Evaluating Construct Validity of Simulation-based OSCE for Summative Assessment in an Anesthesiology Teaching Program

A. Sidi[1], H. Berkenstadt[2], A. Ziv[2], T. Euliano[1], S. Lampotang[1], C. White[1];

Department of Anesthesiology[1], University of Florida College of Medicine, Gainesville, FL; Israel Center for Medical Simulation (MSR)[2], Tel Hashomer, Israel

Introduction

The goal of this study was to use a well-established format of testing and evaluating with simulation,[1-2] in order to assess the construct-related validation, 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 anesthesiology residents in all 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 model of medical competence.[3] We use the term, *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-1, 2 & 3).**

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 1-2 evaluators according to a preset checklist comprised of 12-20 items (Table 1). **Examinees "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:

1) *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_t errors for n residents in all (total) i items = total X_t error items in scenario, as a portion of all items i tested by n residents = $[X_t /(i \bullet n)]$.

2) *Performance/Difficulty Grade* (the ratio of residents who performed an item satisfactorily in the scenario). Difficulty grade was expressed as the non-errors in 1 item, performed satisfactorily by a group of n residents in the scenario. This grade was calculated based on 2 factors, for n residents and X_i errors = 1- (sum of X_i errors in 1 item/ n residents tested) for each item = $[\mathbf{1- (X_i /n)}]$. 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 evaluators (1 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 formula as above for assigning scores to the critical items included in the checklist; **the critical items error rate was calculated as the rate at which the examinees did not perform critical items;**

c) ***Mean general (Global) 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 examinees. The grade of *difficulty (= performance grade)* was similar between scenario #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, where the error rate was relatively high in all PGY levels (**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 (subjective)* score was significantly higher for PGY-4 residents compared to PGY-2s in OR #2, trauma #1 and resuscitation #1 scenarios, and in the OR field (**Table 3, bottom**).

The *critical items error rate* was significantly lower for PGY-4 residents compared to PGY-3s in the OR field; this rate was also significantly lower for PGY-4 residents compared to PGY-2s and PGY-3s 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, but not in the trauma or resuscitation field; this rate was also significantly higher for PGY-4 residents compared to PGY-2s in all 3 fields (**Table 4, bottom**).

Appendix

Example for Operating-Room scenario checklist:

Examinee Code:_____ PGY_____

Scenario – O.R. Type 1
Hypotension developing in the O.R.:
Initial response - Fluid challenge
Differential diagnosis (at least 3)
Decision re monitoring/labs
Assess fluid balance as a cause
Assess drug effect (anesthesia) cause
Assess cardiac reason/diagnosis
Assess pulmonary reason/diagnosis
Assess drug sensitivity/hormonal effect
Assess metabolic reason/MH
Final diagnosis/treatment
Blood Reaction: ↑Pulmonary inspiratory pressure; hypoxemia developing
Clinical diagnosis – Pulmonary sounds
Evaluate mechanical problem (systematically)
Evaluate airway resistance with capnography
Evaluate anesthesia level
Suspect allergic reaction – Skin
Change parameters in mechanical ventilation
Decision about final diagnosis
Treatment - Pharmacological agents/Groups
Agent choice & doses for treating bronchospasm
Scenario – O.R. Type 2
Hypertension developing in the OR:
Initial response - Check anesthesia level
Initial treatment before differential diagnosis - Oxygen for desaturation
Differential diagnosis (at least 3)
First treatment after differential diagnosis – Adjust anesthesia level
Assess & adjust fluid balance
Consider drug effect (anesthesia level)
Consider cardiac reason/diagnosis
Consider pulmonary reason/diagnosis
Consider drug sensitivity/hormonal effect
Consider metabolic reason/MH
Decision on final diagnosis/treatment
Decision – Treat HTN and HR pharmacologically
Drug Combination Treatment: Vasodilators + Betablockers
Additional invasive monitoring – Possible location of AL
"Overshoot" interpretation of AL
Additional invasive monitoring – CVP
Invasive monitoring CVP complication
CVP level interpretation – Fluid challenge directed
Arterial blood gas interpretation
Chest x-ray interpretation

Table 1. Distribution of Residents in Each Postgraduate Year (PGY-2 to PGY-4), Each Domain (Operating Room, Trauma, and Resuscitation), and Each Scenario (Type 1 and Type 2) within Each Domain (including Number of Items Tested), and the Critical Items in Each Scenario and Domain.

Scenario	O.R. Type 1	O.R. Type 2	O.R. Domain	Trauma Type 1	Trauma Type 2	Trauma Domain	Resuscitation Type 1	Resuscitation Type 2	Resuscitation Domain
PGY-2	3	5	8	3	5	8	4	3	7
PGY-3	5	4	9	3	4	7	3	4	7
PGY-4	4	3	7	4	2	6	4	3	7
No. of items	20	20	40	14	12	26	17	17	34
Critical items	3	2	5	5	4	9	11	11	22

Table 2. Descriptive statistics, including the grade of difficulty (top) and rate of errors (bottom) for residents in each postgraduate year (PGY) 2-4 in each field/domain, and in each scenario (#1 and #2) within each field.

