A Study on the Standard Setting, Validity and Reliability of the Standardized Patient Performance Rating Scale – Student Version

Ipek Gonullu (✉ zipgonullu@yahoo.com)

Ankara University School of Medicine <https://orcid.org/0000-0001-6333-6087>

Celal Deha Dogan

Ankara University Faculty of Education

Sengul Erden

Ankara University Faculty of Medicine

Research article

Keywords: Medical education, performance rating scale, standard setting, standardized patient, validation

Posted Date: October 14th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-89934/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

A standardized patient is a healthy person, trained to role-play a patient who gives feedback on the performance of the student. During the clinical skills training of students, the performance of a standardized patient is essential for the effectiveness of clinical education.

Methods

In this study, we have developed and determined the psychometric properties and standard-setting of the 9-item "Standardized Patient Performance Rating Scale – Student Version," which is designed for assessing the individual performance of standardized patients. For the scale development process, 702 medical students and seven educators participated in the study. Exploratory and confirmatory factor analyses were performed. Cronbach's alpha internal consistency coefficient and split-half reliability coefficient were calculated. For the standard-setting study, the extended Angoff method was used.

Results

As a result of the exploratory factor analysis, the scale had a single-factor structure as confirmed by confirmatory factor analysis. Cronbach's alpha internal consistency coefficient was calculated as 0.91. The score of a standardized patient at the borderline was 24.11 out of 45.

Conclusions

"Standardized Patient Performance Rating Scale – Student Version" is a valid and reliable scale for assessing standardized patients' performances, and can be used for the standardized patient selection and evaluation process in the health sciences.

Introduction

The changes in healthcare service delivery, developments in the measurement and evaluation process in education sciences, and humanistic and ethical education principles have led to a search for new ways of teaching clinical skills. Consequently, the use of a standardized patient (SP), a method of learning through simulation, has become mainstream. Using SPs in health education was first implemented in 1963 by Howard Barrows and has become a pillar in today's health education disciplines [1]. SPs are trained to role-play a patient and give feedback on the performance of the student who has examined that patient [2,3]. SP encounters are exceptionally well suited for teaching clinical skills and assessing student performance [4]. SPs help students acquire history taking and physical examination skills, improve their clinical reasoning and communication skills, and gain the ability to integrate them into their existing skills. The SP encounter provides medical students with equal learning opportunities, a learning environment in line with adult learning principles, and the flexibility to implement different learning styles.

SP encounters in a safe environment allow students to overcome the fear of harming a patient before they are required to encounter actual patients [5,6].

The domain of SP performance can be summarized as follows [7]:

- Ability to portray a patient; the SPs must give an accurate medical history; realistically depict the patient's educational level, psychological state, and emotional condition.
- Ability to observe medical student's behavior; the SPs must be able to observe the student's performance while performing the case.
- Ability to recall the encounter; the SPs must recall the details of the student's behavior after the encounter.
- Ability to give feedback; the SPs must give thoughtful, beneficial, and effective feedback on the student's communication skills and interaction style from the point of view of the patient SP was portraying.

In our university, medical students practice interviews (taking a medical history and informed consent, and communication skills) with SPs. During these sessions, a faculty watches the interview behind a mirror glass. After each encounter, both SP and the faculty immediately give oral feedback to the student about his/her performance. All scenarios used during SP interviews are about a patient with an emotional strain, and the students are expected to understand patients' perspectives.

SP-student interactions are usually in the form of patient-doctor interactions. At the end of each interaction, the SP gives feedback on the student's performance. It is important to examine SP's performance to increase the effectiveness of education. In this sense, assessing individual performances of SPs in an educational setting and assisting them in decreasing their weaknesses can improve the quality of the clinical skills training program. Among many stakeholders who can assess SP performance, faculty, SP trainers, and students are the most relevant since they interact with SPs more than others do. Besides faculty and SP trainers, students' assessments about SPs will provide straightforward and intuitive multi-source feedback to help SPs identify their strengths and developmental needs. As students interact with SPs one-on-one during the encounters, it is essential to learn students' perspectives immediately after the encounters. Because more than one student assesses each performance of an SP at different times, it is possible to get comprehensive information about SP performance. Moreover, the progress of SP performance can be monitored through time.

