Introduction

The goal of this study was to use a well-established format of testing and evaluation of residents in order to assess the construct-related validity, by using the progression of scores with the level of training.

This study measured construct validity by evaluating the progression of simulator scores with the level of training, or the years of training, via examination of the residents of residents in 3 years of training postgraduate years (PGY) 2 - 4 in one institution.

Using the simulation-based format of our “practical” exam, we merely tested the upper level of competence – the “does” stage, according to the Miller’s Pyramid of Clinical Competence (3). Yet, the training, here in the broad sense to refer to all determinants of resident training potential, and ability (including all forms of knowledge acquisition).

The term, PGY 2-4, denotes the 3 years of training in an Anesthesiology residency program in the U.S. (i.e., CA, A & B & C).

Materials and Methods

The examination was administered to 50 residents undergoing PGY 2-4 anesthesia training. Residents were tested in 1 of 2 scenarios in each of the 3 major anesthesia fields: operating room (OR), trauma management, and resuscitation (Appendix). Each scenario was evaluated by 2-1 evaluators according to a pretest checklist comprised of 13-20 items. Examiners “passed” the scenario if they successfully performed 70% of the station’s checklist items, including all critical actions/items. For each item in each of the scenarios, we calculated the following:

2) Error rate (the degree to which the residents did not perform the items in a scenario satisfactorily). Error rate was calculated based on 3 factors as the sum of X errors for residents in all (total) items + total X error for items in a scenario, as a portion of all items tested by n residents - X / [n X (frequency)]

3) Performance/Difficulty Grade (the ratio of residents who performed an item satisfactorily in the scenario). Difficulty grade was expressed as the mean error in the item performed satisfactorily by a group of residents in the scenario. This grade was calculated based on 2 factors, for n residents and X errors = 1 (sum of X errors in item / n error items tested for each item = X / [n X (frequency)]). The grade was then calculated, for all items in the scenario tested by a group of residents, as mean ± SD.

3) Critical Items: errors occurred across PGY levels and were analyzed; their frequency was compared between and among groups.

The following scores were computed for each resident and for each scenario:

a) Proportion correct (Total) across all items in the checklist, across the 1-2 evaluations (for correct performance, 0 otherwise); the final score was expressed as the % of items performed satisfactorily out of the total possible items in a scenario;

b) Proportion correct (Critical) we used the same formulas as above for assigning scores to critical items included in the checklist; the critical items error rate was calculated at the rate at which the examiners did not perform critical items;

c) Mean general (Total) subjective evaluation across the examiners was expressed as a general score, on a scale of 1 to 4, with 1 indicating unsatisfactory and 4 indicating excellent performance.

Results

The examination was administered 66 times to 50 residents. The grade of residents was based on OSCE scores, in each of the scenarios, similar between PGY 1 and 2 in every field, and between the different PGY levels (Table 2, top).

The error rate was lower for PGY-4 residents compared to PGY-2s in each field, and also in each scenario except in scenario OR #1 and trauma #2, whereas the error rate was significantly higher in PGY level (Table 2, bottom).

The total proportion correct score was significantly higher for PGY-3 and PGY-4 residents compared to PGY-2s in Trauma #1 scenario (Table 3, top).

The general (factorial) score was significantly higher for PGY-4 residents compared to PGY-2s in OR #2, trauma #1 and resuscitation #1 scenarios, with a critical item error rate which was significantly lower for PGY-3 residents compared to PGY-1 and PGY-2 in the resuscitation field (Table 4, top).

The final pass rate was significantly higher for PGY-3 and PGY-4 residents compared to PGY-2s in the OR field, but not in the trauma or resuscitation field; this rate was also significantly higher for PGY-4 residents compared to PGY-2 and PGY-3 in the resuscitation field (Table 4, bottom).

Discussion

The process of incorporating simulation-based OSCE-driven modules is the testing and certification of anesthesiologists addressed with this work confirms the construct-related validation.

The examination also provided a rare glance at the performance of residents in American institutions, highlighting areas of strength and weaknesses.

The present process may evolve in the future not only as a constructive form of feedback for residency programs and means of establishing simulation-based training as part of the residency curriculum, but also toward the adoption of mannequin patient simulation-based accreditation.

The ASA’s adoption and incorporation of a re-certification simulation course, which must be completed at an ASA-endorsed simulation center, is the first step toward this goal.

NCCDA Part IV requirements: update. American Board of Anesthesiology Inc. September 20, 2010. Available at: http://www.absa.org/for-


Summary

We confirm in this work the construct-related validity for an evaluation of the process of incorporating simulation-based OSCE-driven modules in the testing and certification of anesthesiologists.

References

