



# Dimensions

## Schroeder Measurement Technologies, Inc.

2536 Countryside Blvd., Suite 300, Clearwater, FL 33763 (727) 799-1638

Website [www.smtorg.com](http://www.smtorg.com)

### Technology

## Breaking Down the Language Barrier

**M**any credentialing agencies are calling for the addition of test forms in languages other than English to be administered to non-English speaking candidates. Major issues surround the translation of examinations from English into other languages, ranging from questions about the validity of the translated examinations to concerns about logistical issues of how to efficiently work with English questions and their translated counterparts in the test development process.

Despite the major issues that exist concerning the translation of examination questions, many agencies are driven by legal or even marketing mandates that require them to provide translated examinations.



SMT-Bank Designer  
Steve Schimsky

To meet the need for

translated examinations, SMT has developed a unique facility within our item banking system, SMT-Bank, that eliminates much of the work associated with managing multi-language item banks and avoids many of the errors that can occur. To understand how SMT-Bank deals with the problem of assembling forms in multiple languages, it is useful to first review how forms are assembled using a typical item bank.

In virtually all item banking systems, examination questions are identified by some unique index, and this is usually a simple item

## INSIDE

- 1 Breaking Down the Language Barrier
- 2 Scoring Examinations: Equating and Scaling
- 3 Spotlight on People

number. An examination form is developed by listing the items to be placed on the examination. This is usually done by creating a data file which contains the indices of those items to be placed on the examination. Software, which is part of the item banking system, is then used to create the camera ready copy for an examination that is composed of the questions listed in the file.

When the questions in an item bank are translated, it is necessary to identify these questions by using the same indexing system. When an examination is developed, for example, in Spanish, that is equivalent to a base English examination, a file which contains indices of the Spanish texts of the English items must be created. This file is used to generate the camera ready copy of the Spanish form. In using such an approach, a number of problems can exist and a number of errors can occur.

First, it is necessary to develop a scheme, and this is usually done outside of the item banking system to associate the English questions with their translated equivalents.

Errors in the management of this concordance will result in errors in examination forms. Second, it is necessary to develop files of the question indices for both the English and the translated examinations. If an error is made in constructing either of these files, the examinations will not be equivalent. Third, a separate key is usually generated for both the English and the translated examination form. This is another source of error.

**SMT-Bank's** new multi-language capabilities allow translated items to be automatically associated with their equivalent English items. Hence, when an examination form is developed in English using SMT-Bank, an equivalent examination in a second or even third language can be assembled automatically and the same key is generated for each examination. In other words, once the translated item is associated correctly with its English equivalent, no errors can occur in generating the camera ready copy or the key. Also, when an item is accessed by SMT-Bank, a list of all languages into which the item has been

translated is displayed, and the user may view any translation, as well as move from one translation of the question to another, while at the same point of accession in the item bank. Using this feature, it is even possible to develop multiple translations of the same question into different dialects of the same language. This means, for example, that it is easy to manage an item bank which may contain translations of English items into both the Parisian and Canadian dialects of French. SMT-Bank provides for the statistical data for the translated examinations to be calculated and evaluated separately for each language, thus facilitating the evaluation of translation-based item performance differences.

SMT-Bank has become an invaluable tool for the management and maintenance of multi-language item banks for licensure and certification programs. SMT-Bank can automatically assemble examinations consisting of translated items, and then independently evaluate and analyze their psychometric properties. All of this is accomplished without confusion or error. If your

testing program requires examinations in languages in addition to English, you may find that SMT's technology can save you time and money as well as help you avoid errors.

## Scoring Examinations: Equating and Scaling

Many of us tend to think of examinations in terms of the tests we took in school. When we took an

### TESTING ISSUES

or college, our teacher might have given us a test consisting of 50 multiple-choice items. All students in the class would receive the same examination. Our score was simply the number of questions correct, and sometimes this would be represented as a simple percentage. If we got 40 of the 50 items correct, our score would be 80 percent. The process of scoring these tests was simple and easy to understand.