Although SPs are widely used in health sciences education, studies focusing on their performance are rare and limited in literature and are mostly focused on accurate portrayals of case specifics [8,9]. The literature analysis reveals that there are very few studies using a standard scale to assess SP performance [10-12]. However, there is not a single study that has developed a standard scale specifically for students to assess the quality of SP performance after they interact with them.

When developing these standard scales, it is crucial to set cut-off scores. The standard setting is the methodology for defining achievement and proficiency levels and for identifying cut-off scores corresponding to those levels [13]. If the cut-off scores are not appropriately set, the results of the assessment could be questioned. For that reason, standard-setting is a critical component of the test development process [14].

Consequently, developing a valid and reliable scale that students can use for evaluating SPs' performance in an educational setting and implementing a standard-setting study for this scale will make a significant contribution to both the literature and the standardized evaluation of SP performance. This study aimed to develop and determine the psychometric properties of the "Standardized Patient Performance Rating Scale – Student Version" (SPS-S) and implement a standard-setting study.

Methods

Study Design

The model for this study was methodological research by which the validity and reliability of the developed SPS-S were examined and implemented as a standard-setting study. The study was approved by Ethics Committee of Ankara University School of Medicine.

Participants

In this study, there were two groups of participants: Sample 1 was used for scale development, and Sample 2 was used for the standard-setting study of the scale.

Sample 1: For the scale development process, 702 second and third-year medical students enrolled in a public university were utilized. Those students had previous SP encounter experiences, and a rater training was performed before the SP encounters. The numbers of participants (sample size) were decided based on the assumptions of statistical methods. Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) were computed to test the construct validity of SPS-S. These are the multivariate statistics, which require big sample sizes. According to Comrey & Lee, a sample size of 200 is fair, and a sample size of 300 is suitable for EFA [15].

Moreover, at least 300 cases are needed with low commonalities, a small number of factors, and just three or four indicators for each factor [16]. As EFA and CFA should be conducted with two different groups selected from the same population, we divided participants into two groups. One group is for EFA (N=307), and the other is for CFA (N=395).

Sample 2: The standard-setting study of the scale was carried out with a test-centered approach, and expert opinions were collected in this context. Experts who had at least five years of experience in using and training SPs were selected by criterion sampling and maximum variation sampling methods. In this study, the variation factors were being an SP trainer, using SPs in student training, and working in different departments (Table 1).

As the research is partly a scale development study, the development stages of the scale and information on the "standard-setting form" are explained in this section.

SPS-S Development Steps

The scale development process consisted of seven steps, which included a literature review, conducting interviews, synthesis of the literature review and interviews, developing items, consulting expert validation, preliminary application, and pilot testing [17].

Literature Review: At this stage, the psychological attributes to be measured were defined, and a literature review was performed to determine which characteristics should be present in the SP. During this stage, the research focusing on the development of the measurement tool for SP performance evaluation was investigated [10,11].

Conducting Interviews: The interviews were conducted with seven tutors, 50 students, and two field experts who participated in SP practices. They were asked what they considered to be the main attributes of the good and poor performance of the SP. They focused on the different performance characteristics like persuasion in the role, successful portrayal, respecting the scenario, and giving effective feedback. Both written and verbal answers were collected from these interviews.

Synthesis of the Literature Review and Interviews: The data from the literature review and the interviews were evaluated together. The scope and content of the measurement tool intended to be developed were determined. Nine indicators of performance (1- Persuasiveness of acting, 2- Portraying of a patient, 3- Acting according to the scenario, 4- Using communication skills while giving feedback, 5- Competency in giving feedback, 6- Recalling the encounter, 7- Efficacy of the feedback, 8- Professional attitude while giving feedback, 9- Observing the student performance) were identified for scaling [7].

