



Standardized Objective Exam vs. High Fidelity Simulation: Two Models, Same Content

Araceli Hambleton Fuentes^{1,*} and Jesus Alfonso Beltran-Sanchez²

¹Tecnologico de Monterrey, Escuela de Medicina y Ciencias de la Salud, Ave. Morones Prieto 3000, Monterrey, 64710, Mexico

²Tecnologico de Monterrey, Escuela de Medicina y Ciencias de la Salud, Ave. Morones Prieto 3000, Monterrey, 64710, Mexico

*Corresponding author. Email address: ahamblet@tec.mx

Abstract

The evaluation of learning is becoming more relevant in medical education institutions, which has led to a rethinking of the design, application, and research of the process itself. In this study, we conducted a quantitative study with a quasi-experimental design utilizing pre-test and post-test. Twenty medical students enrolled in two classes (groups) in the subject of Respiratory Physiopathology in the Surgeon Medical Program of the School of Medicine and Health Sciences at Tecnológico de Monterrey participated. Three variables were used: 1) the student experience with the evaluation process, 2) learning engagement and 3) burnout syndrome. Three assessment techniques were applied in each group. In both groups, the Portfolio and the Written Examination were generic. The distinction between the two groups was in the application of the third tool. The PEE class (8 students) received the "Standardized Objective Examination." The PSE class (12 students) received the "High Fidelity Simulation." The results indicated that both techniques promoted a sense of student satisfaction, representing an opportunity to demonstrate the development of their competencies. However, in the results of the final overall knowledge review, it was noted that the PEE Group scored higher. Each method offers its advantages in both the assessment and the learning itself. Their utilization will depend on the availability of each institution to define its use according to their goals.

Keywords: High-Fidelity Simulation; Learning Assessment models; Competences; OSCE; Educational Innovation, Higher Education

1. Introduction

Universities should rethink how to define, measure, and demonstrate mastery of a topic (Adams, Cummins, Davis, Freeman, Hall, & Ananthanarayanan, 2017). In this regard, López and Sicilia (2017) mention that, currently, many universities have chosen to favor a new assessment approach, which is oriented toward learning and promotes the participation of the student in this activity. Furthermore, Argudin (2014) mentions that the evaluation has gone from being understood as the

measurement of learning with an exam that grants a grade to become a more complicated process that goes beyond just measuring. Now, it must begin from an integrative development experience and allow the student to expand his or her strengths; also, it must involve the recovery of learning and, at the same time, be a moment of learning by itself (Ferreiro, 2012).

However, in the Mexican context, most institutions training future health professionals carry out the evaluations of disciplinary competencies through written examinations (Gandia and Romeo 2019). Although there



is evidence of other techniques, the examination remains the predominant technique (León-Bórquez, Lara-Vélez & Abreu-Hernández 2018). Among students, this traditional assessment has generated feelings of not being evaluated in full and dissatisfaction with the evaluation methods. There is also a disparity in the development of disciplinary competencies, calling into question the students' training.

On the other hand, efforts have been made to integrate evaluation techniques that truly measure competencies, putting it on the student to demonstrate the mastery of knowledge, skills, and attitudes. The new way is to seek through the evaluation a systematic process of obtaining information about the student, where the student has the means to assess his current condition, form value judgments, and decide how to increase his learning. The application of Objective Standardized Clinical Evaluation (OSCE) stations and clinical simulation with high fidelity technology models are the most used (López-Sánchez et al., 2013; Urta, Sandoval, and Iribarren, 2017; Royce, Hayes, and Schuartzstein, 2019).

Consequently, a model is required that gives the student the benefit of collecting a considerable amount of evidence about the development of his disciplinary and generic competencies, allowing him, also, to experience reflection and feedback processes (Hambleton and Beltran 2019). Therefore, one research objective aims at identifying the differences between two evaluation models in terms of the impact of the assessment methods on learning, feedback, and the relationship between perceived learning and satisfaction with the evaluation method.

