Operating Policy and Procedure

SON OP 30.093 – Inter-rater Reliability for Clinical Policy

PURPOSE
The purpose of the School of Nursing Inter-rater Reliability for Clinical Operating Policy and Procedure (SON OP) is to have guidelines/policies for performing inter-rater reliability reviews of evaluators in clinical evaluation competency as required by the Board of Nurse Examiners for the State of Texas.

REVIEW
This OP will be reviewed biennially by September 1 of each even numbered year (ENY) by the applicable Associate Dean/Department Chair, with recommendations for revisions forwarded to the Dean of the School of Nursing.

POLICY/PROCEDURE

The faculty of each nursing program will use a uniform format to grade clinical performance and skills lab competency. Measuring instruments (grading rubrics) will be developed for clinical activities utilizing weighted criteria appropriate to the course level. This policy ensures accuracy and consistency of grading of students in a clinical course taught by more than one faculty member.

At the beginning of each semester in which new grading rubrics are used, all course faculty will complete a blind review of student performance in clinical (based on a written scenario depicting student performance) or the skills lab (based on a recorded simulation) utilizing the appropriate grading rubric. Aggregate inter-rater reliability information is collected at the end of each clinical course via the End of Course Summary.

Calculating Inter-rater Reliability

1. Calculate the mean of all faculty grades.
2. Determine the number of agreements and disagreements. Using the grading scale of 100, an agreement is considered to be within the range of +5 to –5 points of the mean. Disagreements are those scores which fall outside the range of +5 to –5 points.
3. Estimate the inter-rater reliability by using the following formula (Woods & Catanzaro, 1988):
   \[
   \text{Inter-rater reliability} = \frac{\text{# of agreements}}{\text{Total # agreements} \& \text{disagreements}} \times 100
   \]
   An 85% inter-rater agreement between faculty is considered acceptable and reliable.

Failure to Assure Inter-rater Reliability

If the inter-rater agreement is less than 85%, the faculty will evaluate where discrepancies in grading exist with subsequent changes to the rubric or assignment followed by re-evaluation. Grading should reflect agreement among faculty after reviewing discrepancies.

In the Event of a Failing Grade

If a failing grade is earned on a clinical component of a course, the faculty responsible for grading will have another faculty who is teaching in the course* evaluate the student as outlined in the course syllabus. It is recommended that two faculty members be present for evaluation when high fidelity simulation is a component of the evaluation and that the simulation is recorded. The repeated evaluation should be graded using an unmarked set of grading criteria identical to the criteria used for the original grade.
*In cases where there is only one faculty member teaching the course, the second faculty member should be someone who is familiar with the grading rubric(s).