471865 - EXPERT VERSUS NON-EXPERT RATERS IN SIMULATION PERFORMANCE EVALUATION

Pamela Morgan, MD, CCFP, FRCPC¹, Roderick Cheung, MD candidate², Vicki LeBlanc, PhD³, Michele Haley, BA¹

1. Department of Anesthesia, Women's College Hospital, Toronto, ON, Canada

2. University of Toronto, Toronto, ON, Canada

3. The Wilson Centre, Toronto, ON, Canada

Introduction: One impediment in using high-fidelity patient simulation is the high cost needed to arrange for an expert to rate the performance of participants.(1) The purpose of this study was to determine whether there is a difference between the use of expert and non-expert raters when evaluating participants of high-fidelity patient simulation using a dichotomously-scored checklist (DSC), a human factors rating scale (HFRS), and a global rating scale (GRS).

Methods: After Research Ethics Board approval, practicing anesthesiologists were invited to participate. Each subject managed two scenarios 6-9 months apart. One of three academic anesthesiologists with simulation expertise facilitated all sessions and evaluated each subject using the DSC, HFRS, and GRS. Two non-expert raters who received training in the use of the 3 performance measures also scored videotapes of each participant using the same tools. Intraclass correlation coefficients (ICCs) were computed for each score to evaluate the inter-rater reliability (IRR). Mixed design analysis of variance (ANOVA) was computed to compare the mean scores given by the experts and the non-experts.

Results: Data from 58 subjects were analyzed. For the DSC, inter-rater reliability between facilitator and novice raters was excellent with $r=.91$, $p<.001$. The HFRS demonstrated acceptable IRR with $r=.69$, $p<.001$ and with respect to the GRS, inter-rater reliability was good, $r=.78$, $p<.001$. The results of the mixed design ANOVA show that non-experts tend to give higher absolute scores than experts when using the DSC and HFRS, but all raters give similar absolute scores when using the GRS. The results of the mixed design ANOVA of mean scores for the DSC, HFRS and the GRS are presented in Figure I.

Discussion: Good inter-rater reliability was demonstrated for the DSC, HFRS, and GRS, similar to results found in other HFS studies. This study has demonstrated that experts and non-experts agree on the ranking or general calibre of the different participants but non-experts were more generous in their assessment of performance using a checklist. Non-experts can be used to rate participants if global scores are used. If checklists are to be used, non-experts can be used to identify the lower ranked participants, who can then be double checked by experts. In this way, the time required of expert raters can be reduced.

References: 1. Can J Anesth 1997;44:1191-5.

HFRS=Human Factors Rating Scale GRS=Global Rating Scale