



## In This Issue

|  |     |
|--|-----|
| How Item Statistics Can Improve Your Examination ..... | 1-2 |
| South Carolina Cosmetology and Barbering .....         | 3   |
| Behind the Scenes .....                                | 3   |
| ReMarks .....  | 4   |

# How Item Statistics Can Improve Your Examination

By: James D. Lehman, Ph.D.

Writing and editing multiple-choice examination questions (items) is part art and part science. Part of the science is the ability to calculate certain statistics on each question in an examination. The data and raw numbers come from each candidate that answered a particular multiple-choice question. Commonly, multiple-choice questions have four possible answer choices – A, B, C, or D. When it comes time to analyze the question data, those A-D responses are likely converted into the numbers 1-4. This is done by the computer for each response to each question by every candidate.

Once the candidate responses are converted into numbers, a statistician specializing in applied statistics in the area of tests and measurements – a psychometrician – uses specialized software to calculate item statistics for each item in the examination. Below you will find how statistics are calculated, and how they are used:

## Percent Correct

Percentages are used in daily life fairly often. Likewise, percentages are also used to indicate how easy or difficult an item is for candidates. So for our candidate population, percent correct tells us if the knowledge or skill assessed by the item is complicated or simple. Simple item knowledge in exam items tends to be factual knowledge: What year was the Declaration of Independence signed? More complex items tend to involve a more obscure knowledge or a higher level of cognitive skill: What papermaking process was used to produce the paper upon which the Declaration of Independence was written?

## Item Discrimination

Item discrimination is a little harder to explain. It is assumed that candidates who score well on an examination have more knowledge, skill, or ability in the area we are testing; whatever that might be. This logic applies to each individual question as well. It is expected that a large proportion of high scoring candidates will answer each question correctly. Conversely, it is expected that a smaller proportion of low scoring candidates will answer each question correctly. When these two things occur for a question, the item discrimination is usually moderate to high and positive (discrimination can range from (-1.00) to (+1.00)). There is usually something wrong with a question when high scoring candidates do not select the correct answer and low scoring candidates DO select the correct answer. Such questions have negative discrimination values and need to be edited or removed from the examination.

## Distractor Statistics

There are several different distractor statistics that are commonly used; two of the most common will be discussed. But first, what is a distractor? In a multiple-choice item with four answer choices, three of those options are incorrect. Those incorrect answer choices are called distractors. Good distractors are plausible but still incorrect. So in a sense, they 'distract' less knowledgeable candidates from selecting the correct answer choice. In order to truly understand each question on an examination, we need also to understand how the distractors are performing. It is important that at least some candidates are choosing each distractor. If not, it suggests that one or more distractors are so obviously incorrect that no candidate will choose them. Distractors should be plausible. If they are plausible, then at least a few candidates will select them as their answer. So an item analysis will show, for each item, the percent of candidates choosing each distractor. Another very helpful statistic that is calculated for each distractor and the keyed response is the average score on the entire examination for candidates choosing each distractor or the key. These distractor averages should fall below that for candidates choosing the keyed response. (continued on pg. 2)

Item statistics can vastly improve an examination by illustrating how well each item in your examination is performing. Reviewing your exam items with statistics in hand is made easier by using minimum values, or a range of values, into which good performing items fall. For example, in licensure and certification testing, it is common to use a threshold of 0.20 for evaluating the item discrimination. Thus, items with discrimination values equaling or exceeding 0.20 would be considered adequate. Items whose discrimination values fall below 0.20 may need to be edited or retired from the exam and replaced with a better item. Items with discrimination values that are negative are particularly poor performers and should not be kept in your examination as is.

