J Med Sci 2020;40(4):167-174
DOI: 10.4103/jmedsci.jmedsci_203_19

ORIGINAL ARTICLE

Is it an Impact Factor for Standardized Patients with Actor Background to Perform in the Taiwan High-Stake Objective Structured Clinical Examination?

Yaw-Don Hsu[1], Hui-Chen Lin[2], Chang-Hsun Hsieh[3], Fung-Wei Chang[4], Jiunn-Tay Lee[2], Chin-Sheng Lin[5], Hsu-Kai Huang[6], Chin Lin[7]

[1]Department of Neurology, Objective Structured Clinical Examination Center, National Defense Medical Center and Tri-Service General Hospital, [2]Department of Neurology, National Defense Medical Center and Tri-Service General Hospital, [3]Division of Endocrinology and Metabolism, Objective Structured Clinical Examination Center, National Defense Medical Center and Tri-Service General Hospital, [4]Department of Gynecology and Obstetrics, Objective Structured Clinical Examination Center, National Defense Medical Center and Tri-Service General Hospital, [5]Division of Cardiology, National Defense Medical Center and Tri-Service General Hospital, [6]Division of Chest Surgery, National Defense Medical Center and Tri-Service General Hospital, [7]National Defense Medical Center, Graduate Institute of Life Science, Taipei, Taiwan

Background: In the past decade, Tri-Service General Hospital has trained many actors-background standardized patients (SPs). However, we also have recruited some nonactor SPs since 2010. Reviewing the literature, the reliability of SPs with or without actor background would be the impact factor on high-stakes OSCE has not been well studied. **Aim:** The purpose of this retrospective observation is to clarify whether a SP with or without a professional actor background will affect the performances of the examination when participating in the Taiwan high-stakes objective structured clinical examination (OSCE). The result will be the policy for selecting which kind of SPs to participate in the high-stack OSCE in our hospital. **Methods:** In this retrospective observation, we analyzed 74 actor background SPs (A-SPs), and 70 no actor background SPs (NA-SPs) who have participated in each spring test of 2015–2017 Taiwan high-stakes OSCE. The data of SPs performance have come from two parts: one is from the examinee with the global rating of the SPs performance and the other is from the examiner with eight-item checklist for the SPs performance. The scoring of both examinee and examiner is a five-point Likert scale. **Results:** The results show that there is no significant difference in SPs performance from the examinee scoring in the subsequent 3 years, stable and satisfactory, and let examinees feel like real patients. The scores from examiners also showed that two groups of SPs performed equally well without significant differences. Their performance is reliable and consistent, simulating to be a real patient in both groups. However, only one interest finding, when compared to the different age subdivision of SPs, showed the aged subgroup of A-SPs and NA-SPs with the significant difference in the item of reliability and consistent (4.67 ± 0.53 vs. 4.41 ± 0.50, $P < 0.05$), the possible reason is related to the less-experienced associate with mild memory decline in aged NA-SPs.**Conclusions:** Our observation gives the essential information, that the well-trained and experienced SP are necessary for participating in the high-stakes OSCE examination, even without the background of the actor. What we need is a professional SP who not only has excellent professionalism but also has a good personality literacy to engage in his/her work.

Key words: High-stakes objective structured clinical examination, national medical licensing examination, standardized patients

INTRODUCTION

From the literature, since 1964, standardized patient (SP) is first introduced to clinical medical education by Dr. Barrows.[1]

Received: October 20, 2019; Revised: November 03, 2019; Accepted: November 17, 2019; Published: February 21, 2020
Corresponding Author: Prof. Yaw-Don Hsu, Department of Neurology, Objective Structured Clinical Examination Center, National Defense Medical Center and Tri-Service General Hospital, Taipei, Taiwan. Tel: 886-287927173; Fax: 886-287927174. E-mail: hsu1487@gmail.com

How to cite this article: Hsu YD, Lin HC, Hsieh CH, Chang FW, Lee JT, Lin CS, *et al.* Is it an impact factor for standardized patients with actor background to perform in the Taiwan high-stake objective structured clinical examination? J Med Sci 2020;40:167-74.

Over the last 40 years, SPs have become a standard education and evaluation method in medical schools. One of the best-established and best-validated uses of simulation in medical undergraduate education and evaluation is the SPs.[2,3] A further previous study has emerged that addresses the adequacy of measurements obtained using SPs as a means of assessment.[4,5] The use of SPs for certification and licensure decisions has been a recent phenomenon.[6-10] The introduction of SP-based certification and licensure examinations in medicine was a great achievement.[8-10] The Taiwan Medical Licensing examination (TMLE) introduced high-stakes objective structured clinical examination (OSCE) incorporating SPs, first for medical graduates seeking licensing in 2013.[11] With the current Taiwan high-stakes OSCE requirements, at least eight out of 12-stations require SPs assistance examinations. It needs a large number of SPs to engage the examination. Our hospital, Tri-Service General Hospital, is one of the 21 Taiwan examination sites, and needs multiple SPs for 3 consecutive days of high-stakes OSCE in each Spring examination. Of course, the success of the whole examination and the accuracy of SP's performance are strongly related to the test results.[12-15] Therefore, a well-trained SP is critical.

