ISSN: 2091-2749 (Print)
2091-2757 (Online)

Situational judgement test: Psychometric analysis of a pilot study for selecting post graduate medical student in residency program

Ashis Shrestha[1], Shital Bhandary[2], Shrijana Shrestha[3]

[1]Asst. Prof., Dept. of General Practice and Emergency Medicine
[2]Assoc. Prof., Dept. of Community Health Sciences
[3]Prof., Dean, School of Medicine, Patan Academy of Health Sciences, Lalitpur, Nepal

Abstract

Introductions: Selection process of post graduate medical students requires combination of knowledge, cognitive abilities and skills. Situational judgement test (SJT) is one on the important tools to measure cognitive abilities and skills. This study aims to measure the internal consistency reliability of SJT tool and item quality for local validation of the non-academic attributes. It also aims to do sub-group analysis of SJT scores based on this tool to provide further evidence for validity.

Methods: This cross-sectional study was conducted at Patan Academy of Health Sciences (PAHS) in May 2017. Two sets of SJTs were developed after iterative discussion and modification to suit the local context. Set A had five options to a given situation, students had to rank these options from most appropriate to least appropriate. Set B had seven options to a given situations, students had to pick three best options. Medical officers and interns who were working at PAHS as medical officers participated in this study.

Results: One hundred and sixteen medical officers participated in the study. Thirty-nine (33.6%) participants were interns who had graduated from PAHS and 77 (66.4%) were medical officers who had graduated from various institute inside and outside the Nepal. The overall Cronbach's alpha for 35 questions was 0.65, that for set A (19 questions) was 0.56 and that for set B (16 questions) was 0.52. Adding SJT questions improved value of Cronbach's alpha for SJT test as 40, 50, 60 and 70 questions gave Cronbach's alpha of 0.68, 0.73, 0.76 and 0.79. Average percentage correct of set A was 76.3% and that of set B was 60.5%. SJT scores were statistically different for medical graduates from Nepal and different universities within Nepal.

Conclusions: Locally developed situational judgement test is found to be reliable instrument for measuring non-academic construct of post-graduate medical entrance examination.

Keywords: Cronbach's alpha, internal consistency reliability, situational judgement test, validity

Correspondence
Dr. Ashis Shrestha
Department of General Practice and Emergency Medicine, Patan Hospital, Patan Academy of Health Sciences, Lalitpur, Nepal
Email: ashisshrestha@pahs.edu.np

Peer Reviewers
Prof. Dr. Jay N Shah
Patan Academy of Health Sciences

Prof. Dr. Nabees Man Singh Pradhan
Patan Academy of Health Sciences

Submitted
13 Jul 2019

Accepted
16 Sep 2019

How to cite this article
Ashis Shrestha, Shital Bhandary, Shrijana Shrestha. Situational judgement test: Psychometric analysis of a pilot study for selecting post graduate medical student in residency program. Journal of Patan Academy of Health Sciences. 2019Dec;6(2):81-7.

Introductions

The selection process of medical education should include combination of merit, equity, fairness and social accountability.[1] Situational judgement test (SJT) is a valid and reliable method for assessing a broad range of non-academic attributes in high-volume selection.[2] It can be designed to measure a variety of non-academic attributes beyond clinical knowledge, which are especially relevant in medical education, training and practice.[3] A systemic review shows that SJTs are more effective selection methods and are generally fairer than traditional interviews, references and personal statements.[4] We did not find any article in google scholar and pub med stating use of situational judgement test in post graduate education in Nepal.

Since SJT is a construct (latent variable) of non-cognitive domain, development and local validation of SJT is an important first step for implementing SJT in post graduate medical entrance examination. In this process, reliability analysis of the tool plays a crucial role as it is a necessary condition for gathering evidence for validity of SJT. Reliability analysis is usually done by conducting pilot study of locally developed tool in terms of internal consistency reliability (average inter-item correlation) and test re-test reliability (stability). Internal consistency reliability measures the average inter-item correlations among the items in the tool and higher value indicates good measurement of the construct under consideration i.e. SJT in this case. Stability, on the other hand, is measured by piloting the same tool with the same participants two times in short duration.

Thus, this pilot study is conducted to measure the internal consistency reliability of SJT tool and sub-group analysis of SJT score for maintaining equity and fairness in the selection process.

Methods

This cross-sectional study was conducted at Patan Academy of Health Sciences (PAHS) in May 2017. Two sets of SJT was administered to interns and medical officers working at PAHS. Ethical clearance was taken from institutional review committee of PAHS.

