology and Assessment Study Collaborative
Caroline A. & Peter S. Lynch School of Education, Boston College

Volume 7, Number 2

Developing a Taxonomy of Item Model Types to Promote Assessment Engineering

Mark J. Gierl, Jiawen Zhou, Cecilia Alves

Editor: Michael Russell
russelmh@bc.edu
Technology and Assessment Study Collaborative
Lynch School of Education, Boston College
Chestnut Hill, MA 02467

Copy Editor: Jennifer Higgins
Design: Thomas Hoffmann
Layout: Aimee Levy

JTLA is a free on-line journal, published by the Technology and Assessment Study Collaborative, Caroline A. & Peter S. Lynch School of Education, Boston College.

Preferred citation:

Gierl, M.J., Zhou, J., Alves, C. (2008). Developing a Taxonomy of Item Model Types to Promote Assessment Engineering. *Journal of Technology, Learning, and Assessment,* 7(2). Retrieved [date] from http://www.jtla.org.

J·T·L·A

Abstract:

An item model serves as an explicit representation of the variables in an assessment task. An item model includes the *stem*, *options*, and *auxiliary information*. The *stem* is the part of an item which formulates context, content, and/or the question the examinee is required to answer. The *options* contain the alternative answers with one correct option and one or more incorrect options or distractors. The *auxiliary information* includes any additional material, in either the stem or option, required to generate an item, including texts, images, tables, and/or diagrams. In this study, we first present a taxonomy for item model development where variables in the stem are crossed with variables in the options to create a matrix of possible item model types. We then provide examples of each stem-by-option combination. Finally, we develop a software engine and apply the software to each item model type to generate multiple instances for each model.

J·T·L·A

Developing a Taxonomy of Item Model Types to Promote Assessment Engineering

Mark J. Gierl
Jiawen Zhou
Cecilia Alves
Centre for Research in Applied Measurement and Evaluation, University of Alberta

Introduction

Developments in cognitive science, mathematical statistics, computer technology, educational psychology, and computing science are creating opportunities for theoretical and practical changes in educational measurement, and related assessment fields. One consequence of these interdisciplinary influences is the emergence of a new area of research called *assessment engineering* (Luecht, 2006a, 2006b, 2007). Assessment engineering is an innovative approach to measurement where engineering-based principles are used to direct the design and development as well as the analysis, scoring, and reporting of assessment results. With this approach, the measurement specialist begins by defining the construct of interest using specific, empirically-derived cognitive models of task performance. Next, item models are created to produce replicable assessment tasks. Finally, psychometric models are applied to the examinee response data collected using the item models to produce scores that are both replicable and interpretable.

Assessment engineering differs from more traditional approaches to test development and analysis in four fundamental ways. First, cognitive models guide item development, rather than content blueprints. Hence, the assessment principles used in test construction are much more specific allowing items to be created quickly and efficiently during the development cycle. Second, explicit item models are created to control and manipulate both the content and difficulty of the items. Content experts use the item models during development thereby producing assessment tasks that adhere to strict quality controls and that meet high psychometric standards. Third, automated test assembly procedures are employed to build assessments that function to exacting specifications. As a result, multiple test forms can be created from a bank of items very efficiently according to both content and statistical specifications. Fourth, psychometric models are employed in a confirmatory – versus exploratory – manner to assess the model-data fit relative to the intended underlying structure of the constructs or traits the test is designed to measure. The outcomes from these model-data fit analyses also provide developers with guidelines for specific modifications to the cognitive and item models, as needed, to facilitate the acquisition of data that supports the intended assessment inferences (Luecht, Gierl, Tan, & Huff, 2006).

The purpose of this study is to *describe and illustrate* an approach for developing item models, which is the second stage in the assessment engineering framework. We introduce a *taxonomy of item model types* intended to help developers identify new models and methods for producing high-quality assessment items. We also illustrate how technology can be used with these models to generate large numbers of test items.

Introduction to Item Modeling: Terminology and Concepts

An item model[1] (LaDuca, Staples, Templeton, & Holzman, 1986; Bejar, 1996, 2002; Bejar, Lawless, Morley, Wagner, & Bennett, 2003) serves as an explicit representation of the variables in an assessment task, which includes the *stem*, the *options*, and oftentimes *auxiliary information*. The *stem* is the part of an item which formulates context, content, and/or the question the examinee is required to answer. The *options* contain the alternative answers with one correct option and one or more incorrect options or distractors. When dealing with a multiple-choice item model, both stem and options are required. With an open-ended or constructed-response item model, only the stem is created. *Auxiliary information* includes any additional material, in either the stem or option, required to generate an item, including texts, images, tables, and/or diagrams.

The stem and options can be divided further into *elements*. These elements are often denoted as strings, S, which are non-numeric values and integers, I, which are numeric values. This terminology is adopted from Bejar et al. (2003). By systematically manipulating these elements, measurement specialists can generate large numbers of instances or items for each model. If the instances are intended to measure content at similar difficulty levels, then the generated items are *isomorphic*[2]. When the goal of item generation is to create isomorphic instances, the measurement specialist manipulates the *incidental* elements, which are the surface features of an item that do not alter item difficulty. Conversely, if the instances are intended to measure content at different difficulty levels, then the generated items are *variants*. When the goal of item generation is to create variant instances, the measurement specialist can manipulate the incidental elements, but must manipulate one or more *radical* elements in the item model. The radicals are the deep features that alter item difficulty, and may also affect the psychometric properties of the test such as dimensionality.

To illustrate these concepts, two examples from Grade 6 mathematics are presented. In both examples, the item model is represented as the *stem* and *options* variables. The stem of the first example contains two integers (I1, I2) while the stem of the second example contains two additional strings (S1, S2). The I1 element includes Ann's payment. It ranges from \$1525 to \$1675 in increments of \$75. The I2 element includes the cost of the lawn, as either $\$30/m^2$ or $\$45/m^2$. As the first example represents only one lawn shape, shape is fixed and, therefore, no string variable is required (Figure 1a, next page). The four alternatives, labelled A to D, are generated in the example using algorithms produced from the integer values I1 and I2 as well as from keyed option A. In the second example shown in Figure 1b (page 8), the shape variable includes a square and a circle in the item stem. As a result, the S1 element describes the shape of the lawn and S2 must be constrained to match S1. Hence, the S2 element presents the appropriate area concept, side length or radius, required for calculating the area of the shape. Because the area calculation for a square differs from a circle, the options are also expected to include algorithms for computing the area of a circle (right-side of options box) in addition to the area of a square (left-side of options box). The area calculation difference between a square and a circle serves as the radical for the Figure 1b example. There is no auxiliary information for this item model (see Bejar et al., 2003, p. 9, for another example in mathematics).

Figure 1a: Item Model in Mathematics Used to Generate Isomorphic Instances

Ann has paid \$1525 for planting her lawn. The cost of lawn is \$45/m^2. Given the shape of her lawn is square, what is the side length of Ann's lawn?

A. 5.8

B. 6.8

C. 4.8

D. 7.3

Item Model Variables

Stem	Ann has paid \$ I1 for planting her lawn. The cost of lawn is \$ I2/m^2. Given the shape of her lawn is square, what is the side length of Ann's lawn?
Elements	I1 Value Range: 1525 – 1675 by 75 I2 Value Range: 30 or 40
Options	A. $= \sqrt{I1/I2}$ B. $= \sqrt{I1/I2} + 1$ C. $= \sqrt{I1/I2} - 1$ D. $= \sqrt{I1/I2} + 1.5$
Key	A

Figure 1b: Item Model in Mathematics Used to Generate Variant Instances

Item Model Variables

Stem	Ann has paid \$ I1 for planting her lawn. The cost of lawn is \$ I2/m². Given the shape of her lawn is S1, what is the S2 of Ann's lawn?	
Elements	I1 Value Range: 1525 – 1675 by 75 I2 Value Range: 30 or 40 S1 Range: "square" or "circular" S2 Range: "side length" or "radius" As S1 = "square", then S2 = "side length" As S1 = "circular", then S2 = "radius"	
Options	*As S1 = "square"* A. $= \sqrt{I1/I2}$ B. $= \sqrt{I1/I2} + 1$ C. $= \sqrt{I1/I2} - 1$ D. $= \sqrt{I1/I2} + 1.5$	*As S1 = "circular"* A. $= \sqrt{I1/I2 * 3.14}$ B. $= \sqrt{I1/I2 * 3.14} + 1$ C. $= \sqrt{I1/I2 * 3.14} - 1$ D. $= \sqrt{I1/I2 * 3.14} + 1.5$
Key	A	

Some Benefits of Item Modeling

Traditional item development using manual processes can be inefficient, largely because items are treated as isolated entities that are individually created, reviewed, and formatted. Because the items are individually authored, they yield unpredictable statistical outcomes (and, therefore, require field testing) because the incidental and radical elements are not easily identified or well understood. Traditional item development can also pose security risks for a testing program because the costs associated with construction, calibration, and maintenance limit the number of operational items that are available at any one time—with fewer operational items available, exposure risks may increase because more examinees are being exposed to each item. Drasgow, Luecht, and Bennett (2006, p. 473), in their seminal chapter in *Educational Measurement* (4th Edition) on technology and testing, provide this summary:

> The demand for large numbers of items is challenging to satisfy because the traditional approach to test development uses the item as the fundamental unit of currency. That is, each item is individually hand-crafted – written, reviewed, revised, edited, entered into a computer, and calibrated – as if no other like it had ever been created before. A second issue with traditional approaches is that it is notoriously hard to hit difficulty targets, which results in having too many items at some levels and not enough at other levels. Finally, the pretesting needed for calibration in adaptive testing programs entails significant cost and effort.