SMT President  
Dr. Lee Schroeder

For professional regulatory examinations, however, calculations are not the same, nor are they as easy. Examinations used for high-stakes decision making must follow more rigorous standards than do the teacher-made examinations from school. One examination used for high-stakes decision making is the Scholastic Achievement Test (SAT). Those of us that took the SAT may remember that it is a battery of two examinations, the Verbal and Mathematical examinations. Each examination had a score ranging from 200 to 800, though none of us answered as many as 800 mathematical questions (or verbal questions) on the SAT. Obviously, the scores were not percentages. In fact, the scores are from a reporting scale that is different from, though related to, the raw score or number of questions correct. What we may not

have noticed is that the candidate sitting next to us received an entirely different set of questions than we did.

Despite the best efforts of professional test developers, no two examinations are exactly the same in terms of difficulty. Thus, without adjustment, some candidates could be advantaged by being assigned easier forms, while other candidates may be disadvantaged by being assigned more difficult forms. This is when equating and scaling become essential to fairness.

The use of scaling and equating in the preparation of professional regulatory examinations has been supported in the courts. For example, a lawsuit involving an SMT client was heard recently where a failing candidate complained about the unfairness of an examination score. The candidate blamed this unfairness on calculations associated with equating and scaling. When these processes were explained by an SMT expert witness, the trial judge found no merit in the candidate's complaint and found in

favor of SMT's client. This is typically the result of such litigation.

### Equating

The process of equating and scaling are complicated and somewhat abstract. In view of this, the following example explains these processes in terms that should be easy to understand.

*Suppose that two different groups of candidates (Group 1 and Group 2) took two different forms (Form A and Form B) of an examination on different dates. This could occur if one group of candidates took a given test form in January and a group composed of different candidates took another form of the examination in February.*

If the average test score for the two groups is different, what conclusions can be drawn about the two groups or the two forms? Do both groups have the same level of knowledge on the two examinations, or is one group more knowledgeable than the other? Are both examinations of the same level of difficulty, or is one

examination more difficult than the other?

Suppose, for example, that the average score for Group 1 was 38 and that the average score for Group 2 was 33. (Assume that both Form A and Form B are 50 items in length.) The following is a list of several possible situations that could have contributed to this 5-point average difference:

- Form A and Form B are equally difficult, but Group 1 is more knowledgeable than Group 2. (The entire 5-point average difference is due to **group differences**.)
- Form A is easier than Form B, but Group 1 and Group 2 have the same level of knowledge. (The entire 5-point average difference is due to **form difficulty differences**.)
- Form A is easier than Form B, and Group 1 is more able than Group 2. (Part of the 5-point difference is due to differences in **form difficulty** and the other part of the difference is due to **group differences**.)

Clearly, we do not know very much about the relative difficulty of the two forms. We are also unaware of the relative levels of knowledge in the two groups when each group takes two different forms of an examination.

A common technique to help understand form and group differences is to include a common set of items in both forms of the examination. These common items are sometimes referred to as an *anchor test*. Suppose in the previous example that 25 questions were in common between Form A and Form B on each examination. This could be represented by the following tables:

Form A		Form B
25 Items Unique to Form A		25 Items Unique to Form B
Anchor Test: 25 Items in Common with Form B	Same Items ↔	Anchor Test: 25 Items in Common with Form A

In these figures, both Group 1 and Group 2 took the Anchor test. We can determine the average scores on the anchor

test, and these averages tell us how Group 1 and Group 2 compare in terms of knowledge of the material being tested.

In addition, from the difference between the two groups on the anchor test, we can determine what portion of the difference in average scores in either examination is due to group differences and what portion is due to form differences. The process of making these calculations is called *equating*.

To further explain the process of equating, consider the following example:

*Suppose two forms of a 50-question examination are administered, Form A to Group 1 and Form B to Group 2. Suppose the average of Group 1 on Form A is 38, and the average for Group 2 on Form B is 33. Also suppose that an anchor test of 25 questions is part of both Form A and Form B and that Group 1 and Group 2 have an average score of 15 on the anchor test. This data is shown in the following table:*

**Form A**

**Form B**

25 Items Unique to Form A	Same Items ↔	25 Items Unique to Form B
Anchor Test: Average is 15 out of 25		Anchor Test: Average is 15 out of 25
<b>Average Score: 38 out of 50</b>		<b>Average Score: 33 out of 50</b>

Because both groups have the same average score on the Anchor Test, we can say that the groups are similarly knowledgeable of the material in the examination. Thus, the difference in the averages for Form A (Average=38) and Form B (Average=33) is due to differences in difficulty between the forms.

