

## Algorithmically Derived Examination Specifications

One of the most important components of a legally defensible examination program is a job analysis study. This article describes a procedure SMT incorporates to develop final test specifications for an examination program. These specifications include testing mode (e.g., multiple choice test, performance, or essay test) and the relative weighting of content area for subsequent assessment. *The Standards for Educational and Psychological Testing (1999)* provide guidance on the establishment of test specifications for credentialing examinations:

*The content domain to be covered by a credentialing test should be defined clearly and justified in terms of the importance of the content for credential-worthy performance in an occupation or profession. A rationale should be provided to support a claim the knowledge or skills being assessed are required for credential-worthy performance in an occupation and are consistent with the purpose for which the licensing or certification program was instituted. (Standard 14.14)*

In general, a job analysis is a study of the important tasks associated with a professional role. Using a survey-based methodology helps generalize the results of the study. SMT uses algorithmically generated information to facilitate decisions on testing type and relative content area weighting. For the purpose of this article and example, the discussion is limited to one rating scale type, the importance scale. However, we typically use multiple rating scales during our studies.

Test specification design should incorporate important tasks that data from a job analysis support. It is this link that helps establish a validity claim or link between test score inferences and the occupation. This data typically comes from the judgments of hundreds or thousands of professionals who complete a task analysis survey.

An argument can be made that relative importance of certain content, as provided by mean importance ratings should be included in the decision of how to weight certain areas of a content outline.

The following discussion presents a three-step process that describes our procedure for delineating content and assigning content weights. In addition, the process helps credentialing organizations decide how to weight different examination form types.

### **Step 1: Determine Eligible Content**

Determine what work-related tasks are to be included in the survey by conducting an exhaustive review of curriculum and other work-related documents. Then, a panel of Subject Matter Experts (SMEs) is convened to develop a task list based on the review. Next, the task list is incorporated into a survey, demographic questions for subsequent

## Algorithmically Derived Examination Specifications

analyses are developed, and the survey is distributed to a representative sample of practitioners.

After the survey has been administered and data has been tabulated, decisions about task inclusion and exclusion should be made based on mean task ratings. Below is an example of some fictitious data describing this step. Relative mean importance ratings are used to determine what content becomes eligible for assessment. The scale in this case is:

How important is this activity to the successful performance of the professional?

1. Minimum Importance
2. Below Average Importance
3. Above Average Importance
4. Maximum Importance

**Table 1, Eligible Content**

Content Area	Mean Importance Rating	Task Included in final outline?
A)	2.56	1
<b>B)</b>	<b>2.49</b>	<b>0</b>
C)	2.67	1
D)	2.98	1
E)	2.75	1
<b>A)</b>	<b>2.41</b>	<b>0</b>
B)	2.65	1
C)	2.85	1
D)	2.98	1
E)	3.01	1
F)	2.78	1
G)	2.85	1
A)	3.28	1
B)	3.45	1

Tasks 1B and 2A in *Table 1* would be excluded from the final examination content outline. The criterion is deliberated and set by the panel of SMEs. The rationale for task exclusion in this sample is that the mean importance rating was below 2.50 (i.e, average importance based on the scale). All other tasks remain eligible for assessment.

### **Step 2: Determine what type of assessment should be used**

During this phase of specification development, a panel of SMEs determine what testing modality should be incorporated. For the purpose of this example, it is assumed that the credentialing agency has decided that three types of assessment will be used. After discussing what types of knowledge may be measured by various assessment modes, the panel of SMEs reviews the eligible tasks to determine what assessment type or what type of knowledge should be measured. In this example, the three assessment types are 1) multiple choice examination, 2) performance assessment, and 3) essay examination. During step one, tasks 1B and 2A were eliminated so SMEs will not place these tasks into

any of the three categories.

## Algorithmically Derived Examination Specifications

As can be viewed in Table 2, task 2G that was eligible for assessment on all three- assessment types will not be included on multiple-choice examination. The SMEs indicated that content area 1 is not eligible for assessment on the performance test, and content area 2 is not eligible for assessment on the Essay. SMT is considering the use of an assessment type scale on future surveys. After determining what is eligible for assessment by examination type, the next step is to determine how examination weights should be distributed by content. Since this defines all future examinations, it is a very important step.

**Step 3: Determine content weighting (how many questions or prompts from each content area should be on each assessment type?)**

**Table 2, Assessment Type**

	Content Area	Mean Importance Rating	Task Included in final outline?	On Multiple Choice Test?	On OSCE Performance Test?	On Essay Test?
1)	A)	2.56	1	1	0	1
	B)	<b>2.49</b>	<b>0</b>			
	C)	2.67	1	1	0	1
	D)	2.98	1	1	0	1
	E)	2.75	1	1	0	1
2)	A)	<b>2.41</b>	<b>0</b>			
	B)	2.65	1	1	1	0
	C)	2.85	1	1	1	0
	D)	2.98	1	1	1	0
	E)	3.01	1	1	1	0
	F)	2.78	1	1	1	0
	G)	2.85	1	0	1	0

As indicated at the beginning of this article, the example provided is limited to an importance scale. Content weighting should be assigned based on relative importance. To that end, we have incorporated an algorithm to help SMEs deliberate content weights. These calculations for our sample are shown in Table 3.

**Table 3, Content Weighting**

Content Area	Mean Importance Rating	Task Included in final outline?		On Multiple Choice Test?	On OSCE Performance Test?		On Essay Test?	
1) A)	2.56	1	1	2.56	0	0	1	2.56
B)	<b>2.49</b>	<b>0</b>						
C)	2.67	1	1	2.67	0	0	1	2.67
D)	2.98	1	1	2.98	0	0	1	2.98
E)	2.75	1	1	2.75	0	0	1	2.75
				10.96		0		10.96
				<b>30%</b>		<b>0%</b>		<b>51%</b>
2) A)	<b>2.41</b>	<b>0</b>						
B)	2.65	1	1	2.65	1	2.65	0	0
C)	2.85	1	1	2.85	1	2.85	0	0
D)	2.98	1	1	2.98	1	2.98	0	0
E)	3.01	1	1	3.01	1	3.01	0	0
F)	2.78	1	1	2.78	1	2.78	0	0
G)	2.85	1	0	0	1	2.85	0	0
				14.27		17.12		0
				<b>40%</b>		<b>62%</b>		<b>0%</b>
3) A)	3.28	1	1	3.28	1	3.28	1	3.28
B)	3.45	1	1	3.45	1	3.45	1	3.45
C)	3.89	1	1	3.89	1	3.89	1	3.89
				10.62		10.62		10.62
				<b>30%</b>		<b>38%</b>		<b>49%</b>

Content weights are based on the sum of the mean importance ratings for a given area divided by the sum of all eligible mean importance ratings.

**Final Specifications:**

Content Area	Multiple Choice	OSCE	Essay
1)	30%	XX	51%
2)	40%	62%	XX
3)	30%	38%	49%

The example presented in this article is simplistic. There may be instances where multiple rating scales (frequency and importance) need manipulation to create one composite score for analyses. The data used during these meetings are used to facilitate SME decisions. While the data is based on a large cohort of respondents, minor adjustments to the final specifications can be made if an SME panel can articulate a

defensible position for such modifications.

In summary, we use job analysis data to delineate examination specifications. It is anchored to a large sample’s perceptions of what they feel is important to the profession. In addition, SMT is incorporating a new continual model of Job Analysis. Look for more information about this in our next issue of Dimensions. For more information contact Reed A. Castle, PhD, SMT’s Director of Research and Development.