

Reliability of the NCLEX[®] examinations

The reliability of the NCLEX examination is assessed via a decision consistency statistic. This statistic is used instead of a traditional reliability statistic such as Cronbach's alpha because it captures the reliability of dichotomous pass/fail decisions rather than the reliability of continuous scores or ability estimates.

The decision consistency statistic incorporates a candidate's ability estimate and standard error with normal theory to obtain two probabilities: the probability that the candidate's "true" ability (as opposed to the estimate of their ability) is above passing and the probability that their "true" ability is below passing.

$$P(\text{Pass}|\hat{\Theta}_j) = \Phi\left(\frac{\hat{\Theta}_j - \Theta_c}{SE_{\hat{\Theta}_j}}\right)$$

$$P(\text{Fail}|\hat{\Theta}_j) = 1 - P(\text{Pass}|\hat{\Theta}_j)$$

Θ_c is the passing standard (pass/fail cutscore), Θ_j is the ability estimate of candidate j , SE_{Θ_j} is the standard error of Θ_j , and $\Phi(Z)$ is the cumulative standard normal distribution evaluated at Z . Once these probabilities are obtained for each candidate, the probability that each candidate would obtain a consistent decision over two NCLEX examinations is calculated.

$$P_j(\text{Consistent Decision}) = P(\text{Fail}|\hat{\Theta}_j)^2 + P(\text{Pass}|\hat{\Theta}_j)^2$$

The average of this probability over all candidates is the estimated decision consistency of the NCLEX examination. The decision consistency of the NCLEX examination is psychometrically sound normally running between .87 and .92. If you would like further information on the decision consistency statistic please see

Lunz, M.E. & Bergstrom, B.A. (1991). Comparability of decisions for computer adaptive and written examinations. Journal of Allied Health, 20, 15-23.

Validity of the NCLEX examinations

There are several different aspects of validity: content validity, face validity, construct validity, predictive validity, and scoring (passing standard) validity. National Council procedures ensure that the NCLEX examination is valid with respect to these aspects of validity.

Content Validity: *Does the NCLEX examination content cover the scope of entry-level nursing practice?*

In order to cover the entire range of nursing content, a vast number of examination questions are written. These questions are written by panels of volunteers from around the country with backgrounds in covering the entire spectrum of different specialties and practice settings. By having this diversity in the item writers, the examination questions cover the entire domain of entry-level nursing practice.

However, it would be impossible for each NCLEX examination to cover the entire domain of entry-level nursing practice. If this were so, then each examination would need to include all of the thousands of questions written by the panels of volunteers. Obviously, this is not feasible. This is where sampling validity comes into play.

Sampling Validity: *What sample of questions should an NCLEX examination contain in order to make valid inferences about candidates' knowledge of the entire scope of entry-level nursing practice?*

In order to determine what sample of questions should be included in each NCLEX examination, an on going evaluation of the scope of entry-level nursing practice is conducted. This evaluation is called a job analysis. Every three years, the results of the on going job analysis are compiled. From these results, the NCLEX[®] Test Plan is derived. The test plan provides a set of content categories that define nursing actions and competencies across all settings for all clients. Every NCLEX question is classified into one of these content categories. When a candidate takes the NCLEX examination, their test questions are chosen such that there are some questions from each content category. Since some content categories are more fundamental to the practice of nursing than others, it is necessary to prescribe the percentage of questions chosen from each content category. Again, the job analysis provides the information that is necessary to derive these percentages. These percentages are provided in the NCLEX Test Plan.

Face Validity: *Even with the detail of the NCLEX Test Plan, there are still hundreds of questions in each content category. How can the National Council ensure that each NCLEX examination contains questions on a wide variety of medications, procedures, conditions, diseases, clients, settings etc.?*

Although the Test Plan ensures that the content of nursing is sampled adequately, it is far from a perfect sampling of the nursing domain. No matter how one slices the domain of nursing, there is always going to be similar content across content categories. For example, several items may be about Radiation Therapy. These different questions may investigate side effects (D10), emotional support for the client (C6), teach about the procedure (D10), handling of radioactive materials (A2), on infinitum. In theory, someone could receive an examination which met all of the specifications outlined in the Test Plan and contain solely Radiation Therapy questions. Well, not exactly. That's because the NC conducts a procedure to ensure that this does not happen. This procedure is called a Face Validity review. During this process, real and simulated examinations are read by experienced test developers to ensure that the balance and juxtaposition of content is on face, representative of the domain of nursing.

Construct Validity:

“To the extent that a variable is abstract and latent rather than concrete and observable, it is called a “construct.” Such a variable is literally something that scientists “construct” and which does not exist as an observable dimension of behavior.” - Nunnally, pp 85.

If entry-level nursing competency was concrete and observable, then it would be easy to measure. However, entry-level nursing competency is latent and abstract. Therefore, the National Council has “constructed” the NCLEX examination scale to produce a measure of an examinee’s entry-level nursing competency.

The National Council uses measurement theory via Rasch measurement theory (Wright and Stone, 1977) to construct the NCLEX examination scale. There is an abundance of research that certifies this theory’s capability to produce valid measures of a latent construct such as entry-level nursing competency.

Scoring Validity: *How does the National Council ensure that NCLEX scores are ?*

During examination administration, each examinee receives at least 15 ‘tryout’ items. These items are not counted towards an examinee’s score. Rather, examinee performance on these items is tracked for all examinees. This allows the National Council to determine the exact difficulty of each item. Using this information and Rasch measurement theory, the National Council can accurately calculate a measure of each examinee’s ability.

Pass/Fail Decision Validity: *How does the National Council ensure that NCLEX passing standard truly reflects the minimum competency needed to practice safe and effective entry-level practice?*

The minimum level of competency that an examinee must attain in order to pass the NCLEX examination is investigated thoroughly on a triennial basis. This level of competency is called the passing standard. The passing standard is established by the National Council’s Board of Directors

after they have reviewed extensive information gathered systematically from several sources including a panel of nursing experts. Through this process, the passing standard maintains currency and validity.