Developing Items: An item pool was created, consisting of 18 items, and two items were assigned to each indicator (Appendix A). Two items for each SP indicator were assigned to prevent the narrowing of the scope of the scale in a situation where an item was removed as a result of expert opinion or item analysis.

All developed items were positively worded. A five-point Likert-type scale was determined in consultation with the experts. The response anchors of the items were defined as "poor (1)", "fair (2)", "good (3)", "very good (4)", and "excellent (5)". Following these steps, a draft version of SPS-S was formed.

Consulting Expert Validation: To obtain an opinion on the 18-item draft scale, seven experts working in the field, four faculty experienced in using and training SPs, two linguists, and one measurement-evaluation specialist were consulted. These experts examined the scale items in the context of content, scope, language, comprehensibility, measurement, and evaluation principles using an evaluation form. On the form, the experts stated their opinions on each item as "applicable," "not applicable," "needs

revision," and included their recommendations for these items. Based on the recommendations from the experts, six items were excluded, and one item was revised; thus, a 12-item pilot version of the scale was created (Appendix B).

Preliminary Application: At this stage, the scale was applied to a group of 81 students to determine the approximate duration of implementation and make changes where necessary. As a result of the preliminary application, no item could be misunderstood, none of the items was left unanswered, the instructions were comprehensible, and the evaluation of an SP took three to five minutes.

Pilot Testing: The 12-item pilot form was applied to a large group (N=702), and the validity and reliability studies were performed. As a result of these analyses, the scale was finalized. Relevant findings are presented in the "Data Analysis" section.

Standard-Setting Form

In this study, the extended Angoff method was used to determine the cut-off score for SPS-S. In this method, the experts estimated the number of scale points that they believe borderline examinees would obtain from each response item [13]. In this context, experts determined the level of performance of the SPs at the borderline by using the "Standard-Setting Form" for each item in the scale. In the "Standard-Setting Form," two sections specify the level of performance of the SP, which is at the borderline for each item. The experts complete one of these sections in the first evaluation session and the other in the second evaluation session after carrying out discussions between the two sessions.

Data Collection

Scale Development Data: SPS-S was applied to the participants following the preliminary application. For this purpose, a group of second- and third-year students enrolled during the 2015-2016 and 2016-2017 academic years was utilized. The students were asked to respond to SPS-S immediately after encountering SPs because it was thought that they would evaluate patients more accurately. The students were informed about the study, and their consent was obtained before the interview with SPs. Care was taken for the students to complete the scale alone without any interaction with their peers or others. Twenty-five SPs participated in the study, 19 females and six males, aged between 32 and 65 years. Each student evaluated the SP he/she interviewed one time during communication skills training only.

Data for Standard Setting: First, the experts were trained by a specialist in the field of measurement and evaluation. During this training, the aim and methods used in the standard-setting were explained, and information was given about the procedures to be performed on the standard-setting method. In the next step, the experts discussed the characteristics they considered should be present in the SP and agreed on the level of competence that a patient at the borderline should have. Then, the first evaluation session began, and the experts were asked to give the scores (min: 1, max: 5) for each item by an SP at the borderline. This process was carried out individually. They were then asked to share their evaluations with

the group and justify them. Then, the experts stated their opinions about each other's evaluations, and a discussion was held. This discussion was followed by the second session, in which the experts were asked once again to provide the scores for the SP at the borderline. The first session, the discussion, and the second session took approximately one hour.

Data Analysis

The Validity of the Scale

Structure Validation: Before EFA, Kaiser-Meyer-Olkin (KMO) and Bartlett test variables were analyzed, and the data was tested for its appropriateness for the factor analysis. The principal components method was used in the factor selection process. In the determination of the number of factors, scree plot and parallel analysis methods were used. Since the scale had a single factor structure, no rotation method was used.

In the CFA process, initially, the data were tested for the multivariate normal distribution assumption. As the data did not satisfy this assumption, the analysis was carried out based on the weighted least squares method, and the standardized coefficients, corresponding t values, and some fit indices were evaluated.