The choice of an SP can be either healthy people or a patient with a specific standard form of disease.[16,17] From the literature review, we have found that their identity could not be uncertain, from patients, actors, lawyers, retired teachers, and home makers.[18,19] Most of the SPs are volunteers.[20-23] Choosing professional actors to participate in the SP program should be carefully trained. If they are not working from a script or within the familiar improvisational outline, they are unlikely to change their performance on the subject than nonactors.[16,24] Some SP training center tells actors that specialized training is required for them to become the SPs.[25] However, some well-trained actors are still the resources of SPs in some medical school portray in psychiatry complaints.[17] In the past decade, our hospital had recruited and trained many retired or part-time actors to become the SPs. They provide clinical simulations to medical students for their clinical skills training.[26,27] However, some of them cannot be suitable for the Taiwan high-stakes OSCE due to their age and gender factors. In order to meet the criteria of high-stakes OSCE, we started to recruit and train non-actor background SPs (NA-SPs) for the shortage of the previous actor background SPs (A-SPs) from 2010.

This retrospect observation, we want to know is, when our SPs participated in 2015–2017 Taiwan high-stakes OSCE, whether the SP has a professional actor background would be an impact factor. Therefore, we enrolled 74 A-SPs and 70 NA-SPs to participate in the 2015–2017 Taiwan high-stakes OSCE. The performance of SPs is assessed in global observation by examinees and in eight-item checklist observation in two different backgrounds SPs by the examiners.

METHODS

Standardized patients

All of SPs recruited to our hospital should meet the selection criteria modified by the University of Texas Health Science Center and Dr. Wang.[28,29] They should demonstrate a passionate and responsible attitude, behave an excellent intellectual ability, communication skills, accountability, patience, appropriate talk, and mental maturity.

In this study, 144 SPs (74 A-SPs and 70 NA-SPs) were enrolled to participate in the spring of Taiwan high-stakes OSCE from 2015 to 2017. The gender distribution includes 50 male and 94 females (27 males and 47 females in A-SPs; 23 males and 47 females in NA-SPs); age distribution is from 27 to 68 years (A-SPs are 27–68 years and NA-SPs are 27–68 years) [Table 1].

Both groups of SPs graduated from senior high school or college, and both groups of SPs had at least 2 years of mock OSCE examination experience in our hospital. Before participating in high-stakes OSCE, all enrolled SPs and examiners should complete the training course designed by the Taiwan Association Medical Education (TAME) OSCE committee and be accredited for the examination.

The test content of clinical skills includes history taking, physical examination, counseling/patient education, managing ability/acute care, communication, interpretation of laboratory data, and procedural skills in total 12 stations.

Table 1: General data of standardized patients engaged in the Taiwan High-Stakes Objective Structured Clinical Examination, Tri-Service General Hospital site (2015-2017)

	A-SP (%)	NA-SP (%)	Total (%)	*P*
Years				
2015	26 (35.1)	22 (31.4)	48	
2016	24 (32.4)	24 (34.3)	48	
2017	24 (32.4)	24 (34.3)	48	
Gender				
Female	47 (63.5)	47 (67.1)	94 (65.3)	
Male	27 (36.5)	23 (33.9)	50 (34.7)	
Total	74	70	144	
Age years				
Range	28-68	27-68	27-68	
Mean±SD	51.22±12.29	49.44±11.81	50.35±12.05	0.379*

*A-SP versus NA-SP. SD=Standard deviation; SP=Standardized patient

Examination process

Taiwan high-stakes OSCE program modifies from Canada,[7] a 12-station of OSCE, eight 10-min encounters with an SP, and four 10-min tests with the simulator for procedural skills. Each course of the examination is held on 3 consecutive days in our hospital. SP trainers and physician examiners conduct SP training and exercise 2 h before each day's examination, and we arranged two SPs (one A-SP and another NA-SP) to take turns performing for each SP encounter station. There was one track run and administered the OSCE twice per day, a total of 24 examinees participated per day in our site, and total 16 SPs per day required to portray real patients. A 3-day examination, to avoid excessive fatigue of SPs, affect the test results, take the 1-day test, and rest another 2 days, so a total of 48 SPs in turn performance. Therefore, the entire course of examination requires 24 physician examiners to evaluate 48 SPs' performances in 3 days' exam in our site each year, and the sum of 72 examiners and 144 SPs in 3 years.