2. State of the art

The "showing how it is done" requires the use of situations that simulate real-life and expand the possibility of timely feedback, generating in the student a higher degree of self-knowledge of their achieved learning and, therefore, greater satisfaction.

2.1. OSCE

The standardized objective assessment has been recognized in the medical field as a precise, objective, and reproducible instrument. The format facilitates giving standardized tests to students on a wide range of skills, while directly observing the students in scenarios that simulate clinical environments. However, it requires a higher use of human and material resources and the establishment of evaluation circuits for each group of students. (Dávila 2014). The standardized objective clinical examination is a versatile tool that evaluates competencies based on evidence through direct observation of a previously established target. It is an accurate and reproducible instrument, which allows standardized tests to be applied to students on a wide range of clinical skills. Unlike the traditional examination, it allows evaluating the most relevant aspects of performance, such as communication skills

and ability to manage the unpredictable behavior of a patient; however, like other evidence, it has been criticized for requiring a large quantity of material and human resources (Zayyan, 2011). Even educators who have written about this examination for more than 30 years reflect upon the analyses of results obtained with this strategy. They point out that when the tool is well designed, it can produce reliable results, but this is worth checking (Swanson and van der Vleuten 2013).

2.2. High-fidelity Simulation

Simulating, as a means of assessment, offers the possibility of verifying the progress students make toward programmed learning objectives and allows times for professional supervision and reflection on mistakes. It has proven to be useful for the development of competencies and learning (Segura, Valencia, and López, 2019) and allows measuring the increase of "know-how" in each student. The high-fidelity-simulator manikins can reproduce physiological parameters like those that occur in normal or deteriorating real-life scenarios. This method allows the student to receive feedback within a safe learning environment that is rich with opportunities. The use of high-fidelity simulators has been shown to improve learning outcomes, such as satisfaction, self-confidence, self-efficacy and knowledge (La Cerra et al. 2019). Other studies show these to be useful for the development of learning and skills (Dávila, 2014; Carvalho et al., 2018; Escudero, Silva, and Corvetto, 2019). High-fidelity simulators are well accepted by students and are used to measure the development of competencies (Hambleton, 2019). Some other advantages are the possibility to evaluate programmed learning objectives, have time for reflection on errors, and facilitate professional supervision (Prettyman, Knight, & Allison, 2017).

2.3. Exam

The "applied knowledge," or the integration of knowledge, is widely evaluated in Medicine through written multiple-choice exams with the use of clinical vignettes or short cases that highlight the critical judgment of the student and the application of knowledge. However, the feedback on these ends up being the bibliographic justification of the answer marked as correct; there is no space for discussion, which leads to the debate about whether this is the best strategy to evaluate cognitive competencies. Also, written examinations prepared with multiple-choice questions have been used globally in multiple fields of education. They remain an essential part of the assessment process, and, according to Bloom's ranking, they should be used in the assessment of cognitive competencies (Javaeed, 2018).

2.4. Related research

La Cerra (2019), in a systematic review based on the Cochrane Handbook for Systematic Reviews of Interventions, compared high-fidelity patient

simulation with other teaching methods and revealed that it had a greater impact on nursing students' knowledge and performance. Further exploratory studies are necessary to understand its effectiveness in improving nursing students' competencies and patient outcomes.

Adib-Hajbaghery and Yzazdabi (2018) conducted a systematic review by searching through Google Scholar, SID, Iranmedex, PubMed, and ERIC databases. They found 15 articles to select from 110 publications. The results indicated that the students were satisfied with the implementation of OSCE.

In a systematized search on databases, Martínez-Castillo and Miranda-Matus (2015) found that students consider high-fidelity patient simulation an enriching experience that fosters the development of thinking skills.

Turner & Dankoski (2008) found that successful OSCEs are often the result of significant planning, coordination of multiple resources, commitment to large-scale testing, and judicious use of assessment data. More research is needed about the best uses of the OSCE method and how to maximize reliability and validity; therefore, advocates of the OSCE.