Item Discrimination: In order to assess the item discrimination, the significance of the differences between the scores of the participants in the upper and lower 27% groups for each item was compared using the Mann Whitney U-test. Since the scores given for each item were within the ranking level, parametric tests (t-test for independent groups, etc.) were not used in this comparison.

Reliability of the Scale

In order to assess the reliability of the developed scale, Cronbach's alpha [18], internal consistency coefficient, and split-half reliability coefficient [19] were calculated.

Standard-Setting Study: An adaptation of the Angoff method for items with more than two possible scores is called the extended Angoff method [20]. Candidates at the borderline are those at the sufficient-insufficient border and those who are considered barely sufficient. Using the extended Angoff method, the experts decided the scores for each item of the SP at the borderline and recorded these estimates. The mean of the estimates given by the experts was calculated for each item. The sum of means gave the cut-off score.

During the data analysis process, SPSS 21.0, Lisrel 8.7, Excel 2016, and Monte Carlo PCA for Parallel Analysis packages were used.

Results

Structure Validation Studies

Exploratory Factor Analysis: EFA was carried out on 307 participants. The KMO value for this study was calculated as 0.92. The Bartlett's test of sphericity was applied to determine if the data met the multivariate normality assumption. The statistical significance of the calculated chi-square value can be interpreted as the distribution that meets the multivariate normality assumption with the factors that can be deduced from the correlation matrix [21].

The results of the Bartlett test indicate that a chi-square value of less than 0.05 is significant, which shows that a multivariate normality assumption has been met, and the factors can be deduced from the correlation matrix [22].

As a result of EFA, the scale had three factors with an eigenvalue greater than 1. When more than 200 samples are reached, it is recommended to examine the scree plot according to the eigenvalues for determining the number of important factors [23]. There is a significant deceleration in the Scree plot after the first factor, and the rate of deceleration decreases and follows a horizontal course after the second factor. Moreover, the eigenvalue of the first factor (5.432) is approximately five times greater than the eigenvalue of the second factor (1.082) before the rotation. The first factor alone, yielding a high variance (45%), was interpreted as the scale having a single-factor structure. The parallel analysis for the factor number also supports the single-factor structure. The Scree Plot is presented in Fig.1.

After establishing that the scale had a single-factor structure, the analysis was performed once again. The single-factor structure explained 59% of the total variance. Since the scale had a single-factor structure, no rotation was performed. In the analysis process, three items with a factor loading value below 0.40 (item 4, 0.19; item 5, 0.18, and item 7, 0.23), were excluded starting from the item with the smallest factor loading and items were renumbered. The factor loading values for the rest of the scale are presented in Table 2.

According to the result, SPS-S had a single factor structure and consisted of nine items.

Confirmatory Factor Analysis: CFA was performed for the verification of the single-factor structure resulting from the EFA of SPS-S developed within the scope of the study. In CFA analysis, the latent variable is the SP performance. This latent variable is abbreviated as "PERFORMAN." Observed variables are items of SPS-S abbreviated as S1 to S9. All variables entered into the model are displayed in Fig. 2 and Fig. 3.

As a result of CFA on the single-factor structure of SPS-S, the t values of the latent variables related to observed variables were greater than the critical value (2.58) and statistically significant at the 0.01 level. Fig. 2 and Fig. 3 show the standardized coefficients and t values for the relationships in the model, respectively.

In the analysis, the software suggested that the errors related to the items s1 and s2 should be associated and that this association could result in a decrease of 26.91 in the chi-square value. This theoretically reasonable modification was accepted, and errors of items s1 and s2 were associated. After that, the ratio

of chi-square value to degree of freedom was calculated as 1.81. This value can be considered as an indicator of a perfect fit [16]. Other calculated indices are as follows: NNFI 0.97, CFI 0.98, GFI 0.97, AGFI 0.96, RMSEA 0.04, and SRMR 0.04. When the results are examined, the values of all fit indices are indicative of a perfect fit [24].

Item Discrimination: For each item included in the scale, the mean ranks were higher in favor of the group in the upper 27%, and these differences were significant ($p < 0.01$).