Way to assess

After the examinations of each day, examinees and examiners fill in the assessment form for SPs' performance (this form was designed and provided by the TAME OSCE committee) for scoring.[30]

The performance of the SP assessed by the examinee was only one item scoring in the performance of all eight-stations SPs encountered on that day. Each examinee rated the score as a composited score of eight SPs whom he encountered, which was indistinguishable from the performance of A-SPs or NA-SPs, and SP's performance assessment question is "SP's performance truly like a real patient." The score is from 1 (very dissatisfied) to 5 (very satisfied).

The examiner to assess the performance of the SP is from the checklists of eight items to evaluate.

- Item 1: "SP's performance is reliable and consistent"
- Item 2: "Physical symptoms and moods simulated by the SP are accurate and real as the patient"
- Item 3: "SP deliberately manipulated the message" (reverse the question)
- Item 4: "SP's body language can comply with the requirements of the role"
- Item 5: "SP can smoothly answer the inquiry question"
- Item 6: "SP is over-question or challenges the student" (reverse the question)
- Item 7: "SP's expressions can meet the requirements of the role"
- Item 8: "SP does not have a dialog with students beyond the plot of the test"

The score of each item is the same from 1 to 5 (very dissatisfied to very satisfied). The statistical analysis method is using ANOVA and two-tailed *t*-test.

RESULTS

The gender distribution includes 50 males and 94 females (27 males and 47 females in A-SPs; 23 males and 47 females in NA-SPs). The age distribution of all SPs is 27–68 years, with a mean age of 50.35 ± 12.05 years; the A-SP age distribution is 28–68 years with the mean age of 51.22 ± 12.29 years, and NA-SP is 27–68 years with mean age of 49.44 ± 11.81 years old, $P > 0.05$ [Table 1].

In high-stakes OSCE, there was only one item for assessing the SP's performance from the examinee. The item is "The SP's performance is truly like a patient." A total of 189 examinees participated in the 2015–2017 examination [Table 2]. The mean score of SP's performance is 3.69 ± 0.64 in 2015, 3.84 ± 0.59 in 2016, and 3.86 ± 0.84 in 2017. The data showed no significant difference between 2015 and 2017 ($P = 0.316$), and it also showed that the examinees were satisfied with the performance of all the SP's in the 3 consecutive years' examination.

Analysis of the examiner's rating on the SP's performance showed that among all the SPs who have participated in the 2015–2017 examination, the scores of eight items were high, and there was no statistically significant difference among the 3 years [Table 3]. This result is similar to that of the examinee's observation and shows that during these 3 consecutive years. The performance of the SPs is quite stable and excellent, as is shown from the scores assessed by the examiners or examinees in 3 years. Based on this result,

Table 2: Global rating of Standardized patients' Performance by Examinees

Year	2015	2016	2017	Total
Examinees (*n*)	65	58	66	189
Scores (mean±SD)	3.69±0.64	3.84±0.59	3.86±0.84	*P*=0.316

SD=Standard deviation

Table 3: Standardized patients' Performance Scores by Examiners between 2015 and 2017

Item/year	2015 (48)	2016 (48)	2017 (48)	*P*
Item 1	4.60±0.54	4.65±0.48	4.58±0.54	0.836
Item 2	4.67±0.48	4.60±0.57	4.65±0.53	0.839
Item 3	4.77±0.42	4.83±0.38	4.79±0.46	0.761
Item 4	4.60±0.54	4.58±0.50	4.65±0.53	0.836
Item 5	4.75±0.48	4.60±0.49	4.67±0.56	0.380
Item 6	4.77±0.42	4.85±0.36	4.85±0.41	0.499
Item 7	4.67±0.48	4.60±0.54	4.58±0.54	0.715
Item 8	4.81±0.39	4.71±0.46	4.88±0.39	0.145
Total	37.65±3.11	37.44±3.15	37.65±2.92	0.929

ANOVA test

we can objectively analyze the differences between A-SP and NA-SP groups in the 2015–2017 examination by examiners' assessments.

Our main concern of this study is to observe whether the performance of A-SPs or NA-SPs in high-stakes OSCE is different. First, we integrated 3-year results of SPs' performance by examiners' assessments; the data showed high scores of each out of eight items of SPs' performance, and no significant difference in the analysis between A-SP and NA-SP [Table 4]. Even though the data were separately compared for different 3 years, the performance of the two groups of SPs was still excellent, and the statistic between the groups also showed no significant difference [Table 5]. Therefore, from the examiner's perspective, eight items of SPs' performance in both groups are excellent without a significant difference in the statistical analysis.