Item discrimination can be used in tandem with the average exam score for candidates picking each answer choice. If the discrimination is low or negative, looking at these averages can show where the problem may lie. For example, there might be one distractor in particular that is attracting high scoring candidates, causing the item discrimination value to be negative. Editing or replacing this distractor, so that it does not 'fool' your high scoring candidates, will likely fix the problem. Using this process of identifying item discrimination problems, along with diagnosing possible solutions on all such items in an examination, can do a great deal to improve the reliability and validity of your examination. Using another interpretation of item discrimination can illustrate this point. That alternate interpretation says that items with very low or negative discrimination are inconsistent with the other items in the examination; they may be poorly written items or they may be measuring a different skill or knowledge area than the other items on the examination. By editing or removing items that measure a different skill or knowledge area, we are left with an examination composed of items that are, in a sense, working together to measure a common skill or knowledge area. Such examinations typically have higher reliability. A more reliable examination has the potential to be more highly related to other measures that contribute to the validity of your examination.

Monitoring the percent correct for each examination item provides another statistical 'window' into how an examination is performing with a candidate population. Here again, examination programs can choose an acceptable range into which items should fall. It is essential that exam questions not be too easy or too difficult. In the extreme, having many questions that 100% of the candidates answer correctly does not reveal who has the right level of expertise to be licensed because such items award points to the strongest and weakest candidates. Also, if an examination has many items with percent correct in the 0% to 50% range, the examination may be too difficult for the candidate population. Such an examination may confer passing status on an unacceptably low proportion of candidates. Establishing

an acceptable percent correct range for your examination should be done with the advice of a psychometrician. Such an individual can also advise you on setting the passing score and review the resulting passing rates when the examination is administered.

When and how often item statistics can be calculated, depends upon the number of candidates tested each year. Larger programs, such as those testing 20,000 candidates per year can field test items before using them to score candidates. With this model, each candidate answers a certain number of new items that have never been tried out. These items do not contribute to the candidates' score or pass/fail status. After being field tested, item statistics are calculated. Additionally, the items are reviewed alongside the statistics. The next step is to make a decision as to whether to use the new item, revise it and field test again, or retire the new item. Larger programs also routinely compute item statistics for items that have already been field tested and are now being used operationally. Item statistics for such questions are monitored to make sure that the statistical measures for each item are still in the acceptable range and have not changed since the last time the item appeared in an operational form of the examination.

Small testing programs with as few as 20 or 30 candidates per year can also make good use of item statistics. Depending on several factors, it may take more than one year to accumulate enough candidate responses to calculate item statistics. During this time, it may happen that poorly performing items are used in operational forms of examinations because they cannot be identified due to limited resources, including the number of candidates. It is still important to compute item statistics as soon as possible. Such an analysis can show which items are performing well and which items need to be revised, retired, or reused based on statistical criteria and expert judgment. The benefit ultimately is the same as that for larger examination programs: improved test reliability and validity. Using these statistics are invaluable if a testing program is called upon to show that their examination, and the policies & procedures used to produce, deliver, and score it are legally defensible.





# Client's Corner

## ***South Carolina Department of Labor, Licensing, and Regulation Board of Cosmetology and Barber Examiners***

Back in 2004, and beginning only with the administration of the National-Interstate Council of State Boards of Cosmetology (NIC) practical examination for both cosmetology and barbering, PCS administered its first examinations in the State of South Carolina, commencing a partnership with the State of South Carolina's Department of Labor, Licensing and Regulation (LLR) Board of Cosmetology and Barber Examiners that has continued to blossom since inception. Both PCS and LLR, over the years, have helped thousands of candidates sit for their cosmetology and barber examinations. PCS currently administers examinations for approximately 3,329 candidates on a yearly basis in South Carolina and processes all initial and re-exam applications, eligibility, and ADA candidate applications for both boards.

All PCS practical examiners are required to go through NIC's annual two-day examiner training program that must be passed by all examiners before being allowed to administer NIC examinations. PCS' Chief Examiner in South Carolina, Delores Rush, has been successfully leading all test administrations for the past seven years. For each board, these practical administrations take place each month in South Carolina's capital city of Columbia. Additionally, once a year in the spring, PCS conducts special test administrations to accommodate the vocational school students from all over the state.

