

Accuracy of a Testing Result

Part 1

Introduction

NBRC credentialing examinations are designed to provide a sufficiently accurate reflection of a candidate's mastery of necessary knowledge. The decision to award a credential is based on the total score obtained by a candidate compared to the cut score. A candidate who has a fail result should focus on the gap between the score obtained and the cut score. Both of these scores are the number of correct responses, which is referred to as a raw score.

Questions and Answers

Q. Why aren't sub-scores reported?

A. If subscores were reported, each would be based on a small number of item responses yielding a less accurate score than the total score. A candidate who fails an examination plus the patients who receive care are best served when the candidate elevates his or her overall ability rather than trying to identify weak areas from less accurate sub-scores. Remediation is best focused on becoming a stronger practitioner, not speculating where gaps are so they can be filled in.

Q. If a candidate was to take an examination again without remediating to elevate his or her knowledge, how likely is it the candidate would subsequently pass?

A. The likelihood is low even for those who only fail by a point or two. The larger the gap between a candidate's test score and the cut score, the less likely a reversal will occur without remediation.

Q. What error factors influence candidate scores?

A. For any examination, a score can be influenced by external factors. Some factors negatively influence scores such as a candidate's anxiety level, distractions, fatigue, and incomplete reading comprehension. These factors can introduce errors decreasing accuracy of some scores as reflections of knowledge mastery. Factors like prior knowledge of examination content and cheating while taking a test erroneously increase some scores.

Q. What does the NBRC do to minimize the influence of error on test scores?

A. Test content is approved by a diverse expert panel. The correct response for each item is verified through data summarized from candidates' responses, which is evidence-based item validation. A continuous quality improvement system identifies items that should be refined. Tests are administered under standard conditions that minimize distractions while securing content from theft. Candidates who take tests are warned by Terms and Conditions statements against disclosing information about examination content. With these systems in place, error influences are minimal so candidates typically achieve the scores they deserve.

Q. Is there an analogy that might help me understand potential error in test scores?

A. Imagine the test as a mountain to be climbed. Preparation plus effort is necessary to get any distance up the mountain and only the elite few will get to the top (achieve the highest score). There are safe thresholds in the upper mountain section that climbers (candidates) can reach. The lowest of these safe thresholds is like the cut score on the test.



To better understand potential error, imagine some climbers get extra weight added to their backpacks as they climb. Some climbers receive occasional outside help at points along the way. The negative influences from extra weight are added unexpectedly. The positive influences may unexpectedly occur like when someone else helps without being asked. Positive influences can occur on purpose when a climber asks for and receives help while climbing. Outside influences contribute to some individuals climbing higher and others climbing lower than their preparation and effort otherwise would have allowed.

Error is the difference between the height that should be reached and the height that is reached. Some climb so close to the expected height that error is not worth discussing. Even among those whose climb is influenced by outside forces, most of what explains the height they reach is their preparation and effort. Hence, anyone who fails to reach the lowest safe threshold is advised to wait and train (remediate) before trying again. The goal is to climb to a safe level so guessing whether one's skills with ropes, ice tools, or survival techniques let one down distracts from the preparation needed to become a generally strong climber and succeed the next time.

Q. What is the Standard Error of Measurement (SEM)?

A. The technical definition is that the SEM is the standard deviation of measurement errors within test scores from a group. Less formally, the SEM describes the degree to which candidates' test scores occur above and below true scores that would result when no error occurred. SEM values decrease as test scores become more accurate. An SEM value can be used to describe a range of scores in which a typical candidate's true score is located.

Q. How can SEM be useful to candidates who have failed?

A. Consider an example in which the SEM is 5.04 at the cut score for a typical set of scores from the Adult Critical Care Specialty Examination. A 99% confidence interval can be calculated by multiplying the SEM (5.04) by 2.58 to produce 13. Practically all (99%) scores influenced by error fall inside a range that is 13 points above and below the cut score. Outside the low boundary of this range, a candidate certainly lacks the ability to pass. Determining whether a candidate's test score falls inside or outside the low boundary can help a candidate decide whether to make another attempt and how much effort to put into remediation before the next attempt.

Because the maximum and cut scores are constant within an NBRC multiple-choice examination over a multi-year period, the low boundary of the error range has been calculated in Table 1 for candidates who have failed. This convenience is unavailable for the Clinical Simulation Examination since different forms have different maximum and cut scores, so Table 1 offers an estimated lower error boundary that can be calculated from the cut score.

Table 1. Guidance Table for Candidates with a Fail Result

Examination	Cut Score*	If a candidate's score is less than the threshold below, the fail result is outside the 99% margin of error from the cut score.
Adult Critical Care Specialty	99	86
Asthma Educator-Certified	105	92
Clinical Simulation	~72% of max raw	raw cut score from one's score report – 32
Neonatal / Pediatric Specialty	77	65
Pulmonary Function Technology	Low 61	Low 50
	High 74	High 63
Therapist Multiple-Choice	Low 86	Low 73
	High 92	High 79
Sleep Disorders Specialty	97	84

*Note- Cut score is the number of correctly answered items required to pass the examination.

Q. How long should a failing candidate wait before another attempt?

A. A candidate who sees that the score he or she reached was within the lower boundary of error could be encouraged to only wait a brief time before another attempt at the test. However, there is no guarantee the first score was negatively influenced by error so anyone who fails should prepare more before another attempt.

Q. What should I expect after I pass the test?

A. The mountain remains as a useful metaphor for what happens in the years after passing a test. First, effort is required to remain at the safe level because the requirements will change that define what safe means. Second, some of the abilities one built up to make the attempt will decline. Those who do nothing to maintain their abilities will find themselves below the everchanging safe threshold as a result.

Changes in technology and evidence-based practices plus new drugs and techniques alter the knowledge base associated with safe practice as time moves forward. If preparation and effort stand still, ground will be lost on the competency front. Hence, expect to document that competency has been maintained during credential renewal.

Summary

The best strategy after failing is to increase general preparation so the next attempt will yield a higher performance. Error in some test scores is unavoidable, but this error is small compared to what a candidate brings to a test. A candidate who fails will benefit by identifying where his or her test score falls in comparison to the low boundary described in Table 1 before deciding whether to make another attempt. The farther a candidate's raw score is from the cut score, the more remediation should be done before another attempt is made. Even after passing a test, expect to continue learning to maintain your practice at a safe level.