Our sample showed that the mean age of SPs in the gender was different, and male SPs was older than female SPs (56.94 ± 10.28 vs. 46.85 ± 11.48, $P < 0.001$, *t*-test). However, the mean age in the A-SPs (51.22 ± 12.29) and NA-SPs (49.44 ± 11.81) revealed no statistically significant difference ($P = 0.379$) [Table 1]. For further analysis, whether the age and/or gender factors would be an impact on the SPs' performance. We observed the results of these two factors in two SPs' groups. First, we divided ages into young and older groups with the cut point at the age of 50 for analysis. It is an interesting finding that adds age factor into two groups of SPs, the data showed older NA-SPs' scores were the lowest than other three groups in item 1 (4.41 ± 0.50) and item 2 (4.47 ± 0.51), but no significant scientific difference among four groups [Table 6]. When we compared two different backgrounds of older SPs, data show significantly lower mean scores of older NA-SPs than that of older A-SPs in item one [4.41 ± 0.50 vs. 4.67 ± 0.53, $P < 0.05$, Table 7], it means that the older NA-SPs with less reliability present less consistent information to the examinees than that of older A-NP, but fortunately, the score was still high (4.41 ± 0.50). We have also analyzed the gender factor into two groups of SPs: the scores analysis showed no significant difference in eight items among the four groups [Table 8].

Table 4: The eight different performance of A-SP versus NA- standardized patients in total 3 years

Item/SP	A-SP (74)	NA-SP (70)	*P*
Item 1	4.68±0.50	4.54±0.53	0.124
Item 2	4.70±0.52	4.57±0.53	0.133
Item 3	4.84±0.37	4.76±0.46	0.250
Item 4	4.66±0.50	4.56±0.53	0.224
Item 5	4.70±0.49	4.64±0.54	0.486
Item 6	4.86±0.34	4.79±0.45	0.234
Item 7	4.68±0.50	4.56±0.53	0.169
Item 8	4.81±0.39	4.79±0.45	0.721
Total	37.93±2.94	37.20±3.12	0.149

t-test, two tails

DISCUSSION

Over the past decade, international medical education reformers have recognized that all medical students would take responsibility for health care after graduation. The high-stakes OSCE was introduced because of the high incidence of malpractice claims, increases in medical errors, and the decrease in patient satisfaction with treatment attributed to poor interpersonal and clinical skills, as well as poor history-taking and physical examination skills.[2,31,32] It has been found that it is not only essential for the medical student to have good medical knowledge but also essential to have excellent clinical skills for patient safety and to improve patient–doctor interaction.[7,33] In 1992, the Medical Council of Canada (MCC) Qualifying Examination Part II[8] and 2004, the United States Medical Licensing Examination (USMLE)

Table 5: The performance of A-standardized patients and NA-standardized patients in the different 3 years

	2015 A-SPs (26)	2015 NA-SPs (22)	2016 A-SPs (24)	2016 NA-SPs (24)	2017 A-SPs (24)	2017 NA-SPs (24)	*P*
Item 1	4.65±0.49	4.55±0.60	4.62±0.49	4.67±0.48	4.75±0.53	4.42±0.50	0.313
Item 2	4.69±0.47	4.64±0.49	4.62±0.58	4.58±0.58	4.79±0.51	4.50±0.51	0.514
Item 3	4.77±0.43	4.77±0.43	4.88±0.34	4.79±0.41	4.88±0.34	4.71±0.55	0.719
Item 4	4.65±0.49	4.55±0.60	4.58±0.50	4.58±0.50	4.75±0.53	4.54±0.51	0.732
Item 5	4.81±0.40	4.68±0.57	4.58±0.50	4.62±0.49	4.71±0.55	4.62±0.58	0.697
Item 6	4.77±0.43	4.77±0.43	4.88±0.34	4.83±0.38	4.96±0.20	4.75±0.53	0.443
Item 7	4.69±0.47	4.64±0.49	4.62±0.49	4.58±0.58	4.71±0.55	4.46±0.51	0.597
Item 8	4.85±0.37	4.77±0.43	4.67±0.48	4.75±0.44	4.92±0.28	4.83±0.48	0.398
Total	37.88±2.88	37.36±3.42	37.46±3.19	37.42±3.17	38.46±2.80	36.83±2.87	0.566

ANOVA test. SP=Standardized patient

Table 6: The performance of A-standardized patient versus NA-standardized patient in different age groups

