

Home Office:

2494 Bayshore Blvd.
Suite 201
Dunedin, FL 34698
800.556.0484

Business Development:

262 Chapman Road
Suite 108
Newark, DE 19702
302.737.0796

Candidate Service Offices in:

- London
- Tokyo
- Hong Kong

SMT and its staff maintain membership in:

- NOCA
- ATP
- CLEAR
- AMC Institute
- APA
- NCME
- ABNS
- ITC

IN THIS ISSUE:

Scaling & Equating |

New Clients |

Staff Spotlight: 3
Darcy St. George

Scoring Tests: Equating & Scaling

When you took an examination in high school or college, the teacher might have given you a test consisting of 50 multiple-choice items. All students in the class received the same examination. The score you received 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.

For professional regulatory examinations, however, calculations are not so simple. Examinations used for high-stakes decision making must follow more rigorous standards than teacher-made tests. If you took the SAT, you may remember that it was a bat-

tery of two examinations—Verbal and Mathematical. Each examination had a score ranging from 200 to 800, even though you didn't answer 800 mathematical on either section. The scores were not percentages. In fact, the scores were derived from a reporting scale that is different than, though related to, the number of correct responses. In all likelihood, the candidate sitting next to you received an entirely different set of questions.

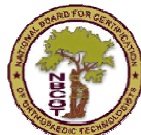
Despite the best efforts of professional test developers, no two examinations are precisely the same in terms of difficulty. Without adjustments, candidates may be advantaged or disadvantaged by being assigned different forms. This is why equating and scaling are essential for fairness.

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 some time ago 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—the typical result of such litigation when appropriate equating and scaling processes are used in test development.

(See Scaling & Equating, pg. 2)

SMT Welcomes New Clients



The National Board for Certification of Orthopaedic Technologists, inc. (NBCOT), a not-for-profit credentialing agency providing certification for the Orthopaedic Technology profession for over twenty years, has selected SMT to provide full test development and computer administration services. The NBCOT serves

the general public by developing, administering, and continually reviewing a certification process that reflects current standards of competent practice in Orthopaedic Technology.

The NBCOT was formed in 1982 to credential those persons with proven measurable knowledge and experience as Orthopaedic Technologists. You may learn more about NBCOT at

www.nbcot.net.



The American Gastroenterological Association (AGA) Institute recently contracted with SMT for delivery of its Gastroenterology Training Examination (GTE) via secure web portals to multiple approved training institutions. The GTE™ serves as an assessment tool and benchmark for

(See New Clients, pg. 2)

New Clients

Additional new SMT clients include:

- Special Care Dentistry Association
- College of Psychologists on Ontario
- National Parks & Recreations Association
- Business Continuity Institute
- National Council on Family Relations

(Continued from pg. 1)

both training directors and fellows. The exam's overall goal is to enhance clinical care and research by advancing quality and consistency across GI training. Training directors and fellows benefit from the exam by receiving detailed score reports allowing them to effectively gauge progress over time.

The National Business Aviation Association (NBAA) has se-

 lected SMT for the ongoing development and administration of its Certified Aviation Manager (CAM) program.

The Certified Aviation Manager Program identifies qualified professionals to lead flight departments and companies using business aircraft. The certification process tests individuals for knowledge and experience in five subject areas: leadership, personnel man-

agement, operations, technical and facilities services, and business management. NBAA and the business aviation community acknowledge Certified Aviation Managers as professionals who have reached a distinct level of industry knowledge. To start the certification process, individuals must first qualify to take the CAM Exam by submitting an application. Learn more at www.nbaa.org/cam

Equating & Scaling

(Continued from pg. 1)

Equating

The process of equating and scaling are complicated and somewhat abstract. In view of this, the following example explains these processes in simplified terms.

Two groups of candidates (Group 1 and Group 2) take different 50-item exam forms (Form A and Form B) of an examination on different dates. The average score for Group 1 on Form A is 38, while the average score for Group 2 is 33.

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? Here are some possible answers:

Form A and Form B are equally difficult, but Group 1 is more knowledgeable than Group 2. (The entire 5-point average difference is the result of **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 the result of **form difficulty differences**.)

Form A is easier than Form B, and Group 1 is more able than Group 2. (Some of the 5-point difference is the result of differences in **form difficulty** and some the result of **group differences**.)