3. Materials and Methods

3.1. Design

A quantitative-type study was conducted with a quasi-experimental design having pre- and post-tests (Edmonds and Kennedy, 2019; Head and Harsin, 2018).

3.2. Sample

Twenty students participated from the Medical Surgeon curriculum of the School of Medicine and Health Sciences at Tecnológico de Monterrey. They all were studying Respiratory Physiopathology under the supervision of the same teacher. The participants were divided naturally into two class groups, where the first group consisted of eight students who were assessed with the Portfolio, ECOE, and Exam (PEE); the second group had 12 students who were evaluated with the Portfolio, High Fidelity Simulation, and Examination (PSE).

3.3. Instrument

The instrument applied had three sections: (1) identification data with student-specific information; (2) assessment experience data, which had four items aimed at collecting information about the student's experience with the evaluation methods, and (3) two scales, adapted and translated into Spanish (Martos, Pérez-Fuentes, del Mar, Gázquez, del Mar and Barragan, 2018). These scales measured Burnout (14 items) and Engagement (11 items). Both constructs used a Likert-type scale, ranging from 0 (totally disagree) to 4 (totally agree). It should be noted that

negative and positive items were added together.

3.4. Procedure

1. Consent from the relevant authorities was requested for the implementation of the methodologies.
2. Students were informed of the implementation of the project.
3. Four learning assessment techniques were applied three times during the period August-December 2019.
 - 3.1 Portfolio: Ten evidence assignments on pathophysiology were collected, including conceptual maps, summaries, and algorithms freely designed by the students.
 - 3.2 Multiple Choice Written Exams: During the period, the same three bi-monthly knowledge assessments were applied to all the students through the Blackboard platform.
 - 3.3 For the PEE group, four stations were designed for (a) Interpretation of normal breathing noises in a simulated patient, and (b) Interpretation of abnormal breathing noises in a noise simulator.
 - 3.4 For the PSE group, an acute respiratory failure scenario was designed in a high-fidelity simulator, where the student was to interpret abnormal respiratory noises and abnormal findings and make a therapeutic decision, followed by the corresponding debriefing.
4. Impact measurement surveys were designed and applied to the students.
5. The results report was prepared.

4. Results and Discussion

An initial measurement was performed to diagnose and determine the equality between the groups. For this, the Mann-Whitney nonparametric U test was applied, thus determining that the groups were homogeneous in most variables except the variables of the relationship between grade assigned and learning and burnout (see Table 1).

Midway through the course, the instrument was again applied. On this occasion, we found that the groups remained comparable. However, we noted that the perception of the students in the PSE group towards their experience with the proposed strategies improved considerably over the first measurement. In contrast, the PEE group remained at the same levels (see Table 2).

Table 1. Pre-test of the PEE group (n=8) and the PSE group (n=12)

Variables	Descriptive statistic				Hypothesis test	
	PEE		PSE		U	p
	M	SD	M	SD		
Positive impact of the evaluation on my learning.	7.5	1.1	5.9	2.0	24.5	.064
Contribution of feedback to my learning.	7.8	1.5	6.5	2.3	31.5	.193
Relation between grade and perceived learning.	7.4	1.7	5.2	2.3	18.0	.019
Satisfaction with the evaluation methods.	7.4	1.7	6.0	2.3	30.5	.169
Engagement.	3.2	0.4	3.0	0.4	40.5	.560
Burnout.	1.1	0.5	2-0	0.7	17.0	.016

Note: Descriptive statistics are shown for better interpretation; however, it is recognized that this type of test only works with frequencies, medians, and average ranges.