Item modeling can help overcome some of the limitations of the traditional approach thereby enhancing test development in two important ways. First, item modeling is cost-effective. The purpose of development is to create multiple models, where each model yields many items. Hence, banks can be created quickly which will minimize item exposure because larger pools of operational items are available for each test administration. The *logic* behind item modeling can also lead to more cost-effective practices because items are treated as classes which require a systematic and strategic development approach compared with treating each item as a single unit. Hence, the cost per item is lower because the unit of analysis is multiple instances per model rather than single instances per content specialist. Also, costly, yet common, errors in item development – including omissions or additions of words, phrases, or expressions as well as spelling, punctuation, capitalization, item structure, typeface, formatting, and language (e.g., English to French translation) problems – can be avoided because only specific elements in the stem and options are manipulated across large numbers of items. That is, the item model serves as a *template* where content specialists manipulate specific, well-defined, elements. The remaining components in the template, once finalized, are not altered during item development. As a result, item modeling should allow content specialists to quickly create large numbers of high-quality operational items that require few revisions during the development stage.

Second, item models provide the foundation necessary for *automatic item generation*. Automatic item generation is a procedure for using item models to create isomorphic instances with known item characteristics, often in real-time, as the examinee is writing the test. The procedure has two requirements: An item class must be described in enough detail to permit a computer to create instances of the class automatically. Also, the variables that affect item difficulty must be controlled across instances so the generated items do not require separate calibration (Drasgow et al., 2006). One key benefit of automatic item generation is that it minimizes, if not eliminates, the need for extensive field testing because the isomorphic instances generated from the parent model are *pre-calibrated* and, thus, do not need to be field tested. Automatic item generation can proceed from either strong or weak theory. If strong theory is used, calibrated items are generated automatically using the design principles articulated in a cognitive model (i.e., step 1 in assessment engineering framework; see also Leighton & Gierl, 2007). The cognitive model provides a detailed description of the variables that affect examinee performance which, in turn, can help pinpoint the item difficulty features. The obvious benefit of strong theory is that the cognitive features of item performance are identified and articulated in such detail that difficulty can be predicted and controlled. Unfortunately, few strong theories currently exist to guide our educational and psychological measurements. As a result, strong theory for automatic item generation has been limited to specific tasks in domains such as mental rotation (Bejar, 1990) and spatial ability (Embretson & Yang, 2007).

In the absence of strong theory, weak theory must be used. Weak theory yields calibrated items generated automatically using design guidelines (rather than design principles) discerned from a combination of experience, theory, and research (rather than cognitive models) (Drasgow et al., 2006). Initially, the guidelines are used to identify a *parent item model*. Then, incidental item features in the parent item model are manipulated to produce isomorphic instances. The benefit of weak theory for automatic item generation stems from its practicality. Parent models can often be identified by reviewing items from previously administered exams. Weak theory is also well-suited to broad content domains where few theoretical descriptions exist about the cognitive knowledge and skills used by examinees to solve items. The main drawback of weak theory is that item difficulty is neither predictable nor easily controlled. However, if data for a parent item model are available, statistical procedures have been developed to account for the variation among the isomorphic instances and estimate their item difficulty levels (e.g., Glas & van der Linden, 2003; Mislevy, Wingersky, & Sheehan, 1994; Sinharay, Johnson, & Williamson, 2003; Sinharay & Johnson, 2005). In short, item modeling can enhance test development practices and provide the necessary foundation for sophisticated psychometric procedures such as automatic item generation.

To create item models systematically and strategically using either strong or weak theory, a generic *item model taxonomy* is required. This type of taxonomy, in fact, is a prerequisite for a functional automatic item generation system because it provides the guiding principles necessary for designing a large number of diverse item models by outlining their structure, function, similarities, differences, and limitations. Unfortunately, the educational and psychological measurement literature contains little discussion on how to develop item models and few examples exist. Therefore, our first step is to create a taxonomy of item model types that could generalize across content areas, as a way of offering test development principles for creating item models. We also provide examples of these item model types.

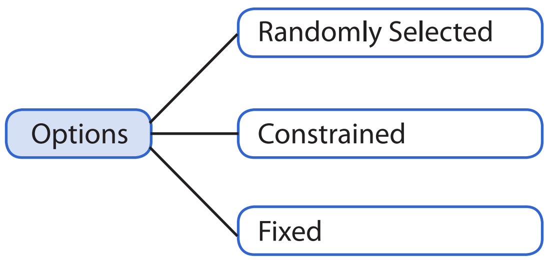
A Taxonomy for Designing Item Models

A taxonomy for item model development requires at least three variables: the *stem*, *options*, and *auxiliary information*. Each variable functions differently. The *stem* is the section of the model used to formulate context, content, and/or questions. It contains four categories, as shown in Figure 2. *Independent* indicates that the n_i element(s) ($n_i \geq 1$) in the stem are independent or unrelated to one another. That is, a change in one element will have no affect on the other stem elements in the item model. *Dependent* indicates n_d element(s) ($n_d \geq 2$) in the stem are dependent or directly related to one other. *Mixed Independent/Dependent* include both independent ($n_i \geq 1$) and dependent ($n_d \geq 1$) elements in the stem. *Fixed* represents a constant stem format with no variation or change.

Figure 2: Categories in the Item Model Stem

The *options*, shown in Figure 3, contain the alternatives for the item model when the multiple-choice format is used. The options contain three categories. *Randomly-selected* options refers to the manner in which the distractors are selected from their corresponding content pools. The distractors are selected randomly. *Constrained* options mean that the keyed option and the distractors are generated according to specific constraints, such as formulas, calculation, and/or context. *Fixed* options occur when both the keyed option and distractors are invariant or unchanged in the item model.

Figure 3: Categories in the Item Model Options



By crossing the stem and options categories, a matrix of item model types can be produced. The *stem-by-options* matrix is presented in Table 1. Ten functional combinations are designated with a checkmark, "✓". The two remaining combinations are labelled not applicable, "NA", because a model with a fixed stem and constrained options is an infeasible item type and a model with a fixed stem and options produces a single multiple-choice item type (i.e., a traditional multiple-choice item).

Table 1: Plausible Stem-by-Option Combinations in the Item Model Taxonomy

	Stem			
Options	**Independent**	**Dependent**	**Mixed**	**Fixed**
Randomly Selected	✓	✓	✓	✓
Constrained	✓	✓	✓	N/A
Fixed	✓	✓	✓	N/A

Next, the ten *stem-by-options* combinations are illustrated. We draw on examples, first, from mathematics to demonstrate the applicability of our taxonomy. For each stem-by-option combination in Table 1, we present an item, followed by the item model template which outlines the stem, elements, options, auxiliary information, and key (cf. Bejar et al., 2003,

p. 9). These 10 models will then be used with the item generator software described in the next section to create new items. We also provide a second set of examples in the Appendix. These examples were drawn from diverse content areas, including science, social studies, language arts, and architecture.

Model #1: Stem: *Independent*; Options: *Randomly Selected*; Auxiliary Information: *None*

The students could see a circular lake from the top of a Tramway.
The distance around the lake is known as its

A. circumference

B. diameter

C. radius

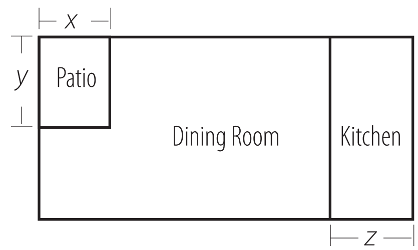
D. area

Item Model Variables

Stem	S1 could see a circular S2 from the top of a S3. The distance around the S2 is known as its
Elements	S1 Range: "Some students", "Bob and Mike", "Anne and her sister", "Some boys", "Some girls" S2 Range: "lake", "pool" S3 Range: "Tramway", "mountain", "building", "tower"
Options	Key: circumference or perimeter Distractors: diameter, radius, area, sector, chord, arc
Auxiliary Information	None
Key	A

Model #2: Stem: *Independent;* Options: *Constrained;* Auxiliary Information: *Diagram*

This is a diagram of the Pizza Place's floor plan. Fire regulations state that each customer in a dining room must have a minimum of 2.2 m² of floor space. What is the maximum number of customers that can be seated in the Pizza Place's dining room when $x = 3.0$, $y = 5.0$, $z = 4.0$, and the restaurant is 10 × 16?



A. 47

B. 59

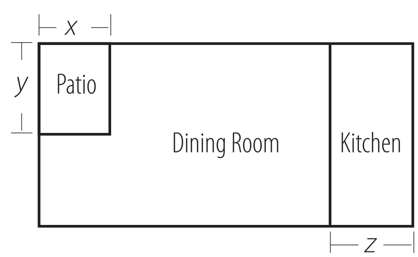
C. 54

D. 72

Item Model Variables

Stem

This is a diagram of the Pizza Place's floor plan. Fire regulations state that each customer in a dining room must have a minimum of 2.2 m² of floor space. What is the maximum number of customers that can be seated in the Pizza Place's dining room when x = I1 , y = I2 , z = I1 + 1, and the restaurant is 2 * I2 by I1 + 2 * I2 + 3?