Without some basis for comparing the groups and forms, there is no way to know whether the difference is the result of the group or the exam form. A common technique to help understand form and group differences is to include a common set of items in both forms of the examina-

tion. These common items are sometimes referred to as an *anchor* test. Suppose, for example, that Form A and B had 25 questions in common, as represented in Figure 1.

In this figure, both Group 1 and Group 2 took the *anchor* test (the 25 items in common). 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*. Continuing the example:

Two forms of a 50-question examination are administered, Form (See Scaling & Equating, page 3)

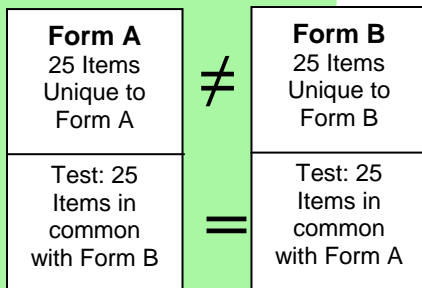


Figure 1

Scaling & Equating

(Continued from pg. 2)

A to Group 1 and Form B to Group 2. The average of Group 1 on Form A is 38, and the average for Group 2 on Form B is 33. An anchor test of 25 questions is part of both forms and that both groups have an average score of 15 on the anchor test (see Figure 2).

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 the result of 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. Further, if the minimum passing score on the test was determined to be 70 percent (35 of 50), a number candidates in Group 2 would fail simply because they took Form B. Clearly, this is unfair (and would be a legitimate basis for candidate challenges).

A simple solution to this problem would be to add five 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.

Such a scoring adjustment is a simplified example of equating. Equating determines how scores from one form may be weighted so as to have a value equal to scores from an-

other form. This eliminates the effects of differences in test difficulty and assures fairness to all candidates.

Scaling

Given that equating is necessary, we must also determine how best to report scores on equated examinations. In the example above, a candidate taking Form B with a score of 33, has displayed the same level of knowledge as a candidate with a score of 38 on Form A.

If we simply set the passing score at 38 for Form B, we will inevitably get candidate questions that will require complex explanations. *Why did I fail with a 36, while my friend passed?* We could, of course, simply add five points to all Form B scores, but then we face a challenge in reporting sub-scores, and candidates would certainly ask why scores on their test weren't

adjusted when others were.

These issues illustrate why it is best not to report number correct or percent scores on certification and licensing examinations. 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. We could, for example, create a scale that ranges from 5 to 15 with the cut-score set at 12 through an appropriate standard-setting procedure. A score of 38 on Form A may be set at 13 on this scale. Therefore, in this example, a score of 33 on Form B would have a scaled score of 13 as well.

Summary

The process of equating and scaling is necessary to fairness for high-stakes examinations. Equating allows us understand whether differences in test scores are the result of 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.

“The process of equating and scaling is necessary to fairness for high-stakes examinations.”

Note: along with the new format of *Dimensions*, we are also offering a new delivery option. If you would prefer to receive *Dimensions* by e-mail, please send a message to dimensions@smttest.com

<p>Form A 25 Items Unique to Form A</p>	<p>Form B 25 Items Unique to Form B</p>
<p>Anchor test: Average is 15 out of 25</p>	<p>Anchor test: Average is 15 out of 25</p>
<p>Average Score: 38 of 50</p>	<p>Average Score: 38 of 50</p>

Figure 2

Form A and 33 for Form B, we will inevitably get candidate questions that will require complex explanations. *Why did I fail with a 36, while my friend passed?* We could, of course, simply add five points to all Form B scores, but then we face a challenge in reporting sub-scores, and candidates would certainly ask why scores on their test weren't



Darcy St. George

As Contract Manager for SMT, Darcy works with each department to manage and monitor the progress of job functions that need to be completed for each project by defining tasks and responsibilities, and assigning them to the appropriate department. She ensures continuous progress by monitoring and communicating deadlines and upcoming events, and providing relevant contract information to key project leads in each depart-

ment. Prior to joining SMT, Darcy St. George performed as a Business Analyst and



Project Manager for Luxottica Retail. While participating in their Give the Gift of Sight Program, Darcy participated on seven international eyeglass missions to deliver eyewear to recipients in developing countries in Asia, Africa and South America. In her free time, Darcy enjoys tennis, biking and traveling.