26349 - SIMULATION PERFORMANCE CHECKLIST GENERATION USING THE DELPHI TECHNIQUE

Pamela Morgan MD, CCFP, FRCPC, Jenny Lam-McCulloch, MSc; Jodi Herold, PhD; Jordan Tarshis, Sunnybrook & Women's College Health Sciences Centre, Toronto, ON, Canada

INTRODUCTION: Performance assessment using high fidelity simulation is problematic, largely due to the difficulty in developing valid and reliable evaluation tools. Checklists are easy to use and reliable but appropriate content is essential for validity. The Delphi technique is a consensus based content generation method used for multiple purposes such as policy development, best-evidence practice guidelines and competency assessments.(1) It involves item generation/endorsement, iteration and feedback, and does not require face to face meetings.

METHODS: After REB approval, 5 university-based anesthesiologists were asked to participate in the development of a checklist for two scenarios to be used for performance assessment of practicing anesthesiologists using high fidelity simulation. Patient information was entered into a Microsoft Excel spreadsheet under the following categories: preoperative assessment, machine/equipment, drugs/medications, monitors, induction, maintenance, critical event #1, and critical event #2. The templates for both scenarios were emailed to the 5 anesthesiologists who, on the first iteration, were asked to generate items they would expect to be performed in each category. Data from the 5 participants was collated anonymously and returned to them. Participants were asked whether each item should be kept, discarded and whether further items should be added. An a priori decision was made to delete items endorsed by 20% of participants. This process of collection, collation and reevaluation was repeated until consensus was reached. Four independent raters used the checklist to assess 3 subjects managing the 2 simulation scenarios. Inter-rater reliability was assessed using KR-20 and repeated measures analysis of variance (ANOVA) was used to determine differences in difficulty between scenarios.

RESULTS: On the first iteration, 78 items were generated for scenario 1 and 83 for scenario 2. After 3 iterations, no further items were added or deleted. For critical events that involved Advanced Cardiac Life Support, current ACLS guidelines were listed. The final checklists included 100 items for scenario 1 and 98 items for scenario 2. The mean interrater reliability was KR-20=0.921 for scenario 1 and 0.903 for scenario 2. Repeated measures ANOVA with scenario as the within-subjects factor, and rater as between subjects factor was $F=3.302$, $p=.107$ indicating no statistically significant difference between scenario difficulty.

DISCUSSION: The Delphi technique can be very useful to generate consensus based evaluation tools. In so doing, content and face validity are high compared to subjective evaluative tools. Since there was no difference in scenario difficulty, these scenarios can be used to determine the effect of educational interventions on performance. The high inter-rater reliability mirrors previous work in performance evaluation using simulation.