Elements

I1 Value Range: 1–3 by 1

I2 Value Range: 10–18 by 1

Options

A. Round Down (4 * I2 * I2 + 4 * I2 – I1 * I2) / 2.2

B. Round Down (2 * I2 * (I1 + 2 * I2 + 3) – I1 * I2) / 2.2

C. Round Down (4 * I2 * I2 + 4 * I2) / 2.2

D. Round Down (2 * I1 * I2 + 4 * I2 * I2 + 6 * I2) / 2.2

Auxiliary Information

Diagram of a Pizza Place's floor plan

Key

A

Model #3: Stem: *Independent*; Options: *Fixed*; Auxiliary Information: *None*

In order to make a particular shade of green paint, Mary uses 24 parts of blue pigment, 12 parts of white, and 6 parts of yellow. What is the simplest ratio of these pigments?

A. 4 : 2 : 1

B. 6 : 3 : 2

C. 12 : 6 : 3

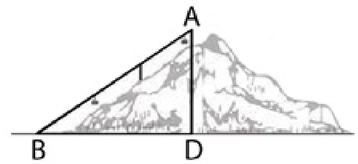
D. 1 : 1/2 : 1/4

Item Model Variables

Stem	In order to make a particular shade of S1 paint, Mary uses 4* I1 parts of S2 pigment, 2* I1 parts of white, and I1 parts of S3. What is the simplest ratio of these pigments?
Elements	S1 Range: "green", "orange", "purple", "brown" S2 Range: "blue", "red" S3 Range: "yellow", "blue", "black" I1 Value Range: 2, 3, 6, or 12 As S1="green", S2="blue", S3="yellow"; As S1="orange", S2="red", S3="yellow"; As S1="purple", S2="red", S3="blue"; As S1="brown", S2="red", S3="black"
Options	A. 4 : 2 : 1 B. 6 : 3 : 2 C. 12 : 6 : 3 D. 1 : 1/2 : 1/4
Auxiliary Information	None
Key	A

Model #4: Stem: *Dependent*; Options: *Randomly Selected*; Auxiliary Information: *Pictures*

Some students visited the sights around Jasper. Omar observed that the Jasper Tramway formed an angle with the surroundings.



What kind of angle is ∠ABD?

A. Acute

B. Right

C. Obtuse

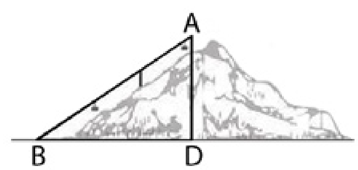
D. Straight

Item Model Variables

Stem

Some students visited S1. Omar observed that the S2 formed an angle with the surroundings.

What kind of angle is ∠ABD?



Elements

S1 Range: "the sights around Jasper", "a pizzeria", "a park", "a watch store"

S2 Range: "Jasper Tramway", "pizza slice", "park's table", "a clock"

As S1="the sights around Jasper", S2="Jasper Tramway"

As S1="a pizzeria", then S2="pizza slice"

AS S1="a park", then S2="park's table"

AS S1="a watch store", then S2="clock's pointer"

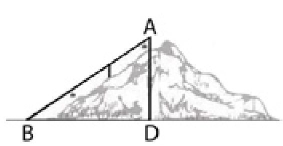
Options

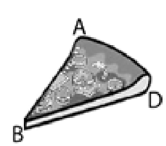
Key: Acute

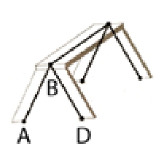
Distractors: Right, Obtuse, Straight, Vertical

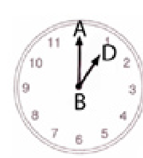
Auxiliary Information

Picture of Jasper Tramway; Picture of pizza slice; Picture of park's table; Picture of a clock







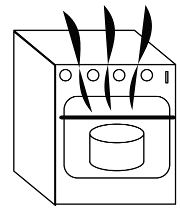


Key

A

Model #5: Stem: *Dependent*; Options: *Constrained*; Auxiliary Information: *Picture*

The thermostat of an oven malfunctioned. First, the temperature dropped 5°C, then it increased 7°C, fell 12°C, and finally decreased a further 30°C before it stabilized 185°C. What was the original temperature?



A. 239°C

B. 225°C

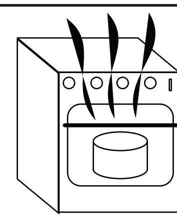
C. 131°C

D. 145°C

Item Model Variables

Stem

The thermostat of an oven malfunctioned. First, the temperature dropped I1 ° S1, then it increased I2 ° S1, fell I3 ° S1, and finally decreased a further I4 ° S1 before it stabilized I5 ° S1. What was the original temperature?



Elements

As S1 = "°C"	*As S1 = "°F"*
I1 Value range: 3 to 18 by 3	I1 Value range: 15 to 30 by 3
I2 Value range: 2 to 20 by 2	I2 Value range: 10 to 30 by 2
I3 Value range: 5 to 15 by 1	I3 Value range: 21 to 30 by 1
I4 Value range: 10 to 40 by 4	I4 Value range: 50 to 60 by 5
I5 Value range: 100 to 200 by 5	I5 Value range: 200 to 300 by 5

Options

A. I1 + I2 + I3 + I4 + I5

B. I1 – I2 + I3 + I4 + I5

C. I1 + I2 – I3 – I4 + I5

D. –I1 + I2 – I3 – I4 + I5

Auxiliary Information

Oven picture

Key

B

Model #6: Stem: *Dependent*; Options: *Fixed*; Auxiliary Information: *None*

To calculate the speed of a motorcycle, if d = distance a motorcycle travels in metres, r = speed in m/s, and t = time in seconds, the formula $d = rt$ would have to be rewritten as

A. $r = d - t$

B. $r = dt$

C. $r = d/t$

D. $r = t/d$

Item Model Variables

Stem	To calculate the speed of a S1, if d = distance a S1 travels in S2, r = speed in S3 / S4, and t = time in S5, the formula $d = rt$ would have to be rewritten as
Elements	S1 Range: "motorcycle", "bike", "car", "truck" S2 Range: "metres", "kilometres" S3 Range: "m", "km" S4 Range: "s", "h" S5 Range: "seconds", "hours" As S1 = "bike", then S2 = "metres" As S1 = "motorcycle", "car", "truck", then S2 = "kilometres" As S2 = "metres", then S3 = "m", S4 = "s", S5 = "seconds" As S2 = "kilometres", then S3 = "km", S4 = "h", S5 = "hours"
Options	A. $r = d - t$ B. $r = dt$ C. $r = d/t$ D. $r = t/d$
Auxiliary Information	None
Key	C

Model #7: Stem: *Mixed*; Options: *Randomly Selected*; Auxiliary Information: *None*

Four of the students had a foot race at their campsite near Jasper. John finished 5 s behind Ryan, Sheila finished 3 s behind John, Danielle was 6 s in front of Sheila.

In what order, from first to last, did the students finish?

A. Ryan, Danielle, Sheila, John

B. Ryan, John, Danielle, Sheila

C. Ryan, Sheila, John, Danielle

D. Ryan, Danielle, John, Sheila

Item Model Variables

Stem

Four S1 had a S2 at their S3. John finished I1 S4 behind Ryan, Sheila finished I2 S4 behind John, Danielle was I3 S4 in front of Sheila.

In what order, from first to last, did the S1 finish?

Elements

S1 Range: "students", "kids", "children"

S2 Range: "foot race", "bike race", "competition", "raffle basket competition", "Miniature Golf Tournament", "balloon race", "Candy Bar Bingo"

S3 Range: "school", "campsite near Jasper", "community league"

S4 Range: "s", "points""

I1 3 to 6 by 1

I2 2 to 5 by 1

I3 I2+2

As S2= "foot race", "bike race", or "balloon race", then S4="s"

As S3= "raffle basket competition", "Miniature Golf Tournament", or "Candy Bar Bingo", then S4="points"

Options	Key: Ryan, Danielle, John, Sheila Distractors: Danielle, Ryan, Sheila, John Danielle, John, Ryan, Sheila Danielle, John, Sheila, Ryan Danielle, Sheila, Ryan, John Danielle, Sheila, John, Ryan Danielle, Ryan, John, Sheila Ryan, Danielle, Sheila, John Ryan, John, Danielle, Sheila Ryan, John, Sheila, Danielle Ryan, Sheila, Danielle, John Ryan, Sheila, John, Danielle John, Ryan, Sheila, Danielle John, Ryan, Danielle, Sheila John, Danielle, Ryan, Sheila John, Danielle, Sheila, Ryan John, Sheila, Danielle, Ryan John, Sheila, Ryan, Danielle Sheila, Danielle, Ryan, John Sheila, Danielle, John, Ryan Sheila, Ryan, Danielle, John Sheila, Ryan, John, Danielle Sheila, John, Danielle, Ryan Sheila, John, Ryan, Danielle
Auxiliary Information	None
Key	D

Model #8: Stem: *Mixed*; Options: *Constrained*; Auxiliary Information: *Table*

Mrs. Kary kept a record of participants in school activities and the total points some teams accumulated.

Team	Participants	Total Points
Red	26	762
Green	33	978
Yellow	22	641
Blue	29	?

Based on the information in the table, the Blue team's total number of points would most likely be

A. 692

B. 768

C. 809

D. 851

Item Model Variables

Stem

S2 kept a record of participants in school activities and the total points some teams accumulated.

Team	Participants	Total Points
S1_1	I1	Round I1 * I2
S1_2	I1 + 7	Round (I1 + 7) * I3
S1_3	I1 – 4	Round (I1 – 4) * I4
S1_4	I1 + 3	?

