Appendix A: Sample Candidate Explanation

Why WOCNCB Uses Scaled Scores

For each WOCNCB examination, test takers receive different sets of test items (called test forms). We do this to ensure that the content of the test is up-to-date and to avoid overexposure of test items. Test development agencies work hard to ensure that all the test forms are comparable in content and difficulty so that test takers are treated equally. This equality is achieved through a two-step process.

First, the different versions of an examination are built to standardized test specifications. Using standardized test specifications ensures that all new versions of the test are comparable in content coverage, that test content is weighted in accordance with the requirements for competent performance in the profession, and that all forms of the examinations adhere to standards for content validity.

Although all forms are developed from the same examination specifications, the difficulty level of the forms may vary slightly due to the different questions used. In order to ensure that test takers are neither penalized nor rewarded for variations in examination difficulty, we equate the forms to adjust for differences in form difficulty.

Because test forms vary in difficulty, the number of items required to pass each test form also can differ. When a test form is first developed, a standard-setting process is undertaken to establish the passing standard, or cut score. For later forms, procedures appropriate to the given examination program are used to equalize the different versions and the scores test takers achieve on the examination. To ensure that candidates can compare apples to apples, a scaled score is provided for reporting scores to candidates and other stakeholders. Scaling allows the passing score to remain constant for all examination forms even as the number of correct answers necessary for passing may vary from form to form. Scaled scores are a mathematical transformation from one scale to another. They maintain the overall performance of candidates but allow comparisons between individuals who take different examinations.

For the WOCNCB examinations, the cut score for any form is established as equivalent to 500 on the score scale. The lowest possible score on the form (typically set to 0 or some value less than chance responding) is set to 200, and the highest score is scaled to 800. Raw score between these numbers are then transformed to the appropriate point on the new scale.

Appendix B: Frequently Asked Questions Document

Scaled scores provide a useful measurement tool for many assessment programs. They are used in numerous national testing programs, including the ACT® and SAT® examinations, which are typically part of the admissions process for colleges and universities. They are particularly useful at providing the basis for long-term, meaningful comparisons of examination results across different test administrations.

Does everyone take the same test?

To maintain the validity, security, and relevance of examinations, old, outdated, or poorly performing items are replaced with new items. Every time this process is performed, the result is a different version of the examination (called a *test form*). As a result, over time, candidates may receive different forms of the examination.

What is a passing score?

The passing score is the number of items candidates must answer correctly to pass the examination. The passing score is established by the credentialing agency based on the results of a criterion-referenced process. During this process, a committee of subject matter experts discuss the minimum level of competency required to pass the examination and participate in an established psychometric standard-setting method, such as the Angoff method.

Are all the forms the same difficulty?

Certification agency staff and psychometric consultants work hard to make sure that the exams are extremely similar in overall difficulty, both within a year and when compared to previous years. However, each form of an exam is not identically difficult. Some forms require a greater number of correct answers to pass than others.

For example, two versions of the same exam (Form A and Form B) have the same number of items, but the items on Form A happen to be slightly more difficult than the items on Form B. As a result, candidates would need to answer a higher percentage of items correctly on Form B than Form A.

If raw scores or percent correct are used, then users of the exam reports are confronted with different passing scores from one form of the exam to another. This can (and does) create confusion among candidates. To resolve any confusion about different passing scores on different versions of the same exam, scaled scores are used.

What are scaled scores?

Scaled scores are raw scores transformed by a numerical procedure. By using scaled scores, certification agencies make sure that they can provide reports to candidates that are consistent and without concern for test security or confusion about the scores.

For every possible raw score on a test form, there is a corresponding scaled score. When multiple forms of a test are used, or when results are compared from year to year, scaled scores are needed to adjust for possible differences in test form length or difficulty.

This transformation is similar to converting from pounds to kilograms. The weight of the object has not changed; only the units being reported have changed.

How is the scaled score computed?

To calculate a scaled score, the raw score required to pass is first set equal to the scaled score passing point (e.g., 500). An analogous situation is with temperature: 0 degrees and 32 degrees both represent freezing on different temperature scales. For a scale with the passing point set at 500, raw scores below the passing point are converted in linear fashion to scaled scores below 500; those above the passing point are similarly converted to scaled scores above 500. Thus, the passing standard on any test form of the examination is always reported as the same scaled score.

Why are scaled scores used?

Scaled scores provide a consistent scale of measurement so that the same scaled score represents the same level of knowledge from one test administration to the next. Testing programs often use multiple test forms to limit exposure of test questions. While the different forms are built to the same test specifications and are designed to be similar in difficulty, rarely are they exactly equal in the level of difficulty. If two candidates take different forms of different levels of difficulty but get the same number of questions correct, the candidate with the more difficult form demonstrated a higher level of knowledge. A scaled score ensures comparability in reporting across test forms.

Why not report raw scores and percent correct?

Scaled scores are intended to make scores more meaningful by defining a scale of measurement that is not tied to a particular form of a test. Percent correct scores are just another version of raw scores and thus do not resolve the issue of comparability of scores from different test forms. Scaled scores help resolve confusion among users of the score reports when there are changes in the testing program, especially for test length and form difficulty.

A major benefit of scaled scores for test takers is that they provide a way for the test takers to compare their level of performance from one administration of an examination to another. If they fail the first time, they want to know by how much. If they fail a second time, they would like to be able to compare the score to their first score. A reporting scale that remains constant across test forms enables candidates to make these comparisons.

Scaled scores cannot be used to compare across test programs.

Do scaled scores change who passes or fails?

Scaled scores do **NOT** affect individual candidate pass or fail decisions. The decision is always made by comparing the number of items answered correctly to the number of items required to pass the test form that was established using the criterion-referenced process.