Variable	Young A-SPs (35)	Young NA-SPs (38)	Older A-SPs (39)	Older NA-SPs (32)	*P*
Item 1	4.69±0.47	4.66±0.53	4.67±0.53	4.41±0.50	0.088
Item 2	4.74±0.44	4.66±0.53	4.67±0.58	4.47±0.51	0.178
Item 3	4.86±0.36	4.76±0.43	4.82±0.39	4.75±0.51	0.689
Item 4	4.66±0.48	4.55±0.55	4.67±0.53	4.56±0.50	0.687
Item 5	4.80±0.41	4.66±0.53	4.62±0.54	4.62±0.55	0.403
Item 6	4.94±0.24	4.79±0.41	4.79±0.41	4.78±0.49	0.265
Item 7	4.71±0.46	4.61±0.55	4.64±0.54	4.50±0.51	0.395
Item 8	4.86±0.36	4.84±0.37	4.77±0.43	4.72±0.52	0.490
Total	38.26±2.37	37.53±3.19	37.64±3.38	36.81±3.04	0.286
Age	39.71±6.51	40.03±5.61	61.54±4.56	60.62±5.94	

ANOVA test. SP=Standardized patient

Table 7: The performances of older A-standardized patients versus older NA- standardized patients

Variable	Older A-SPs (39)	Older NA-SPs (32)	*P*
Item 1	4.67±0.53	4.41±0.50	0.038*
Item 2	4.67±0.58	4.47±0.51	0.134
Item 3	4.82±0.39	4.75±0.51	0.510
Item 4	4.67±0.53	4.56±0.50	0.402
Item 5	4.62±0.54	4.62±0.55	0.942
Item 6	4.79±0.41	4.78±0.49	0.899
Item 7	4.64±0.54	4.50±0.51	0.263
Item 8	4.77±0.43	4.72±0.52	0.655
Total	37.64±3.38	36.81±3.04	0.287
Age	61.54±4.56	60.62±5.94	0.466

*<0.05, *t*-test, two tails. Item 1: SP's performance is reliable and consistent, SP=Standardized patient

Table 8: The performances of difference gender in A-standardized patients and NA-standardized patients

Variable	Male A-SPs (27)	Male NA-SPs (23)	Female A-SPs (47)	Female NA-SPs (47)	*P*
Item 1	4.64±0.49	4.52±0.51	4.68±0.52	4.57±0.54	0.588
Item 2	4.68±0.48	4.60±0.50	4.70±0.55	4.57±0.54	0.644
Item 3	4.84±0.37	4.76±0.44	4.85±0.36	4.74±0.49	0.584
Item 4	4.56±0.51	4.48±0.51	4.70±0.51	4.62±0.53	0.348
Item 5	4.68±0.48	4.56±0.51	4.70±0.51	4.70±0.55	0.682
Item 6	4.92±0.28	4.76±0.44	4.85±0.36	4.79±0.46	0.442
Item 7	4.64±0.49	4.48±0.51	4.68±0.52	4.62±0.53	0.473
Item 8	4.76±0.44	4.72±0.46	4.83±0.38	4.83±0.43	0.662
Total	37.72±2.87	36.88±3.15	38.00±3.06	37.45±3.08	0.505

ANOVA test. SPs=Standardized patients

Step 2 Clinical Skills (USMLE Step 2 CS),[9] both initiated important clinical skills test for qualifying their medical graduated students to reduce patient harm. In 2013, TAME, the Ministry of Examination and the Ministry of Education, three organizations together to held first high-stakes OSCE for TMLE, Step 2 Clinical Skills for all medical graduated students.[34] Taiwan high-stakes OSCE is a multisite, fixed test form model with two administrations per year. Our hospital is certified as a medical center examination site and joined each spring administration. Historically, SP-based assessments were implemented as a part of formative evaluation activities.[1,16,35] There is an increasing confidence that SPs can offer a reliable learning experience for students, can provide valuable and formative feedback, and are sufficiently reliable to be used in the assessment of clinical skills acquisition.[12,36,37] For this purpose, in the initial 10 years, we also applied SP-base assessments to our medical students for clinical skills training.[26,27] In our experience, SP portrays specific patient conditions are undoubtedly successful for clinical skills training, and most of our SPs recruited from the Taiwan Actor Union in the past. For high-stakes OSCE eligibility requirements, being a certified SP to participate high-stake OSCE, must complete the training course designed by TAME OSCE, and practiced at least two times of mock OSCE held in each site.[34] Based on this requirement, each site began to train more certified SPs for backup. Owing to the age and health factors, the number of our original A-SPs is not enough, so it is necessary to train some qualified NA-SPs since 2011.