Based on the information in the table, the S1_4 team's total number of points would most likely be

Elements

S1 Range: "Blue", "Green", "Yellow", "Red", "Gray", "Brown", "Black", "White"
S2 Range: "Mr. Kary", "Mr. Rogers", "Mr. Pitt"
I1 Value Range: 20 to 29 by 1
I2 Value Range: 29.00 to 29.99 by 0.3
I3 Value Range: 29.00 to 29.99 by 0.3
I4 Value Range: 29.00 to 29.99 by 0.3
I5 Value Range: 29.00 to 29.99 by 0.3

Options	A. Round ((I1 –4) * I4 +7) B. Round ((I1 – 4) * I4 + 29) C. Round ((I1 * I2 + (I1 + 7) * I3 + (I1 – 4) * I4) / 3) D. Round ((I1 + 3) * I5)
Auxiliary Information	Table
Key	D

Model #9: Stem: *Mixed*; Options: *Fixed*; Auxiliary Information: *None*

There are three radio stations in a town with about 60,000 potential listeners. Peter was hired to determine the approximate number of listeners each station had. He decided to survey 200 people.

These 200 people are called a

A. sample

B. population

C. frequency

D. census

Item Model Variables

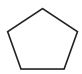
Stem	There are I1 S1 in a town with about I2 potential S2. S3 was hired to determine the approximate number of S2 each one had. He decided to survey I3 people. These I3 people are called a
Elements	S1 Range: "radio stations", "TV stations", "sports teams" S2 Range: "listeners", "watchers", "fans" S3 Range: "school", "campsite near Jasper", "community league" I1 Value range: 2 to 5 by 1 I2 Value range: 20,000 to 80,000 by 10,000 I3 Value range: 200 to 600 by 50 As S1= "radio stations", then S2= "listeners" As S1= "TV stations", then S2= "watchers" As S3= "sports teams", then S2= "fans"
Options	A. sample B. population C. frequency D. census
Auxiliary Information	None
Key	A

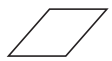
Model #10: Stem: *Fixed*; Options: *Randomly Selected*; Auxiliary Information: *None*

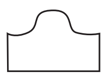
There is a team crest on Henry's jacket that:

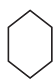
- is a polygon
- has more sides than a triangle but fewer than a hexagon
- is not a quadrilateral

What is the shape of the crest?

A. 

B. 

C. 

D. 

Item Model Variables

Stem

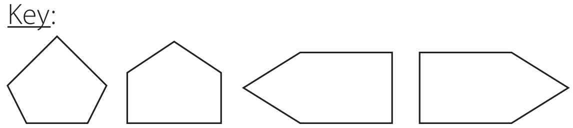
There is a team crest on Henry's jacket that:

- is a polygon
- has more sides than a triangle but fewer than a hexagon
- is not a quadrilateral

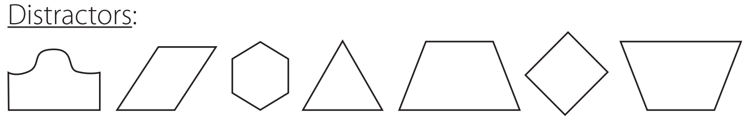
What is the shape of the crest?

Options

Key:



Distractors:



Auxiliary Information

None

Key

A

Generating Items from Models

With 10 functional item models in mathematics, the second step was to develop a software engine that automatically creates and banks items for each model. This type of software serves as a proof-of-concept to demonstrate the practicality and feasibility of our item generation approach[3]. The software is called IGOR (**I**tem **G**enerat**OR**). It was written in Sun Microsystems JAVA SE 6.0. The purpose of IGOR is to generate items for each model. A short description of the item generation software is presented next. This description can also be found in the 'Help' menu of the IGOR software.

Getting Started

This editor is used to generate item models that can be used to dynamically create item banks for use in test construction. An item stem can be defined, along with an optional set of variables, constraints, and response choices to generate a wide range of item models.

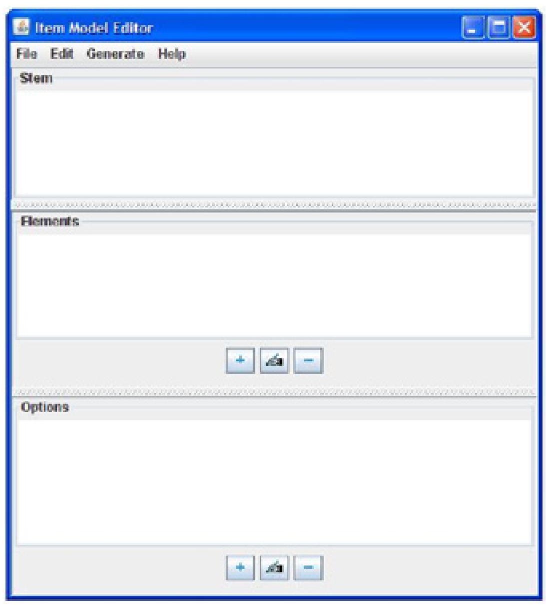
The User Interface

There are three components to the main screen of the editor window:

The ***stem*** is the base from which each derived test item is created. Any text or formula entered here will appear as the stem in each test item.

The ***elements*** are the variables and constraints used to generate each individual test item. Variables can consist of either a number range or a series of text values.

The ***options*** are the possible answers to a given test item. They are classified as either key or distractor.



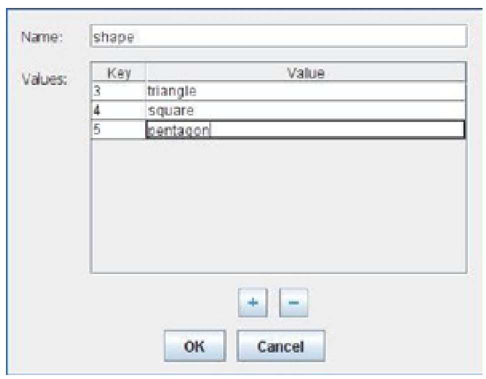
The Elements and Options panels each contain three buttons. The first of these adds a new element or option to its panel. The second opens a window to edit the currently selected element or option. The third removes the currently selected element or option from the model.

Elements

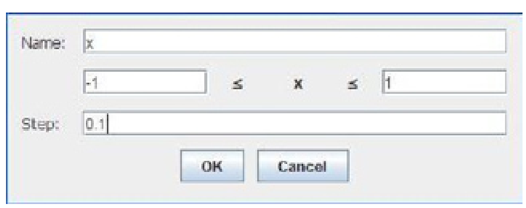
Elements in item models are used to create variety between different test items. This is done using variables, the values of which change from item to item, and constraints, which restrict the ways in which variables may be combined.

Variables and Constraints

There are two types of variables in the item model editor. The first is the text variable. A text variable may have any number of possible lines of text, each of which must be matched with a corresponding numeric key. A key for a given value may be any desired integer, but no two values can have the same key.



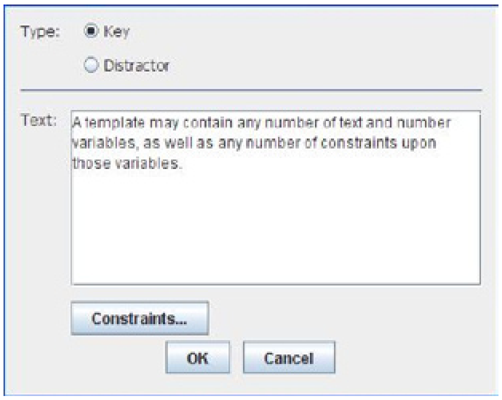
The second type of variable is the numeric variable. A numeric variable may take on any value between some minimum and maximum value, and increases by some fixed step size.

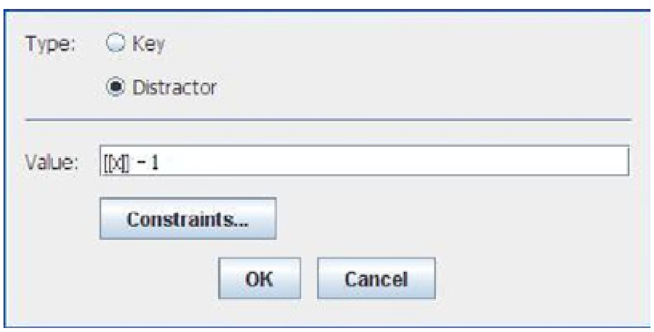


Variables in an item model may have constraints placed on them, restricting the values they may take within a given test item. Constraints may be useful for eliminating some combinations of variables that yield nonsense items or items that are too similar to other items in a bank.

Options

Options represent the different possible choices available in any given test item. They may be classified as either keys, representing correct responses, or distractors, representing incorrect responses, and may have either text or numbers as their values. Options may have restrictions placed on them so they change when a variable has different values.



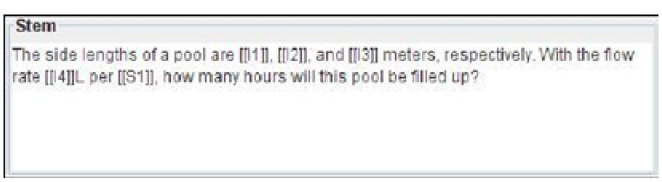


Creating a Simple Item Model

Creating an item model requires the following three steps:

Define a Stem

Define a stem for the item model. The model can contain variables, defined in the Elements panel, equations, and/or LaTeX formulas.