Reliability Studies

Cronbach's alpha internal consistency coefficient was calculated as 0.91, and the split-half reliability coefficient as 0.87. These findings show that the internal consistency coefficient of the scale is at the desired level.

Findings Regarding the Cut-Off Score Obtained by the Extended Angoff Method

In order to determine the cut-off score of SPS-S (nine items), seven experts were consulted. The experts were asked to use the extended Angoff method and give a score between 1 and 5, taking into account an SP at the borderline for each item. This scoring was undertaken in two rounds (R), as shown in Table 3.

When the cut-off scores for each item were examined after the second round, the experts allocated the lowest cut-off point to item 4 (2.857) and the highest cut-off point to item 9 (4.14). Also, the experts stated that an SP at the borderline should have an average of 3.44 points from each item. The standard deviation values were examined to determine the variability between the experts' scores, and the variability between expert opinions was less in round 2 (0.53) than in round 1 (0.59).

The mean of the total points of each expert for each item was taken to calculate the cut-off score. Then, the means were summed to obtain the cut-off score.

Cut-off score = $3.00 + 3.78 + 3.56 + 3.33 + 4.00 + 3.11 + 3.33$

Cut-off score = 24.11

Therefore, according to the findings from the extended Angoff method, for an SP to be sufficient, he/she must obtain at least 24.11 out of 45 from SPS-S.

Discussion

SPs are used for teaching, practicing, and assessing a wide range of communication skills, counseling, and clinical/physical examination skills in medicine. In this context, the authenticity of role-play and quality of feedback provided by SPs is of high importance for the quality of learning during SP contact learning sessions. Also, Objective Structured Clinical Examination (OSCE) that use SPs are subject to many measurement errors based on SPs performances [25]. Therefore, training of SPs and following their

performance and development are important. A standard scale can be used by stakeholders to assess and follow the development of SP performance.

In this study, SPS-S was developed, and validity, reliability, and standardization studies were performed. This scale is only for students who can assess the quality of SP performance after the interaction. As a result of EFA, it was determined that the scale had a single-factor structure confirmed by CFA. The internal consistency of the scale was shown to be at the desired level by the reliability analysis.

The scale consists of nine items scored out of 5, meaning that the lowest achievable score is 9 and the highest is 45. The borderline score of an SP was 24.11 out of a total of 45 determined by the extended Angoff method. In this context, for someone to successfully qualify as an SP, he/she must obtain at least 24 points from SPS-S.

Students are important stakeholders during the assessment of an SP performance. As students interact with SPs one-on-one during the encounters, it is essential to learn their perspectives to obtain straightforward and intuitive multi-source information about the SP performance. The assessment of the SP performance by students might have certain advantages compared to the assessment done by other stakeholders. For example, several students can assess SPs, whereas only a few faculty and even fewer number of SP trainers can. Furthermore, students can assess SPs at different times, so it is possible to monitor SP performance progress through time. Moreover, this scale can be used to identify SPs who need further training early on by picking out the ones who scored less than 20, which will bring efficiency to the SP training and development process.

The strength of this study is that the SPS-S is not case-specific and can be used for various scenarios. Moreover, it is short (9 items) and can easily be completed by students. This scale was personalized for students; it contains items relating to their interactions with SPs and feedback about their performances.

The limitation of the study is that the criterion validity could not be tested because there is no other reliable and valid scale to assess SP performance by students. However, in similar studies, SPS-S can be used as a criterion scale to test validity. Another limitation could be the presence of a recall bias amongst students during their evaluation of the SP performance.

Conclusion

This study provides a valid and reliable instrument to assess SP performance by students. The use of SPS-S, which has been confirmed for validity, reliability, and standard-setting studies, will guide SP trainers during the training. It will also help SPs assess their weaknesses on an individual performance level. For further studies using SPS-S, we recommend the researchers to re-assess validity and reliability using CFA and internal consistency coefficient. Also, the scale can be modified for other stakeholders who will use it to assess SP performance.

