

# 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. A candidate who has a fail result should focus on the gap between the total score obtained and the score required to pass the examination. Throughout this document, the phrase **cut score** describes the score required to pass.

### Questions and Answers

#### **Q. Why aren't sub-scores reported?**

**A.** If subscores were reported, each would be based on a limited number of item responses yielding a less accurate value than the score based on the whole examination. A candidate who fails an examination and the patients who receive care are best served when the candidate elevates his or her overall ability rather than identifies what may appear to be weak areas from less accurate sub-scores. Remediation is best focused on becoming a stronger practitioner, not guessing where gaps are so they can be filled in.

#### **Q. If a candidate was to take an examination again with no remedial effort to elevate his or her knowledge, how likely is it that the candidate would subsequently pass?**

**A.** The likelihood a candidate will pass without remediation 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 and the more intense the remediation will have to be.

#### **Q. What error factors influence candidate scores?**

**A.** For any examination, not just NBRC examinations, a score can be influenced by external factors. Some factors that negatively influence scores include a candidate's anxiety level, distractions, fatigue, concentration, and confusion while reading. Such factors can introduce errors that decrease some scores as reflections of the extent to which knowledge is mastered. Factors like prior knowledge of examination content and cheating while taking a test erroneously increase some scores.

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

**A.** Test content is the product of approval from a diverse panel of experts, not just one or two people. Despite agreement among panel members, validation of the correct response for each item is verified through data summaries of candidates' responses. In other words, the NBRC deploys evidence-based item validation. A continuous quality improvement system identifies items that should be refined. Tests are administered under standard conditions that minimize distractions and secure content. 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 deserved to achieve.

**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. The first things to know are that it takes preparation plus effort to get any distance up the mountain and no climber is really expected to get to the top (no candidate has recently achieved a perfect 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.



Some climbers get extra weight added to their backpacks as they climb. Some climbers receive occasional outside help at points along their climb. The negative influences like the extra weight are added unexpectedly. The positive influences may unexpectedly occur like when someone else helps without being asked. Positive influences 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 have been reached and the height that was reached for each person whose performance was influenced. Others 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 before trying again. The goal is to climb the whole mountain 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.** For a technical definition, 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 are scattered above and below true scores that would result when no error occurred. SEM values decrease as test scores become more accurate. An SEM value is commonly 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 intended to illustrate how SEM can be useful. The SEM is 5.17 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 (5.17) by 2.58 to produce 13.3. Practically all (99%) scores influenced by error fall inside a range that is 13.3 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 lower error boundary can help a candidate decide whether to make another attempt and how much effort to put into remedial learning 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 a candidate's cut score.

*Table 1. Guidance Table for Candidates with a Fail Result*

<b>Examination</b>	<b>If candidate's score is less than the value below, a fail result is outside the margin of error.</b>
Adult Critical Care Specialty	86.7
Clinical Simulation	cut score from a candidate's score report – 33.0
Neonatal / Pediatric Specialty	65.4
Pulmonary Function Technology	Low cut score 48.2 High cut score 62.3
Therapist Multiple-Choice	Low cut score 73.8 High cut score 80.2
Sleep Disorders Specialty	82.6

**Q. How long should a failing candidate (climber) wait before another attempt?**

**A.** A candidate who sees that the score (height) he or she reached was within the lower error boundary of the safe threshold might be encouraged to only wait a brief time before another attempt at the test (mountain). However, only those inside the error boundary who are very close to the cut score (lowest safe threshold) should think about trying the test (mountain) again quickly because most of what explained their performances is within them.

**Q. What should I expect after I reach the safe threshold (pass the test)?**

**A.** The mountain continues as a useful metaphor for what happens in the years after a successful climb. First, effort is required to remain at the safe level because the requirements will change. Second, some of the abilities one builds 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.

Bringing this discussion back to practitioner competency, changes in modern 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.

**Summary**

The best strategy after failing a test is to increase preparation in a general way so the effort given during the next attempt will yield a higher performance. Error in some test scores is unavoidable, but this error is small in its influence compared to the abilities within candidates. Candidates who fail a test would benefit by identifying where their test scores fell in comparison to the low error boundary before deciding whether to make another attempt. The farther a candidate's test score is from the cut score, the more preparation should be done before another attempt is made. Even after passing a test, expect to continue learning to remain in the safe competency zone.