Create Elements

Create elements such as variables that change from item to item and constraints that restrict the values of these variables. A multiple-choice item model requires a stem, a key, and one or more distractors.

Elements	
Variable (numeric)	I1
Range	25.0 to 30.0
Step size	5.0
Variable (numeric)	I2
Range	10.0 to 15.0
Step size	5.0
Variable (numeric)	I3
Range	3.0 to 5.0
Step size	2.0
Variable (numeric)	I4
Range	300.0 to 50000.0
Step size	49700.0
Variable (text)	S1
Value 1	minute
Value 2	hour
Constraint	([[S1]] == 1 AND [[I4]] == 300) OR ([[S1]] ...

Create Options

Create options for the item model. Options are divided into two groups. The first group, key, consists of the correct response for the test item. The second, distractors, consists of the incorrect responses for the test item. A multiple-choice item model requires at least one key and one distractor.

Options	
Key (numeric)	[[I1]] * [[I2]] * [[I3]] * 1000 / ([[I4]] * 60)
Constraint	S1 = 1
Key (numeric)	[[I1]] * [[I2]] * [[I3]] * 1000 / ([[I4]] * 60)
Constraint	S1 = 2
Distractor (numeric)	[[I1]] * [[I2]] * [[I3]] * 1000 / [[I4]]
Constraint	S1 = 1
Distractor (numeric)	[[I1]] * [[I2]] * [[I3]] / [[I4]]
Constraint	S1 = 1
Distractor (numeric)	[[I1]] * [[I2]] * [[I3]] / [[I4]] + 10
Constraint	S1 = 1
Distractor (numeric)	[[I1]] * [[I2]] * [[I3]] * 1000 / [[I4]] - 5
Constraint	S1 = 2
Distractor (numeric)	[[I1]] * [[I2]] * [[I3]] / [[I4]] + 5
Constraint	S1 = 2
Distractor (numeric)	[[I1]] * [[I2]] * [[I3]] * 1000 / [[I4]] + 10
Constraint	S1 = 2

Options

Stems, text variables, and text options may contain equations and formulas. They may be evaluated as numerical or logical values or used to generate an image of the equation in the test bank output.

$$\int_{t=0}^{2\pi} \frac{\sqrt{t}}{1+\cos^2 t}\, dt$$

Equations are marked with /(to indicate the start of the formula and /) to indicate the end of the formula.

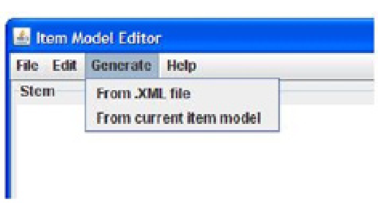
LaTeX Formulas

LaTeX formulas may be used to display mathematical content. Such formulas are marked with the characters /(to indicate the start of a formula and / to indicate the end of a formula). Supported LaTeX tags include, but are not limited to:

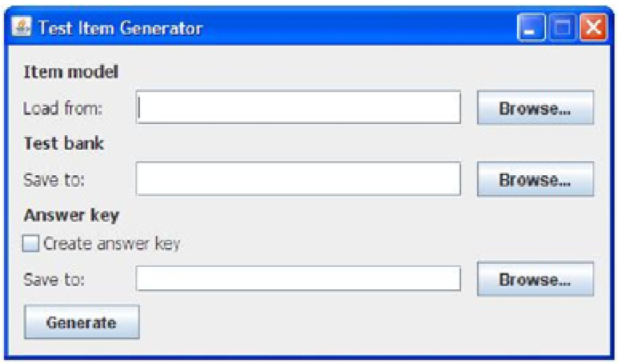
\frac{numerator}{denominator}	$\frac{1}{2}$
\sqrt[root]{argument}	$\sqrt[3]{9}$
\cdots, \ddots, \ldots	$\cdots \ddots \ldots$
Greek letters \alpha to \omega	$\alpha\beta\ldots\omega$
\leftarrow, \Leftarrow, \uparrow, \leftrightarrow ...	$\leftarrow\Leftarrow\uparrow\leftrightarrow$
\cap, \cup	$\cap\cup$

Our program uses JMathTex for LaTeX formulas and supports all LaTeX features in the JMathTex library.

Once an item model is created, a test bank can be generated and saved to an HTML file. This file is produced using the Generate menu in the editor.



To generate items from a model, the user will be presented with the dialogue box shown below. In this box, the user must specify the item model file, the item bank output file, and the answer key file. If the option 'Create answer key' is not selected, then the resulting test bank will always display the correct answer as the last option (or alternative). If the option 'Create answer key' is selected, then the resulting test bank will randomly order the options. Once the files have been specified in the dialogue box, the program can be executed by selecting the 'Generate' button.



When IGOR was applied to the 10 mathematics item models, 331,371 items were created. Generation capacity is dependent on several factors including the model, the number of elements in the stem of the model, and the range specified for the elements. For this demonstration, item generation ranged from a low of 8 items for model 6 to a high of 202,860 items for model 5. A summary of each model and their generation capacity is presented in Table 2.

Table 2: Mathematics Items Generated from 10 Item Model Types

Mathematics Item Model	Number of Items Generated
1 (Independent; Randomly Selected)	1,280
2 (Independent; Constrained)	27
3 (Independent; Fixed)	16
4 (Dependent; Randomly Selected)	16
5 (Dependent; Constrained)	202,860
6 (Dependent; Fixed)	8
7 (Mixed; Randomly Selected)	364
8 (Mixed; Constrained)	122,880
9 (Mixed; Fixed)	3,780
10 (Fixed; Randomly Selected)	140
Total	**331,371**

Summary

An item model (LaDuca et al., 1986; Bejar, 1996, 2002; Bejar et al., 2003) serves as an explicit representation of the variables in an assessment task. For a multiple-choice item, the developer can manipulate information in the *stem*, *options*, and *auxiliary information*. In the current study, we introduced an item model taxonomy and applied the taxonomy to different content areas. We also developed and applied a software program called IGOR to the math item models to generate instances for each model. Singley and Bennett (2002, p. 366) claimed that an item model serves as "*a schematized description of a class of questions from which draft test items are automatically generated for the test developer to review and revise*." In other words, item modeling can supplement, rather than replace, traditional test development procedures by offering a systematic and strategic approach for manipulating content so items can be produced automatically. This type of integrated approach to test construction, where content expertise and technology are efficiently combined to direct the design, development, and production of test items, is one example of how assessment engineering can alter educational and psychological testing. The role of the content specialist is critical for the *creative task* of designing and developing meaningful models. At the same time, the role of technology is critical for the *algorithmic task* of combining large number of elements in each model to produce items which, in turn, are stored in banks. By combining content expertise and technology, item modeling could be used to generate content for an entire test. Item modeling can also be used to supplement existing test content by promoting generative processes for some item types which, when combined with items created using the more traditional approach, could produce the content for the final test form. A testing program may design two item models from our taxonomy to generate some operational items. But, through development and experience, item modeling and the supporting generative processes may expand to account for larger percentages of test content over time.

Theory and Item Modeling

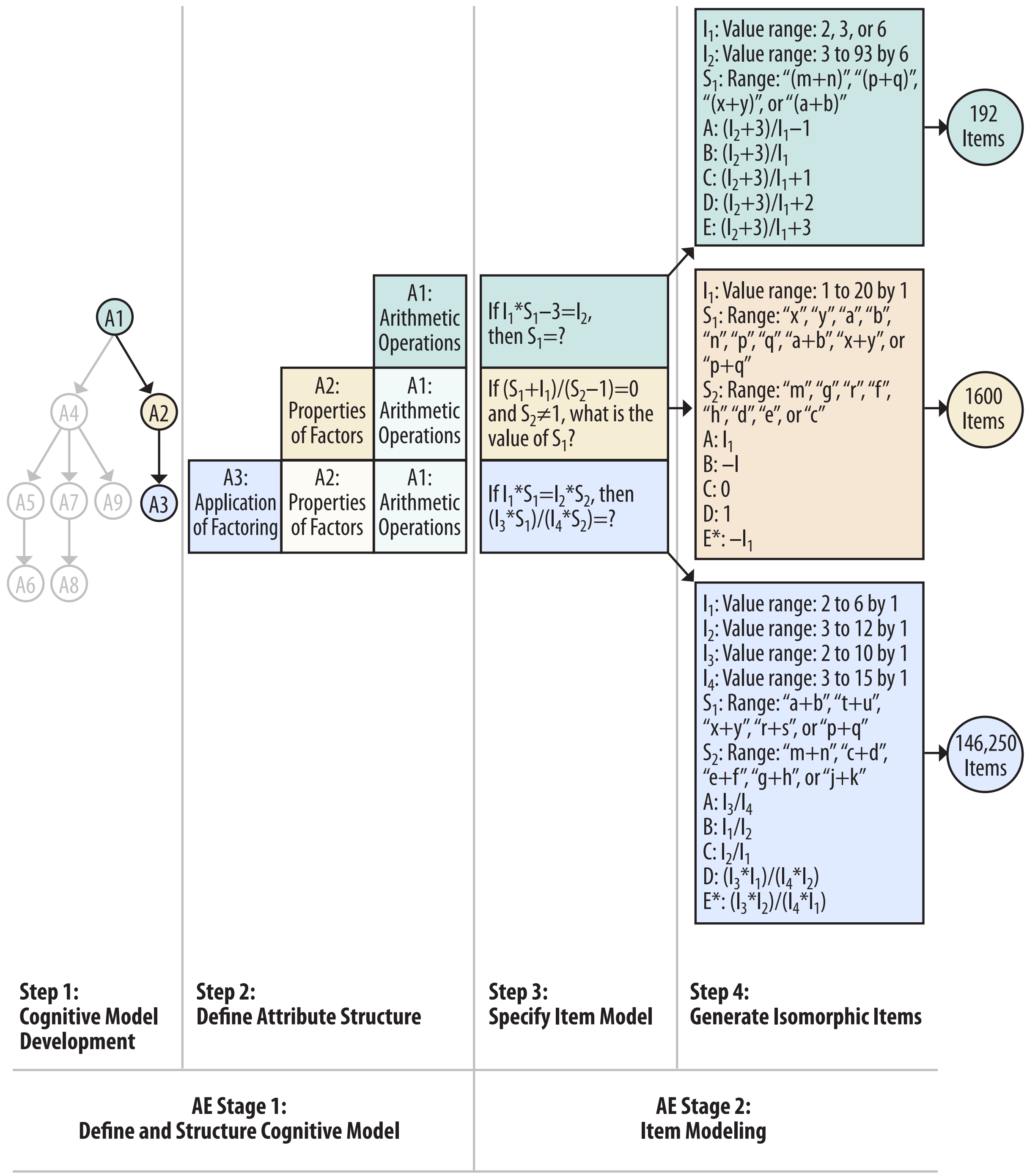
The approach to item modeling, as presented in this paper, is very practical. We omitted the cognitive model step, identified first in the assessment engineering framework, because our focus was on *item model design*. Our sample models were developed from weak theory by reviewing large numbers of items from previous test administrations to illustrate how the four stem and three option categories could be combined to produce diverse yet functional models in a practical testing context. The practicality of weak theory can also be used to generate items automatically, as we noted earlier, because design guidelines can produce parent models that, when combined with specialized statistical procedures, yield item parameter estimates for the isomorphic instances.