In conclusion, this scale can be used for the selection and evaluation of SPs in the field of health sciences.

Abbreviations

Standardized Patient - SP

Standardized Patient Performance Rating Scale – Student Version - SPS-S

Exploratory Factor Analysis - EFA

Confirmatory Factor Analysis – CFA

Kaiser-Meyer-Olkin - KMO

Objective Structured Clinical Examination – OSCE

Declarations

Ethics approval and consent to participate

The study was approved by Ethics Committee of Ankara University School of Medicine.

Approval number: 01-15-16

A written informed consent was obtained from study participants.

Consent for publication

Not applicable

Availability of data and materials

The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests

Funding

This work was supported by the Scientific Research Project Committee (BAP) of the Ankara University under Project ID 16A0230005. The funding body is responsible with providing payment to SPs.

Authors' contributions

The study was designed and developed by IG, DCD and SE. SE was responsible for data collection. DCD was responsible for data analysis interpretation was undertaken by IG, DCD and SE. IG undertook the preliminary drafting of the paper, which IG, DCD and SE revised together and approved the final version of the manuscript.

Acknowledgements

Not applicable

References

1. Barrows HS. Simulated (standardized) patients and other human simulations: A comprehensive guide to their training and use in teaching and evaluation. Chapel Hill, NC: Health Sciences Consortium;1987
2. Barrows HS. An overview of the uses of standardized patients for teaching and evaluating clinical skills. *Acad Med*. 1993;68:443-443.
3. Cleland JA, Abe K, Rethans JJ. The use of simulated patients in medical education: AMEE Guide No 42. *Med Teach*. 2009;31(6):477-486.
4. Talwalkar JS, Cyrus KD, Fortin AH. Twelve tips for running an effective session with standardized patients. *Med Teach*. 2020;42(6):622-627.
5. Becker KL, Rose LE, Berg JB, Park H, Shatzer JH. The teaching effectiveness of standardized patients. *J Nurs Educ*. 2006;45(4):103-111.
6. Bokken L, Rethans JJ, Jöbsis Q, Duvivier R, Scherpbier A, van der Vleuten C. Instructiveness of real patients and simulated patients in undergraduate medical education: a randomized experiment. *Acad Med*. 2010;85(1):148-154.
7. Wallace P. Coaching standardized patients: For use in the assessment of clinical competence. New York: Springer Publishing Company; 2006.
8. Baig LA, Beran TN, Vallevand A, Baig ZA, Monroy-Cuadros M. Accuracy of portrayal by standardized patients: results from four OSCE stations conducted for high stakes examinations. *BMC Med Educ*. 2014;14(1):97.
9. Erby LA, Roter DL, Biesecker BB. Examination of standardized patient performance: accuracy and consistency of six standardized patients over time. *Pat Educ Couns*. 2011;85(2):194-200.
10. Wind LA, Van Dalen J, Muijtjens AM, Rethans JJ. Assessing simulated patients in an educational setting: the MaSP (Maastricht Assessment of Simulated Patients). *Med Educ*. 2004;38(1):39-44.
11. Bouter S, van Weel-Baumgarten E, Bolhuis S. Construction and validation of the Nijmegen evaluation of the simulated patient (NESP): assessing simulated patients' ability to role-play and provide feedback to students. *Acad Med*. 2013;88(2):253-259.
12. Perera J, Perera J, Abdullah J, Lee N. Training simulated patients: evaluation of a training approach using self-assessment and peer/tutor feedback to improve performance. *BMC Med Educ*.

2009;9(1):37.

