

## GGSON Undergraduate Program Handbook

### SON OP 30.275 - Test Revision

- PURPOSE** Texas Board of Nursing Rules of the Nurse Practice Act provide that evaluation of methods and tools used to measure students' cognitive, affective, and psychomotor achievement shall be developed and implemented using sound educational principles.
- REVIEW** This handbook shall be reviewed by the Associate Dean by September 1<sup>st</sup> of even numbered years in collaboration with the Curriculum Committee. Final approval of the GGSON Undergraduate Program Handbook is granted by the Dean.

### **POLICY/PROCEDURE**

This policy provides guidelines for faculty to follow in evaluating the reliability and validity of objective tests.

Faculty members are responsible for reviewing each test item using provided computerized statistical analysis. Course faculty should consider all of the following factors when making decisions about the need for revision of individual items. Each question should be reviewed in order to make the decision whether to keep a question, discard a question, or attempt to improve the question. Course faculty are responsible for reviewing the computer analysis regarding the following:

*Reliability:* a reliability index between 0.7 and 1.0 is preferred. Tests with a lower reliability index need to be evaluated for revision (i.e. test length, item types, item quality, and conditions of administration). On Computer Services reports, this is the KR coefficient. This statistic is not available on the A Plus software.

*Test Mean:* a mean of 75 or better is preferred. Tests with a lower mean need to be evaluated for revision.

*Item Analysis:* item analysis can provide information about the difficulty of the item, the discriminatory power of the item and information about the response pattern of each option.

### Guidelines to Assist Faculty Members in Test Analysis

- Index of difficulty: the majority of questions should reflect a 0.50 or better level of difficulty. Any item falling below 50% should be reviewed and revised if indicated. The higher the index of difficulty, the easier the question.
- Discriminatory Power: discriminatory power reflects how an item discriminated between the upper and lower half of the group taking the test. The value of a test item is judged by its ability to discriminate. Such discrimination determines a test's reliability. The highest discriminatory power for an item is 1.00. A power of 1.00 is achieved when all of the students who score in the upper half of the exam select the correct answer and none of the students who scored in the lower half of the exam select the correct answer. An item with negative discrimination is selected by more students who scored in the lower half of the exam. Items with a discriminatory power of .30 to .50 are considered to be excellent items. A discrimination power of .30 is preferred. Items that reflect an extremely high or low discrimination power and items that have a negative discrimination should be reviewed and revised as indicated. On Computer Services reports, this statistic is known as the Point Biserial.

- Response Distribution: all test items should be assessed to determine the pattern of student responses. If only one option was selected, the item was too easy. If only two options were selected, the multiple choice question with four options has become a true false question with only two options. Faculty should use discretion in revising items too easy or items that had only two responses selected.
- Student Feedback: student feedback provided during test review is utilized at faculty discretion to revise test items.

Faculty cannot “curve” all students’ scores on an exam to raise the mean exam score.

Other recommendations for test revision:

A test item is nullified by giving credit for all choices. Nullifying an item is not the same as “throwing out” an item. Nullifying does not decrease the total number of items.

A negative discrimination index alone is not sufficient reason to nullify an item. Such items must also be reviewed by faculty responsible for teaching the content to validate the necessity of nullifying the item.

An item’s discrimination data from a historical perspective are usually more important than the most recent discrimination data because they reflect the test item’s worth over time and are based on a higher number of responses to the item.

Overall analysis should take into consideration the content and measurement of course and module objectives presented in the syllabus.

Morrison, S., Smith, P., & Britt, R. (1996). Critical thinking and test item writing. Livingston, TX: Century II Printing.