M = mean, SD = Standard deviation, U = Mann-Whitney statistical, p = significance

Table 2. Intermediate measurement of the PEE (n=8) and PSE groups (n=12)

Variables	Descriptive statistic				Hypothesis test	
	PEE		PSE		U	p
	M	SD	M	SD		
Positive impact of the evaluation on my learning.	9.0	1.1	8.4	1.5	38.0	.425
Contribution of feedback to my learning.	9.1	.8	8.8	1.3	45.0	.808
Relation between grade and perceived learning.	7.3	3.	7.8	1.8	47.0	.937
Satisfaction with the evaluation methods.	8.5	1.1	8.5	1.8	42.0	.633
Engagement	3.2	0.4	3.1	0.3	35.5	.333
Burnout	1.1	0.7	1.7	0.7	26.5	.097

Note: Descriptive statistics are shown for better interpretation; however, it is recognized that this type of test only works with frequencies, medians, and average ranges.

M = mean, SD = Standard deviation, U = Mann-Whitney statistical, p = significance

At the end of the course, the final measurement was made. The findings indicate that the PSE group had an equal or better perception of the evaluation techniques; likewise, one can see a gradual increase in engagement and a reduction in burnout (see Table 3).

An intragroup analysis was also carried out to determine statistically significant differences between the different measurement times. For this purpose, Friedman's non-parametric test was applied. In Table 4, it is possible to observe that the PSE group managed to have significant changes during the run of the course, ending with scores higher than the initial ones. However, it is noteworthy that the PEE group maintained the original high scores, without those being affected.

The final student grade results (0-100 points) were compared between the two groups. The results

indicated that there was a statistically significant difference. The PEE group obtained the higher scores (PEE: Mean= 77.25, SD= 9.32; PSE: Mean= 67.5, SD= 8.14; $U' = 19.5$, $p = .026$).

Table 3. The final measurement of the PEE group (n=8) and the PSE group (n=12)

Variables	Descriptive statistic				Hypothesis test	
	PEE		PSE		U	p
	Mean	SD	Mean	SD		
Positive impact of the evaluation on my learning.	8.6	1.5	8.7	2.1	42.5	.654
Contribution of feedback to my learning.	8.9	1.4	9	2.0	41.0	.552
Relation between grade and perceived learning.	7	2.7	7.8	2.1	37.0	.387
Satisfaction with the evaluation methods.	7.6	1.6	7.8	1.4	44.5	.778
Engagement	3.0	0.5	3.1	4.0	35.0	.312
Burnout	1.0	0.8	1.7	8.0	25.5	.163

Note: Descriptive statistics are shown for better interpretation; however, it is recognized that this type of test only works with frequencies, medians, and average ranges.

M = mean, SD = Standard deviation, U = Mann-Whitney statistical, p = significance

5. Conclusions

The perception of the students in the PSE group toward their experience with the proposed strategies improved considerably over the first measurement. This finding aligns with the work of Martínez-Carrillo & Matus-Miranda (2015), who found High fidelity simulation is a helpful learning strategy that fosters the development of thinking skills and student competencies. López-Sánchez et al. (2013) implemented high-fidelity simulation also as a learning tool with the positive results of higher ending scores, like the PSE group.

We conclude in our research that both methods are beneficial to the students in terms of educational experience, as they promote a better sense of learning. However, the evidence found suggests that the PSE may yield better results in terms of the final measurement of learning, which ended up being evaluated by an exam. This suggests that the PSE model is better accepted because it allows students to learn through the evaluation, measurement, and development of their disciplinary and generic competencies, and they like the moments of individual reflection and group feedback. It is also noteworthy that both methods maintained student engagement in learning and reduced burnout syndrome.

We recommend replicating this study with the inclusion of two more groups, one being a control group, that is, without any intervention, and another that combines both methods in a single group. We also recommend observing and conducting interviews with the students to delve more deeply into the impact of the methodologies and the students' perceptions of the experience.

Acknowledgment

The authors wish to acknowledge the technical support of Writing Lab, TecLabs, Tecnológico de Monterrey, Mexico, in the production of this work.