13. Cizek GJ, Bunch M. Standard Setting: A Guide to Establishing and Evaluating Performance Standards on Tests. USA: Sage Publications; 2007.
14. Bejar II. Standard setting: What is it? Why is it important. R&D Connections, 2008;7:1-6.
15. Comrey A, Lee H. A first course in factor analysis. Hillsdale, NJ: Lawrence Erlbaum Associates; 1992.
16. Tabachnick BG, Fidell LS. Using multivariate statistics. Pearson Education Limited; 2013.
17. Artino Jr AR, La Rochelle JS, Dezee K J, Gehlbach H. Developing questionnaires for educational research: AMEE Guide No. 87. Med Teach. 2014;36(6):463-474.
18. Pallant J. SPSS survival manual. UK: McGraw-Hill Education; 2016.
19. Cronbach LJ. Essentials of physiological testing. New York: Harper & Row Publisher; 1984.
20. Hambleton RK, Plake BS. Using an extended Angoff procedure to set standards on complex performance assessments. Appl Meas Educ. 1995;8(1):41-55.
21. Çokluk Ö, Şekercioğlu G, Büyüköztürk Ş. Multivariate statistical SPSS and LISREL applications for social sciences. Ankara: Pegem Academy publications; 2010. (Original work published in Turkish).
22. Field AP. Discovering statistics using SPSS. 3rd ed. London: Sage Publications; 2009.
23. Stevens JP. Applied multivariate statistics for the social sciences. 5th ed. NY: Routledge; 2009
24. Hooper D, Coughlan J, Mullen MR. Structural equation modelling: Guidelines for determining model fit. Electron J of Business Research Methods. 2008;6(1):53-60.
25. Khan KZ, Ramachandran S, Gaunt K, Pushkar P. The objective structured clinical examination (OSCE): AMEE guide no. 81. Part I: an historical and theoretical perspective. Med Teach. 2013;35(9):e1437-e1446.

Tables

Table 1 Distribution of the study groups by department

Expert	Department	SP Trainer	Using SPs in Student Training
1.	Medical Education	+	+
2.	Medical Education	+	-
3.	Infectious Diseases	-	+
4.	Child Health and Diseases	-	+
5.	Psychiatry	-	+
6.	Radiology	-	+
7.	Forensic Medicine	-	+

Table 2 Factor loading values for the scale items

Items	Factor 1
Item 1	.729
Item 2	.744
Item 3	.789
Item 4	.720
Item 5	.773
Item 6	.723
Item 7	.812
Item 8	.815
Item 9	.824

Table 3 Expert Ratings Obtained from the Extended Angoff Method

Expert		Item Number										
		1	2	3	4	5	6	7	8	9	$\bar{X}$	sd
1	R1	3	4	4	3	3	2	3	4	4	3.33	0.70
	R2	3	3	2	2	3	3	3	4	4	3.00	0.70
2	R1	4	3	4	3	4	3	4	3	4	3.56	0.52
	R2	4	4	3	3	4	4	4	4	4	3.78	0.44
3	R1	4	4	4	2	3	3	4	3	4	3.44	0.72
	R2	4	4	4	3	3	3	4	3	4	3.56	0.52
4	R1	3	4	4	3	4	3	3	3	4	3.44	0.52
	R2	4	4	3	3	3	3	3	3	4	3.33	0.50
5	R1	4	4	4	4	4	5	4	4	5	4.22	0.44
	R2	4	4	3	4	4	4	4	4	5	4.00	0.50
6	R1	3	3	4	2	2	2	3	2	3	2.67	0.70
	R2	3	3	4	2	3	3	3	3	4	3.11	0.60
7	R1	3	3	3	4	4	3	4	3	4	3.44	0.52
	R2	3	3	3	3	4	3	4	3	4	3.33	0.50
$\bar{X}$	R1	3.43	3.57	3.86	3.00	3.43	3.00	3.57	3.14	4.00	3.44	0.59
sd	R2	0.53	0.53	0.38	0.82	0.79	1.00	0.53	0.69	0.58		
		3.57	3.57	3.14	2.86	3.43	3.29	3.57	3.43	4.14		
		0.53	0.53	0.69	0.69	0.53	0.49	0.53	0.53	0.38		

Figures

Figure 1

Line chart of eigenvalues

Chi-Square=47.05, df=26, P-value=0.00695, RMSEA=0.045

Figure 2

Standardized coefficients

Chi-Square=47.05, df=26, P-value=0.00695, RMSEA=0.045

Figure 3

t-values