In the training process, for the consistency and accuracy of SPs performance as well as their portrayal of the case, great preparation of guidelines and curriculum for SPs is needed. A-SPs were carefully trained for preventing variable their presentation to the case. However, some NA-SPs were found to take longer and more intensive training to achieve the preset standards. Based on this, we would like to take this opportunity to see the outcome of the two groups of SPs after our training. Before joint high-stakes OSCE, we had evaluated two groups of SPs in our pretest, and found that there was no statistical difference in the performance of A-SP and NA-SP; the data is not presented here. However, as per our examiners' feedback, some of the NA-SPs do have little weaknesses that need to be improved. According to these opinions, we have more rigorous training for our SPs to prevent any errors before engaged to high-stakes OSCE. In this observation, we should know how SPs would behave when they participated in the past 3 years of high-stakes OSCE.

In our site, we arranged one A-SP and one NA-SP in the same station to play the same role during their participation in the high-stakes OSCE from 2015 to 2017. Except for 2015, there were two stations due to the age and gender need of scenario, so both stations were portrayed the cases by the A-SPs only, and the SPs' performance was assessed separately

by the examinees and examiners. According to the examinee's questionnaire, concern on SP's assessment is only one item: "Is SP's performance truly like a patient?" Moreover, give the composite score on the day of the 8-stations SPs encountered in the examination. Each examinee had encountered the 8-station SPs with different backgrounds of SPs combinations, and the rating scores could not indeed distinguish the SP performance from the different backgrounds, but the scores can still give the information of our SPs' performance each year from the examinees' viewpoint. In 2015–2017, among the total of 189 examinees, the mean scores of SP's performance were 3.69 ± 0.64, 3.84 ± 0.59 and 3.86 ± 0.84, respectively, and showed no significant difference among the 3 years, $P = 0.316$. This result indicates that the SP's performance in our site is stable and acceptable to examinees in consecutive 3 years.

Examiner's assessment of the SP is to use the checklist to score eight items separately in each SP. The results showed our SP's performance is also stable in each of 3 years, and each score of the eight items is high, indicating that SP's performance in the examination is satisfactory to the examiners. This excellent result shows that our SPs are of good validity and reliability. It is not only crucial to the examinee, but also an impact on the success of the whole test. Data analysis showed two groups of SPs' scores without significant difference in each of the checklist, even the mean score on NA-SPs was little lower than that of the A-SPs. However, the scores were high in the two groups. The difference is not shown well in the two groups, which may be due to two reasons: the first one is our SPs' performance is good validity and reliability for getting high scores and the other reason may be due to the sign of the ceiling effect. The possible explanation of the sign of high ceiling effect is related to less complicate of the scenario in high-stakes OSCE. For preventing the inconsistent performance of SPs in different sites, the scenario of cases designed not too tricky to perform, which may be the reason that each SP can get a high score on each item. Data show that the SP's performance over these consecutive 3 years has let the examiners feel satisfied in each of the eight check items, and all of these results had reached the goal of high-stakes OSCE.

The issue of concern to A-SP's performance is whether they will go beyond the cases or over self-expressed by the traits of the actors. However, the data of the third, sixth, and eighth items shown that the problem does not exist as these issues have been educated and banned in the training curriculum.

The data also showed one phenomenon that the examinee's rating appeared to be lower than that of the examinee. The possible reason to answer is that the timing of examinees to score after finish their test, we thought their emotions would be complicated in anxiousness and stress to unknown their achievements in the test, resulting in the assessment. However, the mean scores from examinees were still not bad and even reached a satisfying level in consecutive 3 years. The examiners themselves had no such problems, and their scorings were more objective and detail in the checklist assessment.

Our data also showed another thing, that the mean age of all male SPs was significantly older than that of female SPs. However, each score of eight items was not a statistically significant difference in the two genders [Table 8]. It seems to imply that the aged SPs were also not the impact factor to influence the test. Whether this is the case, for further analysis, we take a tangent to the age, divided into young and older two groups. As to which age is ideal for point of a tangent, according to the study of Joshua Hartshorne, an MIT cognitive science researcher and the lead author of a study looking at how intelligence changes as we age. The age of 50 is the beginning of decline in short-term memory, working memory, and image recognition.[38] In our training experience, more extended time and more repetition in SPs' training after the age of 50 are needed. If we add the age factor only into A-SPs and NA-SPs groups and analyze the scores in each of the eight items. The data showed that the older NA-SPs was the weakest group, and the young A-SPs was the most energetic group, even the data were not a significant difference among the four groups. However, the reliability and constant to present to examinees (item 1) score was lowest in older NA-SPs, see Table 6. When we compared older A-SPs with older NA-SPs two groups only, there is an interesting finding that data showed significant differences in item 1; older NA-SPs were less reliable and consistent in performance than older A-SP, see Table 7. Definitely, memory issue is the problem in aged SPs, but if the experience is good enough like older A-SPs, even training need more time for remembering what they learn, they still got excellent scores. From this data analysis, the age factor still wants to be concerned in high-stakes OSCE if training is not enough.