In this case, candidates in Group 2 taking Form B would receive an average score of 33, while candidates in Group 1 with an equal level of knowledge as those in Group 2, but taking Form A, would receive an average score of 38. This would be unfair to all candidates in Group 2.

Further, if the minimum passing score on the test was set at 70 percent, many candidates would pass if they take Form A, but fail if

they take Form B. This would be extremely unfair to candidates in Group 2.

A simple solution to this problem would be to add 5 points to the scores of candidates who take Form B. This would make a correct answer on Form B have more weight or a higher value than a question on Form A. This formula would convert a score of 33 on form B to a score of 38, making it have an equivalent meaning to scores on Form A.

The above provided scoring adjustment is an example of equating. Equating determines how scores from one test may be weighted so as to have equal meaning with scores from another test. This eliminates the effects of differences in test difficulty. Since test forms do differ in difficulty, equating is important to ensure fairness to candidates.

**Scaling**

Given that equating is necessary, we must also know how to report scores on equated examinations. In the example above, a candidate taking Form B with a score of 33, has the same level of knowledge as

## SPOTLIGHT ON PEOPLE

be represented in various ways, such as:

- Add 5 points to all Form B scores, thus reporting an earned score of 38 for candidates who get 33 questions correct. In this case, how are sub-scores reported? Do candidates who take Form A wonder why their scores are not adjusted? What do we tell them?
- Subtract 5 points from all Form A scores, thus reporting an earned score of 33 for candidates who get 38 questions correct. Do candidates who take Form A wonder why their scores are adjusted? What do we tell them?

Actually, there is no way to report equal raw or percent scores on equated examinations without creating some confusion. To prevent confusion, the process of scaling is used to report scores from equated examinations. This process begins with the adoption of an arbitrary scale. To further explain the process of scaling we could, for example create a

scale that may run from 5 to 15 with the cut-score set at 12. A score of 38 on Form A may be set at 13 on this scale. Further, all scores equal to 38 on future forms would also be set at 13. Therefore, in this example, a score of 33 on Form B would have a scaled score of 13 as well.

### Summary

This article was written to explain why the process of equating and scaling are necessary to fairness for high-stakes examinations. Equating helps us understand whether differences in test scores are due to form difficulty or group differences. Scaling provides a means of representing test scores from test forms of different levels of difficulty. Both equating and scaling assure candidates the highest level of fairness.

## A New Vice President

In this issue of *DIMENSIONS*, we introduce

you to Matthew Wenger who was recently promoted from Director of Marketing and Contract Management to Vice President. Mr. Wenger received his Bachelor's degree in Business Administration, specializing in marketing, from the University of Toledo and continued with graduate work at the University of South Florida in business management. He brings the perspective of having managed large test administration operations and having coordinated the development and delivery of high-quality test development products.

Mr. Wenger is a recognized expert in program management and marketing. He has served as Area Operations Manager for Southland, as well as Representative and District Sales Manager for the Florida Lottery.

Additionally, Mr. Wenger served as Regional Director, Program Manager, and Executive Program Director at the National Assessment Institute (NAI).

Mr. Wenger's experience and skill are clearly significant assets to SMT. Currently serving as Vice President, he

has three major responsibilities.

1. Mr. Wenger develops and markets technical and price proposals to ensure that SMT's services are appropriately tailored to meet the needs of prospective clients.



2. Mr. Wenger oversees the contract management function. In this capacity, Mr. Wenger ensures that each client contract is being appropriately managed by working with internal department heads to coordinate that various outputs are provided to clients in the appropriate quality, quantity, and timeliness.
3. Finally, Mr. Wenger serves an operational role at SMT. These activities include tasks such as budget development, forecasting, and managing day-to-day responsibilities within the company.

In addition to Mr. Wenger's busy schedule, he is a member of the Council of Licensure Enforcement and Regulation (CLEAR), the American Marketing Association, and the Financial Management Association.

SMT-Vice President  
Matt Wenger

For more information about SMT, our product and service offerings or to discuss your testing program, please contact :

Matthew J. Wenger  
Vice President  
Schroeder Measurement  
Technologies  
2536 Countryside Blvd., Suite 300  
Clearwater, FL 33763  
(727) 799-1638

*This newsletter and all information contained herein are a Schroeder Measurement Technologies, Inc., 1998. Website [www.smtorg.com](http://www.smtorg.com)*

--	--	--