Scenario		OR1	OR2	OR	Tr1	Tr2	Tr	Res1	Res2	Res
Grade of Difficulty	PGY-2	0.62± 0.33	0.58± 0.29	0.60± 0.30	0.62± 0.29	0.71± 0.27	0.67± 0.28	0.84± 0.31	0.67± 0.26	**0.75 0.29**
	PGY-3	0.66± 0.25	0.63± 0.33	0.65± 0.29	0.83± 0.22	0.70± 0.22	0.79± 0.22	0.90± 0.20	0.81± 0.27	**0.86 0.24**
	PGY-4	0.69± 0.24	0.77± 0.33	0.73± 0.29	0.93± 0.12	0.73± 0.41	0.85± 0.29	0.96± 0.10	0.84± 0.29	**0.90 0.22**
	All	0.66± 0.22	0.64± 0.22	0.65± 0.22	0.80± 0.16	0.71± 0.23	0.77± 0.19	0.90± 0.17	0.78± 0.22	**0.84± 0.20**
Error Rate	PGY-2	38	40	39	38	33	35	16	33	**24**
	PGY-3	34	35	34	17*	27	22	10	19	**15**
	PGY-4	31	22*	27*	7*	29†	14*	4*	16*†	**9***
	All	**34**	**35**	**34**	**19**	**30†**	**25‡**	**10**	**22†**	**16‡§**

Note that the error rate for all residents in trauma and resuscitation fields was lower than < the OR field, and lower in the resuscitation field than < in the trauma field.

*P < 0.05 compared to PGY-2 ; †P < 0.05 compared to scenario 1;

‡P < 0.05 compared to OR in the same PGY; §P < 0.05 compared to trauma in the same PGY.

Abbreviations: OR, operating room; Tr, Trauma; Res, Resuscitation.

Table 3. The total score (top), and the general evaluation score (bottom) for residents in each PGY in each field, and in each scenario within each field.

Scenario		OR1	OR2	OR	Tr1	Tr2	Tr	Res1	Res2	Res
Total (Proportion Correct) Score	PGY2	0.57± 0.03	0.62± 0.09	0.60± 0.08	0.62± 0.18	0.67± 0.22	0.65± 0.19	0.83± 0.09	0.67± 0.14	**0.76± 0.13**
	PGY3	0.61± 0.18	0.65± 0.13	0.63± 0.15	0.84†± 0.08	0.73± 0.20	0.82± 0.16	0.90± 0.07	0.81± 0.12	**0.88± 0.10**
	PGY4	0.65± 0.08	0.77± 0.08	0.70± 0.10	0.93± 0.10†	0.71± 0.06*	0.80± 0.14	0.96± 0.06	0.82± 0.06	**0.86± 0.07**
	All	0.61± 0.12	0.67± 0.11	0.64± 0.12	0.81± 0.18	0.70± 0.18	0.75± 0.18	0.89± 0.09	0.78± 0.11	**0.84± 0.11**
General Evaluation Score	PGY2	1.83± 0.29	1.60± 0.22	1.68± 0.26	1.67± 0.29	2.30± 1.15	2.06± 0.94	2.00± 0.00	1.50± 0.50	**1.79± 0.39**
	PGY3	2.00± 0.21	1.88± 0.85	1.94± 0.68	2.33± 0.58	2.13± 1.03	2.21± 0.81	3.00± 0.00	1.50± 0.71	**2.38± 1.09**
	PGY4	2.60± 0.25	2.33± 0.29†	2.50± 0.71†	3.50± 1.00†	2.25± 1.06	3.08± 1.11	3.25± 0.87†	1.83± 0.29	**2.41± 0.86**
	All	**2.17± 0.72**	**1.88± 0.57**	**2.02± 0.65**	**2.60± 1.05**	**2.23± 0.98**	**2.40± 1.01**	**2.73± 0.75**	**1.6± 0.52**	**2.19± 0.86**

*P < 0.05 compared to scenario 1 ; †P < 0.05 compared to PGY2

Abbreviations: OR, operating room; Tr, Trauma; Res, Resuscitation.

Table 4. The critical items error rate (top) and the ***final pass rate*** (bottom), for residents in each postgraduate year (PGY) in each field.

Field/Domain		OR	Tr	Res	
Critical Items Error Rate	PGY-2	0.15	0.14	0.27	
	PGY-3	0.16	0.13	0.05*	
	PGY-4	0.03†	0.06	0.03*	
	All	0.12	0.11	0.06	
		OR	**Tr**	**Res**	**3 fields combined**
Final Pass Rate	PGY-2	0	0.25	0	**2/23=0.09**
	PGY-3	0.33*	0.29	0	**5/23=0.22**
	PGY-4	0.71*	0.67	0.29	**11/22=0.50***
	All	**0.33**	**0.29**	**0.10**	

*P < 0.05 compared to PGY-2; †P < 0.05 compared to PGY-3

Discussion

The process of incorporating simulation-based OSCE-driven modalities in 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 residencies, highlighting areas of **strength and weakness**.

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.**

† MOCA Part IV requirements update. American Board of Anesthesiology, Inc. September 21, 2010. Available at: http://www.asahq.org/For-Members/Education-and-Events/Calendar-of-Events, and at:

http://www.asahq.org/education. Accessed 8/10/11.

Summary

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

References

1. Ziv A, Rubin O, Sidi A, Berkenstadt H. Credentialing and certifying with simulation. Anesthesiol Clin 2007; 25:261-9.

2. Berkenstadt H, Ziv A, Gafni N, Sidi A. Incorporating simulation-based OSCE into the Israeli national board examination in anesthesiology. Anesth Analg 2006; 102:853-8.

3. Miller GE. The assessment of clinical skills/competence/performance. Acad Med 1990; 65 (9 Supp):S63–7.