Despite the practicality of weak theory, strong theory is still the preferred approach. With strong theory, calibrated items are generated automatically using the design principles articulated in a cognitive model (Drasgow et al., 2006). A cognitive model in educational measurement refers to a "simplified description of human problem solving on standardized educational tasks, which helps to characterize the knowledge and skills students at different levels of learning have acquired and to facilitate the explanation and prediction of students' performance" (Leighton & Gierl, 2007, p. 6). The cognitive model provides a detailed description of the variables that affect examinee performance which, in turn, can help pinpoint the item difficulty elements. Although few comprehensive and practical cognitive models exist for test development, educational and psychological researchers are beginning to identify and articulate these models as well as to create psychometric methods for evaluating their statistical properties. For instance, Leighton, Gierl, and Hunka (2004) introduced a procedure called the *attribute hierarchy method* (AHM), which is used to classify examinees' test item responses into a set of structured attribute patterns associated with different components from a cognitive model of task performance. Attributes include different procedures, skills, and/or processes that an examinee must possess to solve a test item. The attributes are structured using a hierarchy so the ordering of the cognitive skills is specified. As a result, the attribute hierarchy serves as an explicit cognitive model. This model, in turn, provides a framework for designing item models and for linking examinees' test performance to specific inferences about psychological skill acquisition.

The AHM has been used to develop and analyze a cognitive diagnostic assessment in high school algebra (e.g., Gierl, Wang, & Zhou, 2008). Cognitive diagnostic assessment is a form of testing that employs a cognitive model to first develop items that measure specific knowledge and skills and then uses this model to direct the psychometric analyses of the examinees' item response patterns to promote specific diagnostic inferences. The AHM can also be used for computer adaptive testing (Gierl & Zhou, 2008). Computer adaptive testing is an innovative form of assessment that matches the difficulty of a test item to the ability estimate for an examinee. The matching is accomplished by first presenting an examinee with an item of average difficulty and then, depending on the examinee's response, an item of greater or lesser difficulty is presented until the algorithms controlling item administration meet a specified stopping criterion.

Both cognitive diagnostic assessment and computer adaptive testing make heavy demands on the items in the bank. One benefit of using a cognitive model-based procedure, like the AHM, lies in its facility to guide item development. The AHM requires the developer to specify a *reduced incidence matrix* of order (*k*, *i*) where *k* is the number of attributes and *i* is the number of items specified in the hierarchy. The reduced incidence matrix serves as the cognitive test specifications because it identifies all possible attribute-by-item combinations in the cognitive model. Item models can therefore be developed to measure each attribute-by-item combination (see Gierl, Wang, et al., 2008, p. 34). For example, three items are required to measure three algebra attributes, A1 to A3, in Figure 3 (next page) (the cognitive model in Figure 3 contains 9 attributes but, for illustration purposes, only the first three attributes are presented). Item 1 measures attribute A1, which includes basic arithmetic operation skills; item 2 measures attribute A2, which includes knowledge about the properties of factors in addition to basic arithmetic operation skills (i.e., A1); item 3 measures attribute A3, which includes the application of factoring in addition to the properties of factors (i.e., A2) and basic arithmetic operation skills (i.e., A1) (this algebra model is described in more detail in Gierl, Wang, et al., 2008).

Figure 3: Stages 1 and 2 in the Assessment Engineering Framework



In addition to creating each attribute-specific item, multiple instances of these items must also be created to produce a large functional bank for continuous diagnostic and adaptive testing. Research on item modeling, as described in the current paper, can support this process where a parent item model is developed for each attribute in the cognitive model and isomorphic instances are generated. Hence, the parent model yields large numbers of items for each attribute in the cognitive model that will function similarly within or between test forms. In the Figure 3 example, the cognitive model prescribes the hierarchy of skills in step #1. The cognitive model is also used to specify the attribute structure for item development in step #2. Taken together, steps #1 and #2 constitute stage 1 in an assessment engineering framework. In step #3, a parent item model is created to measure each attribute according to the specific hierarchical ordering of increasing cognitive complexity. The appropriate elements in the stem and options are then manipulated in step #4 to generate isomorphic instances for each item model. These last two steps, together, serve as stage 2 in assessment engineering. In sum, both the cognitive and item modeling stages are required to produce an efficient *item generation system* necessary for the development and production of large numbers of items – these items, in turn, provide the foundation for a modern 21st century assessment program where continuous diagnostic and adaptive testing are not only possible, but potentially viable.

Endnotes

1. Item models have been characterized in different ways. For example, they have been described as schemas (Singley & Bennett, 2002), blueprints (Embretson, 2002), templates (Mislevy & Riconscente, 2006), forms (Hively, Patterson, & Page, 1968), and shells (Haladyna & Shindoll, 1989).
2. For Glas and van der Linden (2003) isomorphic instances are called item *clones*.
3. It is common to pair conceptual and technological advances in the item generation literature. For example, Singley and Bennett (2002) developed an item generation program called *Mathematics Test Creation Assistant* to illustrate how schema theory could be used to automatically produce math items.

References

Bejar, I.I. (1990). A generative analysis of a three-dimensional spatial task. *Applied Psychological Measurement, 14*, 237–245.

Bejar, I.I. (1996). *Generative response modeling: Leveraging the computer as a test delivery medium* (ETS Research Report 96–13). Princeton, NJ: Educational Testing Service.

Bejar, I.I. (2002). Generative testing: From conception to implementation. In S. H. Irvine & P. C. Kyllonen (Eds.), *Item generation for test development* (pp.199–217). Hillsdale, NJ: Erlbaum.

Bejar, I.I., Lawless, R., Morley, M.E., Wagner, M.E., Bennett, & R.E., Revuelta, J. (2003). A feasibility study of on-the-fly item generation in adaptive testing. *Journal of Technology, Learning, and Assessment, 2*(3). Available from http://www.jtla.org.

Drasgow, F., Luecht, R.M., & Bennett, R. (2006). Technology and testing. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 471–516). Washington, DC: American Council on Education.

Embretson, S.E. (2002). Generating abstract reasoning items with cognitive theory. In S. H. Irvine & P. C. Kyllonen (Eds.), *Item generation for test development* (pp. 219–250). Mahwah, NJ: Erlbaum.

Embretson, S.E., & Yang, X. (2007). Automatic item generation and cognitive psychology. . In C. R. Rao & S. Sinharay (Eds.) *Handbook of Statistics: Psychometrics, Volume 26* (pp. 747–768). North Holland, UK: Elsevier.

Gierl, M.J., & Zhou, J. (2008). Computer adaptive-attribute testing: A new approach to cognitive diagnostic assessment. *Zeitschift für Psychologie – Journal of Psychology, 216*, 29-39.

Gierl, M.J., Wang, C., & Zhou, J. (2008). Using the attribute hierarchy method to make diagnostic inferences about examinees' cognitive skills in algebra on the SAT©. *Journal of Technology, Learning, and Assessment, 6*(6). Available from http://www.jtla.org.

Glas, C.A.W., & van der Linden, W.J. (2003). Computerized adaptive testing with item cloning. *Applied Psychological Measurement, 27*, 247–261.

Haladyna, T., & Shindoll, R. (1989). Items shells: A method for writing effective multiple-choice test items. *Evaluation and the Health Professions, 12*, 97–106.

Hively, W., Patterson, H.L., & Page, S.H. (1968). A "universe-defined" system of arithmetic achievement tests. *Journal of Educational Measurement, 5*, 275–290.

LaDuca, A., Staples, W.I., Templeton, B., & Holzman, G.B. (1986). Item modeling procedures for constructing content-equivalent multiple-choice questions. *Medical Education, 20*, 53–56.