On behalf of the past, the policy to recruit professional actor background persons only to serve as SPs in our hospital could be adjusted after this observation. The results have proved that as long as well-trained, experienced, coupled with professionalism and pleasant personality, whether the background is an actor or not, can become a competent SP.

Although the data do not show the training process before the test, the SP of the actor background should be able to convey facial emoticons, mood, tone adjustment, or body language performance during the preparation for exams, regarding the actor's expertise. However, some of the symptoms of the disease still require guidance from the professional trainers and examiners so that they will not distort or exaggerate the performance. Besides, during the training process, each

different background SP would discuss and learn with each other to achieve consistency, and our A-SPs had contributed their expertise to assist our NA-SP with excellent achievement in high-stakes OSCE.

To avoid the memory issue of the older SP, in the training of the scripted dialog, the training time should lengthen for repetition, or the training method needs to be adjusted. Fortunately, because high-stakes OSCE is a 10-min station, the script content is not complicated for causing problems. Furthermore, the two groups of SPs had been screened to have a good experience and continuous training and practice in mock OSCE. Therefore, as long as training is completed before the high-stakes OSCE, all SPs should present correct and complete information to the examinees.

Finally, we would like to thank the A-SPs for their excellent clinical practice training in this decade for our medical students. At the same time, they are also helpful to the hospital in training new members of NA-SPs with enthusiasm and professionalism to smoothly pass the SP transition period, and finally got the qualified examination site of the high-stakes OSCE in Taiwan.

CONCLUSIONS

To be competent SP in high-stakes OSCE, SP's professional background as an actor or not is not an essential factor. On the contrary, well-trained, experienced, coupled with professionalism and good personality SPs, are an essential factor. We also look forward to seeing all SPs as Dr. Wilson's insights[39] that their fidelity to their task and their level of courtesy and professionalism is excellent, and their performance is good quality and flexibility on high-stakes OSCE. A professional SP can play a proper role in high-stakes OSCE, just like an excellent actor plays a proper role in the show.

Professional SP will be the goal that hospitals or medical schools should pursue. If the hospital or medical school can let a professional SP being the staffing, we believe the professional SP can not only play an excellent job in the high-stakes OSCE but also can allow hospitals to promote the clinical training of medical students and other medical professionals in the field of medical education.

Ethical approval

The study was approved by Tri-Service General Hospital Approval number: TSGHIRB No.: 2-107-05-086

Declaration of SPs consent

The authors certify they have obtained all appropriate consent forms. During data entry and analysis, the SPs were given different code numbers which could not be traced to their identification to ensure confidentiality and privacy.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. Barrows HS, Abrahamson S. The programmed patient: A technique for appraising student performance in clinical neurology. J Med Educ 1964;39:802-5.
2. Levine AI, Swartz MH. Standardized patients: The "other" simulation. J Crit Care 2008;23:179-84.
3. McGraw RC, O'Connor HM. Standardized patients in the early acquisition of clinical skills. Med Educ 1999;33:572-8.
4. Cohen DS, Colliver JA, Marcy MS, Fried ED, Swartz MH. Psychometric properties of a standardized-patient checklist and rating-scale form used to assess interpersonal and communication skills. Acad Med 1996;71:S87-9.
5. Colliver JA, Swartz MH, Robbs RS, Lofquist M, Cohen D, Verhulst SJ. The effect of using multiple standardized patients on the inter-case reliability of a large-scale standardized-patient examination administered over an extended testing period. Acad Med 1998;73:S81-3.
6. Reznick RK, Blackmore D, Dauphinée WD, Rothman AI, Smee S. Large-scale high-stakes testing with an OSCE: Report from the Medical Council of Canada. Acad Med 1996;71:S19-21.
7. Boulet JR, Smee SM, Dillon GF, Gimpel JR. The use of standardized patient assessments for certification and licensure decisions. Simul Healthc 2009;4:35-42.
8. Medical Council of Canada. Medical Council of Canada Qualifying Examination Part II (MCCQE Part II). Medical Council of Canada; 2008.
9. Federation of State Medical Boards, Inc., and National Board of Medical Examiners. United States Medical Licensing Examination: Step 2 Clinical Skills (CS) Content Descriptions and General Information. Federation of State Medical Boards, Inc., and National Board of Medical Examiners; 2008.
10. National Board of Osteopathic Medical Examiners. Bulletin of Information. National Board of Osteopathic Medical Examiners; 2008.
11. Liu KM, Tseng HM. The application of a high-stakes OSCE in a national medical licensure examination. Natl Elite 2010;2:111-29.
12. Van der Vleuten CP, Swanson DB. Assessment of clinical skills with standardized patients. State of the art. Teach Learn Med 1990;2:58-76.