Development of Tool

Two sets of SJTs were developed by a group of experts identified by PAHS and each sets had twenty questions initially. After iterative discussions of each item among the experts, one question from set A and four questions from set B were removed as they were found not suitable for the local context. So, 19 questions (Q1 to Q19) were included in set A and 16 questions (Q21 to Q36) in set B (total 35 questions). Set A had five options to a given situation where students had to rank these options from most appropriate to least appropriate. Set B had seven options to a given situations, students had to pick three best options. Answer keys were developed based on common consensus of the group, which is the accepted method for SJT.[5] Scoring for set A was done using Pascal's triangle method where maximum score was 20 per question. So, maximum marks that a student can score in set A was 380. Scoring for set B is done based on true answer and each true answer was given score of four, so maximum score per question was 12. So, maximum marks that a student can score in set B was 192. The total marks of set A and B was 572.

Sample selection process

There were 56 interns and 76 medical officers working at PAHS. A notice was sent to all interns and medical officers for participation. Medical officers who had completed their undergraduate education from institutions other than PAHS was also be included. Participants registered for the exam voluntarily.

Variables and Data analysis

The first part of the SJT included general information like: gender, age, number of years worked after graduating and graduating university. This data was used for sub-group analysis using T test and ANOVA. Second part had answer sheet, each item in the answer

sheet was analyzed for internal consistency reliability (Cronbach's alpha) of SJT tool. Test retest reliability testing was not done in this study. P value of less than 0.05 was taken as statistically significant.

Results

Out of 132 interns and medical officers who were working at PAHS during this study, 116 participated in the study. Male participants were 63 (54.3%) and female were 53 (45.7%). Sixty participants (51.8%) were above 25 years of age and 75 (64.7%) had graduated within past two years. Thirty-nine (33.6%) participants were interns who had graduated from PAHS and 77 (66.4%) were medical officers who had graduated from various institute inside and outside the Nepal. There were 50 (43.1%) participants who were either PAHS graduates or interns. Participants from various countries were as follows: Nepal 87(75%); China 13 (11.2%); Bangladesh 12 (10.3%); Philippines 4 (3.4%).

Table 1. Mean, percentage correct, scaled mean if deleted, total correlation and Cronbach's alpha of individual item (n=116)

Q. No	Mean of each item	Standard deviation	% correct of each item	Scale mean if deleted	Coefficient of variation (%)	Corrected item total correlation	Coefficient alpha if item deleted
1	15.1	3.0	75	395.8	20.0	0.15	0.63
2	17.2	2.4	86.4	393.5	13.8	0.30	0.62
3	15.5	2.2	77.4	395.3	14.5	0.18	0.63
4	13.5	2.0	67.5	397.6	15.2	(0.05)	0.64
5	15.1	2.4	75.7	396.2	15.9	0.15	0.63
6	16.0	2.6	80.2	394.9	16.4	0.26	0.62
7	16.2	2.5	81.4	395.1	15.7	(0.09)	0.65
8	13.2	2.9	66.5	397.7	21.6	0.03	0.61
9	15.1	2.1	75.6	396.2	14.0	0.10	0.63
10	15.8	2.2	79.1	395.3	13.8	0.22	0.62
11	16.3	1.8	81.8	395.0	11.2	0.07	0.63
12	14.1	3.1	70.4	397.1	21.8	0.09	0.62
13	13.6	2.9	67.8	397.9	21.4	0.13	0.63
14	17.3	2.2	86.6	391.1	12.7	0.08	0.61
15	17.1	2.2	85.6	394.0	13.0	0.21	0.62
16	14.4	3.0	71.6	396.5	20.8	0.37	0.64
17	15.5	2.3	77.7	395.7	14.9	0.13	0.61
18	14.3	3.3	71.6	397.0	23.3	0.32	0.63
19	14.2	3.0	71.3	396.4	21.0	0.20	0.62
21	6.7	2.4	55.6	404.3	35.4	(0.07)	0.64
22	6.5	2.7	54.4	404.6	41.0	0.31	0.61
23	7.08	2.7	59.1	404.3	37.6	0.09	0.63
24	5.6	2.8	46.8	405.6	49.9	0.20	0.62
25	8.4	2.57	70.2	402.7	30.5	0.03	0.63
26	8.7	2.4	72.2	402.5	28.1	0.23	0.62
27	8.2	3.0	68.7	402.9	36.7	0.22	0.62
28	5.7	2.6	47.4	405.6	45.8	0.19	0.62
29	7.8	3.0	68.8	403.7	37.8	(0.06)	0.65
30	8.2	2.9	68.7	402.9	34.9	0.14	0.63
31	6.0	2.7	49.7	405.2	43.0	0.28	0.62
32	9.2	2.9	77.5	401.7	31.6	0.31	0.61
33	5.7	2.7	47.4	405.5	47.6	0.24	0.62
34	8.4	2.9	70.5	402.6	34.3	0.23	0.62
35	6.5	2.3	54.1	404.8	35.2	0.18	0.63
36	6.8	2.5	56.6	404.4	37.4	0.25	0.62

* Q1 to Q19 = Set A; Q21 to Q36=Set B, Q no 20 was not present in original set of questionnaire. Values in bracket are negative correlation.