Leighton, J.P., & Gierl, M.J. (2007). Defining and evaluating models of cognition used in educational measurement to make inferences about examinees' thinking processes. *Educational Measurement: Issues and Practice, 26*, 3–16.

Leighton, J.P., Gierl, M.J., & Hunka, S. (2004). The attribute hierarchy method for cognitive assessment: A variation on Tatsuoka's rule-space approach. *Journal of Educational Measurement, 41*, 205–237.

Luecht, R.M. (2006a, May). *Engineering the test: From principled item design to automated test assembly*. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Dallas, TX.

Luecht, R.M. (2006b, September). *Assessment engineering: An emerging discipline*. Paper presented in the Centre for Research in Applied Measurement and Evaluation, University of Alberta, Edmonton, AB, Canada.

Luecht, R.M. (April, 2007). *Assessment engineering in language testing: From data models and templates to psychometrics*. Invited paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.

Luecht, R.M., Gierl, M.J., Tan, X., & Huff, K. (2006, April). *Scalability and the development of useful diagnostic scales*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Mislevy, R.J., & Riconscente, M.M. (2006). Evidence-centered assessment design. In S. M. Downing & T. Haladyna (Eds.), *Handbook of test development* (pp. 61–90). Mahwah, NJ: Erlbaum.

Mislevy, R.J., Wingersky, M.S., & Sheehan, K.M. (1994). *Dealing with uncertainty about item parameters: Expected response functions* (ETS Research Report 94-28-ONR). Princeton, NJ: Educational Testing Service.

Singley, M.K., & Bennett, R.E. (2002). Item generation and beyond: Applications of schema theory to mathematics assessment. In S. H. Irvine & P. C. Kyllonen (Eds.), *Item generation for test development* (pp. 361–384). Mahwah, NJ: Erlbaum.

Sinharay, S., Johnson, M.S., & Williamson, D.M. (2003). Calibrating item families and summarizing the results using family expected response functions. *Journal of Educational and Behavioral Statistics, 28*, 295–313.

Sinharay, S., & Johnson, M. (2005). *Analysis of data from an admissions test with item models*. (ETS Research Report 05-06). Princeton, NJ: Educational Testing Service.

Appendix

The appendix contains a second set of item model examples, as outlined in Table 1 (page 12), drawn from diverse content areas.

Model #1: (Biology) Stem: *Independent;* Options: *Randomly Selected;* Auxiliary Information: *None*

After a head injury, a mountain biker was assessed by a doctor. One of the tests the doctor did was to gently tap just below her kneecap. Also, he shone a light into each of her eyes and made observations. In both tests, he was trying to quickly rule out any neurological damage.

Assume that the biker's nervous system was not injured and that the biker's body responded in a normal way to the gentle tap. The neurological pathway that was followed when this response was elicited was

A. receptor, sensory neuron, interneuron, motor neuron, effector

B. effector, sensory neuron, interneuron, motor neuron, receptor

C. receptor, motor neuron, interneuron, sensory neuron, effector

D. effector, motor neuron, interneuron, sensory neuron, receptor

Item Model Variables

Stem	After a head injury, a S1 was assessed by a doctor. One of the tests the doctor did was to gently tap just below her kneecap. Also, he shone a light into each of her eyes and made observations. In both tests, he was trying to quickly rule out any neurological damage. Assume that the S1's nervous system was not injured and that the S1's body responded in a normal way to the gentle tap. The neurological pathway that was followed when this response was elicited was
Elements	S1 Range: "mountain biker", "truck driver", "driver", "motorcycle operator"
Options	Key: receptor, sensory neuron, interneuron, motor neuron, effector Distractors: all other combinations of the neurological processes
Auxiliary Information	None
Key	A

Model #2: (Social Studies) Stem: *Independent*; Options: *Constrained*; Auxiliary Information: *Map*

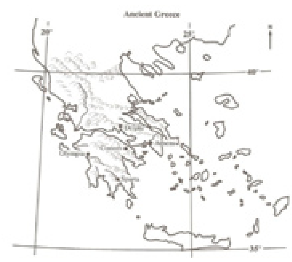
The approximate location of Athens is

A. 24°N and 38°E

B. 26°N and 42°E

C. 38°N and 24°E

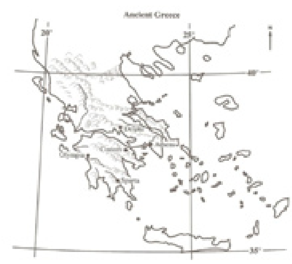
D. 42°N and 26°E



Item Model Variables

Stem

The approximate location of S1 is



Elements

I1 Value Range: 22–24 by 1
I2 Value Range: 37–39 by 1
S1 Range: “Athens”, “Delphi”, “Corinth”, “Olympia”, “Sparta”
As S1 = “Athens”, then I1=24, I2=38
As S1 = “Delphi”, then I1=23, I2=39
As S1 = “Corinth”, then I1=23, I2=38
As S1 = “Olympia”, then I1=22, I2=38
As S1 = “Sparta”, then I1=23, I2=37

Options

A. I1 °N and I2 °E
B. (I1 +2)°N and (I2 +4)°E
C. I2 °N and I1 °E
D. (I2 +4)°N and (I1 +2)°E

Auxiliary Information

Map of Ancient Greece

Key

C

Model #3: (Language Arts) Stem: *Independent*; Options: *Fixed*; Auxiliary Information: *Reading Passage*

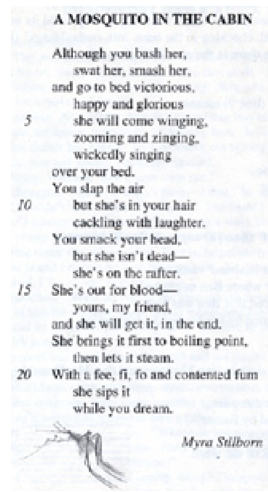
The word "it" as used in line 19 refers to

A. water

B. blood

C. food

D. air



A MOSQUITO IN THE CABIN

Although you bash her,
swat her, smash her,
and go to bed victorious,
happy and glorious
she will come winging,
zooming and zinging,
wickedly singing
over your bed.
You slap the air
but she's in your hair
cackling with laughter.
You smack your head,
but she isn't dead—
she's on the rafter.
She's out for blood—
yours, my friend,
and she will get it, in the end.
She brings it first to boiling point,
then lets it steam.
With a fee, fi, fo and contented fum
she sips it
while you dream.

Myra Stilborn

Item Model Variables

Stem	The word "it" as used in S1 refers to
Elements	S1 Range: "line 17", "line 18", "line 19", "line 21"
Options	<u>Key</u>: blood Distractors: water, food, air
Auxiliary Information	The text "A mosquito in the cabin". Author: Myra Stilborn.
Key	B

Model #4: (Language Arts) Stem: *Dependent*; Options: *Randomly Selected; Auxiliary Information: Reading Passage*

The word "converted" in the last sentence of Step 4 means

A. added

B. exposed

C. changed

D. compared

MAKE A COMPOSTER

You'll Need:
- *a wooden packing or orange crate with holes drilled in the sides*
- *a large plastic garbage bag*
- *soil*
- *kitchen garbage, including fruit and vegetables, coffee grounds, tea leaves, eggshells, and/or garden waste such as grass clippings. Don't include milk, cheese, yogurt, meat or fish. These will attract animals. And don't include foil, glass, metal or plastic, which don't decompose.*
- *a thermometer*
- *a pitchfork or spade*

1. Your wooden crate will serve as your composter. Put it outside in the garden or on a balcony. (For balconies, set the composter in a low-sided box lined with plastic. Your composter may make some mush, and you don't want it to leak all over the balcony.)
2. Now you need to expose your compost to soil and the organisms it contains so that they can break down your left-over food and turn it into rich earth. Put your left-overs into the composter and cover them with a layer of soil.
3. Keep your composter wet but not soaked. Water it every week, if it doesn't rain. If it rains too much, cover the top with a plastic garbage bag.
4. After two weeks, test the temperature of your pile by shoving the thermometer into the middle of it. Wait for three minutes, then check on the reading. Your pile should be about 60–70°C. If the temperature is lower, try adding more soil or water, or stir the compost up to add air. When the garbage is converted to soil, its temperature should drop to 40–50°C.
5. Keep doing Steps 3 and 4 for a few months, checking the temperature and turning the whole pile with a spade or pitchfork every month. At the end of four months, you'll have some rich fertile soil, where once there was only garbage. Use your new earth to perk up potted plants, fertilize your vegetable garden, or sprinkle on your flowerbeds.

Beth Savan

Item Model Variables

Stem	The word " S1 " in the last sentence of S2 means	
Elements	S1 Range: "converted", "perk up", "fertilize" S2 Range: "Step 4" or "Step 5" As S1="converted", then S2="Step 4" As S1="perk up", or "fertilize", then S2="Step 5"	
Options	*As S1 = "converted"* Key = "changed", "altered", or "transformed" Distractors="added", "exposed", " compared", "combined", "inserted", "adjoined", "showed", "contrasted", or "assessed"	*As S1 = "perk up" or "fertilize"* Key = "lift up", "boost", "uplift", "enrich", or "raise" Distractors= "prepare", "produce", "fix", "build up", "water", "adjoined", "showed", or "contrasted"
Auxiliary Information	The text "Make a composter". Author: Beth Savan.	
Key	One of the key alternatives presented in the option section will be randomly selected.	