Quincy Neuble, designated PCS coordinator for LLR, keeps up-to-date on the LLR's latest rules and regulations changes, guides candidates through the application process, and provides superior customer service to all the LLR's cosmetology and barbering constituency she has contact with. She is also responsible for ensuring that all contractual agreements are being met. Quincy has been in her current role since 2008 and is backed up at all times by the entire PCS cosmetology and barbering department which consists of Francine Rananto, Matt Levy, Susan Young, Melanie Proctor and Jamey Russell. All coordinators are cross trained in each state's program to ensure that all cosmetology and barbering candidates receive the highest customer service possible. Melanie Thompson, South Carolina's Board Chair, states: "The past couple of years have brought many changes to the South Carolina Board of Cosmetology. One thing that has remained consistent is the relationship we have with PCS. Their professionalism, commitment to our industry, and eagerness to help us resolve any problem, big or small, has proven to be invaluable."

PCS looks forward to serving LLR and its constituency for years to come.



SC Cosmetology Board Chair,  
Melanie Thompson



## **Behind the Scenes of PCS**

### **Xiomara Gilmore**

Xiomara, originally from San Pedro Sula, Honduras, joined the PCS team in 2007 as a bilingual Customer Service Representative. She has worked in the customer service industry for the past twelve years. Currently, she serves as the Puerto Rico Multiboard Coordinator. In her spare time, she enjoys salsa dancing, cooking, being outside, traveling, and experiencing new things.



### **Quincy Neuble**

Quincy, a Nashville, Tennessee native, joined the PCS team in 2008. She received her formal education from Tennessee State University. Currently, Quincy is the cosmetology and barbering coordinator for the states of South Carolina, Alabama, and New Mexico. When Quincy is not in the office, she enjoys spending time with her 2 year old daughter, exploring the great outdoors, being adventurous, traveling and trying new things.



# ReMarks

*from the desk of Mark Setash*

I travel fairly frequently and, over the past year, I've had the good fortune of flying Southwest Airlines the majority of the time. If you've ever done the same, and particularly if you have flown with Southwest several times, you expect and quickly become accustomed to the "schtick" of the flight attendants. While mostly humorous, it also speaks to the spirit of the airline, instilled by Colleen Barrett, Southwest's President Emeritus. If you visit their website, you'll see their mission is "dedication to the highest quality of Customer Service delivered with a sense of warmth, friendliness, individual pride, and Company Spirit." A fine mission, in my opinion, and one we seek to emulate at PCS.

Furthermore, Southwest promotes their principle of LUV. It's also their stock symbol. So what is this LUV? Southwest explains - "Southwest has been in LUV with our Customers from the very beginning. Therefore, it's fitting that we began service to San Antonio and Houston from Love Field in Dallas on June 18, 1971. As our Company and Customers grew, our LUV grew too! With the prettiest Flight Attendants serving "Love

Bites" on our planes, and determined Employees issuing tickets from our "Love Machines," we changed the face of the airline industry throughout the 1970s. Over the ensuing years, our LUV has spread from coast to coast and border to border thanks to our hardworking Employees and their LUV for Customer Service."

Widely recognized as a fine organization and a highly desirable place to work, I think there are lessons for all of us in what Southwest does. On my most recent flight, even though I routinely "tune-out" the preflight banter, the flight attendant stood just as we approached the runway to announce, "we're next to depart, please check your seat belts one more time as the pilot's going to try something different." Heads popped-up all over the plane and the smile on his face reassured them, and me, that he was only gently teasing and using humor to get attention. I laughed to myself as we began a perfect takeoff and ascent. My hat's off to you Southwest. If imitation is really the sincerest form of flattery, all of us at PCS hope we can imitate some of what you do.

*Mark Setash*

**President and CEO**