13. Tamblyn RM. The Use of Standardized Patients in the Evaluation of Clinical Competence: The Evaluation of Selected Measurement Properties. Doctoral Thesis. Montreal, Quebec, Canada: McGill University, Department of epidemiology; 1989.
14. Tamblyn RM, Klass DJ, Schnabl GK, Kopelow ML. The accuracy of standardized patient presentation. Med Educ 1991;25:100-9.
15. Tamblyn RM, Klass DJ, Schnabl GK, Kopelow ML. Sources of unreliability and bias in standardized-patient rating. Teach Learn Med 1991;3:74-85.
16. Barrows HS. An overview of the uses of standardized patients for teaching and evaluating clinical skills. AAMC. Acad Med 1993;68:443-51.
17. Adamo G. Simulated and standardized patients in OSCEs: Achievements and challenges 1992-2003. Med Teach 2003;25:262-70.
18. Carney PA, Dietrich AJ, Freeman DH Jr., Mott LA. A standardized-patient assessment of a continuing medical education program to improve physicians' cancer-control clinical skills. Acad Med 1995;70:52-8.
19. University of Toronto. Faculty of Medicine. Available from: http://www.spp.utoronto.ca/StandardizedPatients. [Last retrieved on 2018 Feb 08].
20. Blue AV, Chessman AW, Gilbert GE, Mainous AG 3rd. Responding to patients' emotions: Important for standardized patient satisfaction. Fam Med 2000;32:326-30.
21. Robins LS, Zweifler AJ, Alexander GL, Hengstebeck LL, White CA, McQuillan M, *et al*. Using standardized patients to ensure that clinical learning objectives for the breast examination are met. Acad Med 1997;72(10):S91–3.
22. Tsai SS, Yang CW, Yeh CC, Chang SC. The recruitment and training standardized patient in National Taiwan university hospital. J Med Educ 2007;11:174-81.
23. Kao SB. Mastering Standardized Patients. Taiwan: Leader Book Co. Ltd.; 2013. p. 178-9.
24. Hazelkorn HM, Robins LS. Actors play patients. Using surrogate patients to look into private practice. Public Health Rep 1996;111:129-32.
25. University of Pittsburg. Available from: http://www.omed.pitt.edu/standardized/. [Last retrieved on 2017 Jun 10].
26. Chang BY, Shieh JY, Wang JC, Sheen MJ, Kao SY, Chang SY, *et al*. Evaluating the Clinical Competence of Emergency Medicine Interns Using Standardized Patients. J Med Educ (醫學教育) 2007;11:27-38.
27. Shieh JY, Chang BY, Kao SY, Wang JC, Ho ST, Chang SY. Standardized Patient and Medical Education. J Med Educ 2004;8(3):243-54.
28. The Standardized Patient Program: University of Texas Health Science Centre San Antonio; 2000.
29. Wang WD. Using standardized patient for a tool for teaching, learning and assessment. J Med Educ 1998;2:378-85.
30. Liu KM, Tsai SS. The summary of the first Taiwan national medical clinical skills examination. Kaohsiung Med Univ Enews 2013;229;10-1.
31. Vincent C, Young M, Phillips A. Why do people sue doctors? A study of patients and relatives taking legal action. Lancet 1994;343:1609-13.
32. Beckman HB, Markakis KM, Suchman AL, Frankel RM. The doctor-patient relationship and malpractice. Lessons from plaintiff depositions. Arch Intern Med 1994;154:1365-70.
33. Levine AI, Schwartz AD, Bryson EO, Demaria S Jr., Role of simulation in U.S. physician licensure and certification. Mt Sinai J Med 2012;79:140-53.
34. Liu KM, Kao MC. The implementation of OSCE in Taiwan national licensing examination: Past, present and perspectives. Natl Elite 2016;44:104-12.
35. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. Br Med J 1975;1:447-51.
36. Stillman PL, Burpeau-Di Gregorio MY, Nicholson GI, Sabers DL, Stillman AE. Six years of experience using patient instructors to teach interviewing skills. J Med Educ 1983;58:941-6.
37. Colliver JA, Williams RG. Technical issues: Test application. Proceedings of the AAMC's consensus conference on the use of standardized patients in the teaching and evaluation of clinical skills. Acad Med 1993;68:454-60.
38. Hartshorne JK, Germine LT. When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. Psychol Sci 2015;26:433-43.
39. University of Kentucky School of Medicine. Available from: https://meded.med.uky.edu/standardized-patient-professional-standards. [Last retrieved on 2017 Jun 26].