Model #5: (Chemistry) Stem: *Dependent*; Options: *Constrained*; Auxiliary Information: *Graph*

The researchers made a graph of data collected along a river downstream from the plant. They know that rainbow trout cannot live in water with a pH of 5.5 or less.

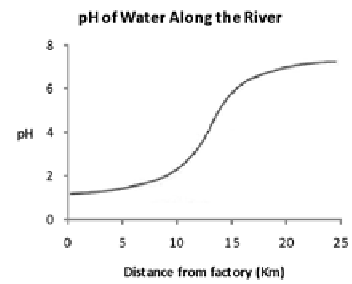
Based on information in the graph, rainbow trout in the river could survive downstream from the plant at a distance of

A. 0 to 10 km

B. 5 to 15 km

C. 10 to 20 km

D. 15 to 25 km

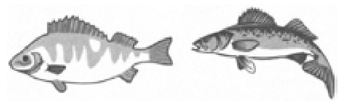


Item Model Variables

Stem	The researchers made a graph of data collected along a river downstream from the plant. They know that S1 cannot live in water with a pH of I1 or less. Based on information in the graph, S1 in the river could survive downstream from the plant at a distance of
Elements	I1 Value Range: 5.5, 5.6, 5.4, or 6.7 S1 Range: "rainbow trout", "fathead minnow", "pearl dace", "lake trout" As S1="rainbow trout", then I1=5.5 As S1="fathead minnow", then I1=5.6 As S1="pearl dace", then I1=5.4 As S1="lake trout", then I1=6.7
Options	A. 0 to I1 *2 km B. I1 to I1 *3 km C. I1 *2 to I1 *4 km D. I1 *3 to I1 *5 km
Auxiliary Information	Graph of pH of Water along the River
Key	D

Model #6: (Biology) Stem: *Dependent;* Options: *Fixed;* Auxiliary Information: *Picture*

Within a lake, you observe two different species of fish.



The process most likely responsible for the development of different species of fish is

A. Artificial selection

B. Artificial breeding

C. Selective breeding

D. Natural selection

Item Model Variables

Stem

Within a S1, you observe two different species of S2 .

[insert Picture P1]

The process most likely responsible for the development of different species of S2 is

Elements

S1 Range: "jungle", "lake", "mountain", "sea"
S2 Range: "monkey", "fish", "pine", "turtle"
P1 Range: picture of two species of monkeys, picture of two species of fish, picture of two species of pines, or picture of two species of turtles
As S1="jungle", then S2="monkey", P1=picture of two species of monkeys
As S1="lake", then S2="fish", P1=picture of two species of fish
As S1="mountain", then S2="pine", P1=picture of two species of pines
As S1="sea", then S2="turtle", P1=picture of two species of turtles

Options

A. Artificial selection
B. Artificial breeding
C. Selective breeding
D. Natural selection

Auxiliary Information

Picture of two species of monkeys; Picture of two species of fish; Picture of two species of pines; Picture of two species of turtles

Key

D

Model #7: (Physics) Stem: *Mixed;* Options: *Randomly Selected;* Auxiliary Information: *None*

While driving to the mountains, you see a glider. As the glider flies across a meadow, you see it rise sharply. This lift is created by the method of heat transfer known as

A. transmission

B. conduction

C. convection

D. radiation

Item Model Variables

Stem	While driving to the mountains, you see a S1. As the S1 flies across a S2, you see it rise sharply. This lift is created by the method of heat transfer known as
Elements	S1 Range: "glider", "hang glider", "bunch of leaves", "kite", "hawk"" S2 Range: "meadow", "pasture" As S1="glider" or "hang glider", then S2="meadow"; As S1="bunch of leaves", "kite", or "hawk", then S2="pasture"
Options	Key: radiation Distractors: transmission. conduction, convection, diffusion, absorption
Auxiliary Information	None
Key	D

Model #8: (Mathematics) Stem: *Mixed*; Options: *Constrained*; Auxiliary Information: *None*

Ann has paid $1525 for planting her lawn. The cost of lawn is $45/m^2. Given the shape of her lawn is square, what is the side length of Ann's lawn?

A. 5.8

B. 6.8

C. 4.8

D. 7.3

Item Model Variables

Stem	Ann has paid $ I1 for planting her lawn. The cost of lawn is $ I2/m^2. Given the shape of her lawn is S1, what is the S2 of Ann's lawn?
Elements	I1 Value Range: 1525 – 1675 by 75 I2 Value Range: 30 or 40 S1 Range: "square" or "circular" S2 Range: "side length" or "radius" As S1 = "square", then S2 = "side length" As S1 = "circular", then S2 = "radius"
Options	*As S1 = "square"* A. $= \sqrt{I1/I2}$ B. $= \sqrt{I1/I2} + 1$ C. $= \sqrt{I1/I2} - 1$ D. $= \sqrt{I1/I2} + 1.5$
	As S1 = "circular" A. $= \sqrt{I1/I2 * 3.14}$ B. $= \sqrt{I1/I2 * 3.14} + 1$ C. $= \sqrt{I1/I2 * 3.14} - 1$ D. $= \sqrt{I1/I2 * 3.14} + 1.5$
Auxiliary Information	None
Key	A

Model #9: (Biology) Stem: *Mixed*; Options: *Fixed*; Auxiliary Information: *Graph*

On a newly formed island, successful populations of grasses and a species of mouse appeared. Later, a species of hawks flew in. The hawks feed on mice. The population levels of mice and hawks are represented in the graph.

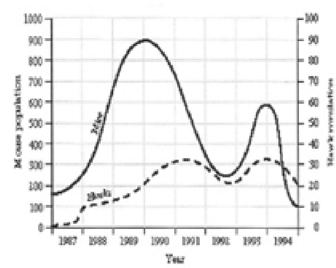
In 1991, the data for the mice indicates that

A. r is negative because $b < d$

B. r is negative because $b > d$

C. r is positive because $b < d$

D. r is positive because $b > d$



r = per capita population growth rate (b–d)

b = per capita births

d = per capita deaths

Item Model Variables

Stem

On S1 , successful populations of grasses and a species of S2 appeared. Later, a species of S3 flew in. The S3 feed on S2 . The population levels of S2 and S3 are represented in the graph.

In I1 , the data for the S2 indicates that

Elements

I1 Range: "1990", "1991", "1994"

S1 Range: "a newly formed island", "distant forest", "isolated jungle"

S2 Range: "worms", "beetles", "mice", "snakes", "fish", "lizards", "insects", "bugs", "frogs"

S3 Range: "hawks", "eagles", "ravens"

As S3="hawks", then S2="worms", "beetles", "mice", "snakes", or "frogs"

As S3="eagles", then S2="fish", "snakes", or "lizards"

As S3="ravens", then S2="lizards", "insects", "bugs", or "frogs"

Auxiliary Information

Graph with Yearly Populations

Key

A

Model #10: (Architecture) Stem: *Fixed*; Options: *Randomly Selected*; Auxiliary Information: *None*

What is one purpose of a bid repository?

A. To distribute bid documents and addenda
B. To administer construction contracts
C. To be a storage facility for typical construction details
D. To ensure substantial performance

Item Model Variables

Stem	What is one purpose of a bid repository?
Options	Key: "To distribute bid documents and addenda", "To manage the distribution of bid information to trade contractors", or "The timely collection of trade contractor bids" Distractors: "To administer construction contracts", "To be a storage facility for typical construction details", "To ensure substantial performance", "To extend the tendering process timeline", "To assist Owners in the selection of architects", "To collect bonds from prospective contractors", "To tabulate and monitor LEED credits", "To assist the Owner in identifying a financial partner", or "The fund construction deficiencies"
Auxiliary Information	None
Key	A

Author Note

The research reported in this study was conducted with funds provided to the first author by Learner Assessment at Alberta Education. We would like to thank Alberta Education for their support. However, the authors are solely responsible for the methods, procedures, and interpretations expressed in this study. We would also like to thank Andrew Turnbull for his help in developing the computer software, IGOR Version 1.0.

Author Biographies

Mark J. Gierl is Professor of Educational Psychology and Canada Research Chair in Educational Measurement, Centre for Research in Applied Measurement and Evaluation, University of Alberta, Edmonton, Alberta, Canada, T6G 2G5; mark.gierl@ualberta.ca. He earned his Ph.D. in quantitative methods from the University of Illinois, Urbana-Champaign, in 1996. His primary research interests are cognitive diagnostic assessment; assessment engineering, including construct mapping, automated item generation, and automated test assembly; differential item and bundle functioning; psychometric methods for evaluating test translation and adaptation.

Jiawen Zhou is a Ph.D. student in the Centre for Research in Applied Measurement and Evaluation at the University of Alberta in Edmonton, Alberta, Canada. Her research interests include cognitive diagnostic assessment, automated item generation, and computer adaptive testing.

Cecilia Alves is a Ph.D. student in the Centre for Research in Applied Measurement and Evaluation at the University of Alberta in Edmonton, Alberta, Canada. From 2001 to 2006 she worked as a Psychometrician for a testing agency in Brazil (CESPE/University of Brasilia). Her interests have been focused on cutting-edge methodologies that can be used to overcome problems of traditional test/item development methods. She has conducted research in the areas of cognitive diagnostic assessment, automated item generation, and automated test assembly.

J·T·L·A

The Journal of Technology, Learning, and Assessment

www.jtla.org

Technology and Assessment Study Collaborative
Caroline A. & Peter S. Lynch School of Education, Boston College