Table 4. Differences between the evaluation times of the PEE (n=8) and PSE groups (n=12)

Variables	Descriptive statistic						Hypothesis test	
	Pre		Intermediate		Post		X ²	p
	M	SD	M	SD	M	SD		
PEE								
Positive impact of the evaluation on my learning.	7.5	1.1	9.0	1.1	8.6	1.5	8.3	.016
Contribution of feedback to my learning.	7.8	1.5	9.1	0.8	8.9	1.4	6.5	.040
Relation between grade and perceived learning.	7.4	1.7	7.3	3.0	7.0	2.7	1.1	.582
Satisfaction with the evaluation methods.	7.4	1.7	8.5	1.1	7.6	1.6	1.7	.431
Engagement	3.2	0.5	3.2	0.4	3.0	0.5	2.5	.291
Burnout	1.0	0.5	1.1	0.7	1.0	0.8	0.2	.895
PSE								
Positive impact of the evaluation on my learning.	5.9	2.0	8.4	1.5	8.7	2.1	12.3	.002
Contribution of feedback to my learning.	6.5	2.3	8.8	1.3	9.0	2.0	15.3	.000
Relation between grade and perceived learning.	5.2	2.3	7.8	1.8	7.8	2.1	6.7	.035
Satisfaction with the evaluation methods.	6.0	2.3	8.5	1.8	7.8	1.4	7.9	.020
Engagement	3.0	0.5	3.1	0.4	3.1	0.4	1.3	.526
Burnout	2.0	0.7	1.7	0.7	1.7	0.8	3.2	.197

Note: Descriptive statistics are shown for better interpretation; however, it is recognized that this type of test only works with frequencies, medians, and average ranges.

M=mean, SD= Standard deviation, X²= Friedman Chi-square, p=significance

References

- Adams, S., Cummins, M., Davis, A., Freeman, A., Hall C., & Ananthanarayanan, V. (2017). *NMC Horizon Report: 2017 Higher Education Edition*. Texas: The New Media Consortium.
- Prettyman, A. V., Knight, E., & Allison, T. E. (2018). Objective Structured Clinical Examination From Virtually Anywhere! *Journal for Nurse Practitioners*, 14(8), e157-e163. <https://doi.org/10.1016/j.nurpra.2018.05.007>
- Adib-Hajbaghery, M., & Yazdani, M. (2018). Effects of OSCE on learning, satisfaction, and test anxiety of nursing students: a review study. *Iranian Journal of Medical Education*, 18, 70-83. <http://ijme.mui.ac.ir/article-1-4539-en.html>
- Argudin Y. (2014). *Educación basada en competencias. Nociones y antecedentes*. México: Trillas
- Carvalho, M. H., Colaço, S., Rafael, H., Baixinho, C. L., Félix, I., Saraiva, C., & Rebelo, I. (2018). Aprender com a Simulação de Alta Fidelidade. *Ciência & Saúde Coletiva*, 23(1), 51-59. <https://doi.org/10.1590/1413-81232018231.23072017>
- Dávila A. (2014). Simulación en educación médica. *Investigación en educación médica*, 3(10), 100-105.
- Edmonds, W. A. & Kennedy, T. D. (2019). Quantitative Methods for Experimental and Quasi-Experimental Research. In: W. A. Edmonds and T. D. Kennedy, eds. *An Applied Guide to Research Designs: Quantitative, Qualitative, and Mixed Methods*. Thousand Oaks, CA: Sage, 29-34
- Escudero, E., Silva, M., & Corvetto, M. (2019). Simulation: A Training Resource for Quality Care and Improving Patient Safety. In S. Çelik Durmuş (Ed.), *Nursing - New Perspectives* (1st ed.). IntechOpen. <https://doi.org/10.5772/intechopen.88918>
- Ferreiro, R. (2012). *Cómo ser mejor maestro: El método ELI*. México: Trillas
- Gandia, P., & Romeo, A. (2019). La exigencia cognitiva en los exámenes tipo test en contexto universitario y su relación con los enfoques de aprendizaje, la autorregulación, los métodos docentes y el rendimiento académico. *European Journal of Investigation in Health, Psychology and Education*, 9(3), 177-187.
- Hambleton, A. & Beltrán, J.A. (2019). "POSE": una aproximación a la evaluación de competencias preclínicas. 6to Congreso Internacional de Innovación Educativa, 1673-1678. Dic 16-18, Monterrey, México.
- Head, K. J. & Harsin, A. M. (2018). Quasi-Experimental Design. In: K. J. Head and A.M. Harsin, eds. *The SAGE Encyclopedia of Communication Research Methods*. Thousand Oaks, CA: Sage, 1384-1387
- La Cerra, C., Dante, A., Caponnetto, V., Franconi, I., Gaxhja, E., Petrucci, C., Alfes, C. M., & Lancia, L. (2019). Effects of high-fidelity simulation based on life-threatening clinical condition scenarios on learning outcomes of undergraduate and postgraduate nursing students: A systematic review and meta-analysis. *BMJ Open*, 9(2), 1-11. <https://doi.org/10.1136/bmjopen-2018-025306>
- León-Bórquez, R., Lara-Vélez, V. M. & Abreu-Hernández, L. F. (2018). Educación médica en México. *FEM: Revista de la Fundación Educación Médica*, 21(3), 119-128
- López, V. & Sicilia A. (2017). Formative and shared