Table 2. Comparison of scores with demographic variable (n=116)

Variables	Category	N	Set A	Set B	Total
Gender	Male	63	288.7	113.7	402.4
	Female	53	289.6	117.5	407.2
P value (T test)			0.767	0.171	0.327
Age	≤ 25 years	60	292.0	117.1	409.1
	>25 years	56	286.1	113.8	399.8
P value (T test)			0.058	0.245	0.059
Year since graduation	< 2 years	75	291.3	115.7	407.1
	≥ 2 years	41	285.2	114.9	400.1
P value (T test)			0.067	0.771	0.176

Table 3. Comparison of scores of participants from different countries and universities (n=116)

Variables	Category	N	Set A	Set B	Total
Designation	Medical officer	77	285.7	112.9	398.5
	Intern	39	297.1	120.7	417.8
P value (T test)			0.001	0.010	0.0001
PAHS	Medical officer	11	301.1	121.1	422.1
	Intern	39	297.1	120.7	417.8
P value (T test)			0.419	0.994	0.600
Country where medical graduation was done	Bangladesh	12	277.5	105.3	382.9
	China	13	275.6	108.6	384.3
	Nepal	87	292.5	117.7	410.6
	Philippines	4	287.5	119.0	406.5
P value (ANOVA)			0.0001	0.015	0.0001
University of graduation inside Nepal	KU	20	285.9	114.2	399.2
	PAHS	50	297.9	120.8	418.8
	TU	17	287.0	112.9	399.9
P value (ANOVA)			0.003	0.078	0.002

Participation from various universities inside Nepal were as follows: PAHS 50 (43.1%); Kathmandu University (KU) 20 (17.2%); Tribhuvan University (TU) 17(14.7%)

Two students did not answer question number 4 and 17 respectively so after excluding these two participants, Cronbach's alpha was calculated for 35 (total) questions that were responded by 114 participants. The overall Coefficient alpha (internal consistency reliability) for 35 questions was 0.65, that for set A (19 questions) was 0.56 and that for set B (16 questions) was 0.52. Spearman-Brown prediction formula suggested that adding questions could improve value of Coefficient alpha for SJT test with 40, 50, 60 and 70 questions as 0.68, 0.73, 0.76 and 0.79 respectively. There were four negatively correlated items out of total 35 items (Table 1).

Graded mean of score in set A was 15.2 which is 75% of the total score (20) and that of set B was 7.2 which is 60% of the total score (12). Average percentage correct of set A was 76.3% and that of set B was 60.5%. Mean score of, set A was 289.16 (SD 17.0) with minimum score of 244 and maximum score of 326; set B was 115.5 (SD 15.1) with minimum

score of 76 and maximum score of 152; aggregated mean score of set A and set B was 404. 6 (SD 26.2) with minimum score of 348 and maximum score of 476. The p value was more than 0.05 while comparing gender, age and year since graduation (Table 2).

Scores of graduates from Nepal was more than that scored by graduates from abroad like Bangladesh, China and Philippines. Similarly, scores of graduates from PAHS was higher than the graduates from Kathmandu and Tribhuvan universities for set A (Table 3).

Discussions

The overall value of Chronbach alpha for 35 questions was 0.65, that for set A (19 questions) was 0.56 and that for set B (16 questions) was 0.52. This means that 35 question included in the SJT test explained 65% variation of non-cognitive domain. The usual practice is to take Chronbach's alpha above 0.80 in internal consistency reliability analysis, however the values of 0.7 or or 0.6 were also seen to be acceptable in some studies when heterogeneous items were used.[6] A meta analysis which included 39 studies published between 1990 and 2011 on SJT reported the alpha from 0.03 to 0.60.[7] Similarly another study done by McDaniel et al[8] summarized that value of Coefficient alpha varied between 0.43 and 0.94. Coefficient alpha in our and various studies were below 0.80, this might be due to heterogeneity of the items used in SJT. Alpha value of 0.8 and higher is appropriate measure of internal consistency reliability where items are homogenous.[9] The items that we used in SJT were heterogeneous under various domains like, professionalism, communication, ethics and empathy. However, considering ranges of Coefficient alpha mentioned in various studies[6-8] the overall value of Coefficient alpha and the value for of set A and set B that we have obtained can be taken as acceptable for our context.