- assessment in higher education. Lessons learned and challenges for the future. *Assessment & Evaluation in Higher Education*, 42(1), 77-97.
- López-Sánchez, M., Ramos, L., Pato, O., & López, S. (2013). La simulación clínica como herramienta de aprendizaje. *Cirugía Mayor Ambulatoria*, 18(1), 25-29.
- Martínez-Castillo, F., & Matus-Miranda, R. (2015). Desarrollo de habilidades con simulación clínica de alta fidelidad. Perspectiva de los estudiantes de enfermería. *Enfermería Universitaria*, 12(2), 93-98.
- Martos, Á., Pérez-Fuentes, M., del Mar, M., Gázquez, J. J., del Mar, M., & Barragán, A. B. (2018). Burnout y Engagement en estudiantes de Ciencias de la Salud. *European Journal of Investigation in Health, Psychology and Education*. 8(1), 23-36.
- Urrea, U., Sandoval, S. & Iribavent, F. (2017). El desafío y futuro de la simulación como estrategia de enseñanza en enfermería. *Investigación en Educación Médica*, 6(22), 119-125.
- Royce, C., Hayes, M., & Schuartzstein, R. (2019). Teaching Critical Thinking: A case for instruction in cognitive bases to reduce diagnostic errors and improve patient safety. *Academic Medicine*. 94(2), 187-194.
- Segura, N., Valencia, J, López, M. (2019). Desarrollo del pensamiento crítico mediante simulación de alta fidelidad en estudiantes de medicina. *Investigación en Educación Médica*, 7(28), 55-63.
- Swanson, D. B., & van der Vleuten, C. P. M. (2013). Assessment of Clinical Skills With Standardized Patients: State of the Art Revisited. *Teaching and Learning in Medicine*, 25(1). <https://doi.org/10.1080/10401334.2013.842916>
- Turner, J. L., & Dankoski, M. E. (2008). Objective structured clinical exams: A critical review. *Family Medicine*, 40(8), 574-578.
- Javaeed, A. (2018). Assessment of Higher Ordered Thinking in Medical Education: Multiple Choice Questions and Modified Essay Questions. *MedEdPublish*, 7(2), 1-8. <https://doi.org/10.15694/mep.2018.0000128.1>
- Zayyan, M. (2011). Objective structured clinical examination: The assessment of choice. *Oman Medical Journal*, 26(4), 219-222. <https://doi.org/10.5001/omj.2011.55>