A study shows that type of response instructions influenced the internal consistency reliability. The Coefficient alpha is highest when candidates are asked to rate each response.[9] Questions used in set A had five options in each question, and every option had to be ranked in the order of sequence. Therefore, it is consistent with our observation of higher coefficient alpha of set A than that of set B, 0.56 vs 0.52. Similar response was observed in above study by Polyhart and Erhart[9] wherein candidates had to choose two responses had coefficient alpha of 0.60 while one response items had coefficient alpha of 0.24. This shows that questions used in set A has more internal consistency reliability. However, if numbers of question are increased in both set, alpha score will increase. This was confirmed by Spearman-Brown prediction formula result where coefficient alpha increased to 0.79 when SJT questions were increased to 70.

The corrected item-total correlation, also known as item quality/discrimination/partials, revealed four problematic items (Table 1: Qno. 4,7,21,29) as they had negative correlation with the total scores minus that item (construct). This means high scorers in SJT did poor in these four items because of ambiguity and low scorers did well as they guessed them well. Further, coefficient alpha increased substantially once these items are dropped from the test (Table 1). This means four items with negative correlation and 7 more items (Table 1: Q no 8,9,11,12,14,23,25) with item quality between 0 and 0.13 (similar items) must be dropped in the final SJT tests suggesting requirement of a large question bank for the effective use of this non-cognitive domain test in the future.

Analysis of SJT score revealed graded mean score of set A (15.2 out of 20 which is 75%) was higher than that of set B (7.2 out of 12 which is 60%). Average percentage correct in set A (76.3%) was also higher than that of set B (60.5%). Further coefficient of variation shows that set A questions were more homogenous (12.7-23.3) than set B (28.1-49.9), which suggest that set B questions were difficult but more discriminatory than set A.

There was no significant difference between the score of set A, set B and total score in the group male to female; age less than or equal to 25 to age more than 25; experience of less than two year and more than or equal to two years (Table 2). This means SJT used is fair to age, gender and experience. However, there are various studies[10,11] which shows that female outperforms male in SJT. Graduates from Nepal performed better than graduates from abroad, similarly graduates from PAHS performed better than graduates from other universities of Nepal (Table 3), this difference might only be due to educational method focusing on communication, ethics and empathy and level of exposure to different situations during their undergraduate training at PAHS and Nepal respectively. A study suggests that there is a significant relation between vocational interest of candidates and experience to such situtation.[11] This difference may not be seen when SJT is administered to the actual post-graduate applicants as the motivation of the candidates will be different than the participants in this pilot study, which will confirm the equity and fairness of this test.

We found that SJTs can be designed to test a broad range of non-academic attributes depending on the selected context. As a relatively low-fidelity assessment, SJTs are a cost-efficient methodology compared with high-fidelity assessments of non-academic attributes like multiple mini-interviews/admission OSCEs.[4] This study suggests that this tool has acceptable internal consistency reliability but can be improved based on item quality analysis, which in turn requires a good and large question bank. So, this study can be taken as baseline psychometric analysis of SJT for testing non-academic attributes in post-graduate medical education student selection in Nepal. Same principles can be applied for selecting undergraduate medical education too though SJT items development and validation must be done by experts who are familiar with those students.

The main limitation of this pilot study was the sample size, which can be one of the confounding factors though 116 out of 132 graduates and interns working at that time participated in the study. Another confounding factor can be a single centre study though 66 out of 116 participants had completed their graduation from various universities.

Conclusions

Situational judgement test is a reliable tool for measuring non-academic attributes in post graduate medical entrance examination. The internal consistency reliability of the tool can be improved by increasing the number of quality items and making it more homogenous.

Acknowledgements

Contributions that need acknowledging but do not justify authorship, such as general support by a departmental chair; Acknowledgments of technical help; and Acknowledgments of financial and material support, which should specify the nature of the support. This should be included in the title page of the manuscript and not in the main article file

References

1. Prideaux D, Roberts C, Eva K, Centeno A, Mccrorie P, Mcmanus C, Patterson F, Powis D, Tekian A, Wilkinson D. Assessment for selection for the health care professions and specialty training: consensus statement and recommendations from the Ottawa 2010 Conference. Med Teach. 2011;33(3):215-23. DOI PubMed GoogleScholar
2. Patterson F, Lievens F, Kerrin M, Zibarras L, Carette B. Designing selection systems for medicine: the importance of balancing predictive and political validity in high-stakes selection contexts. International Journal of Selection and Assessment. 2012;20(4):486-96. DOI GoogleScholar
3. Koczwara A, Patterson F, Zibarras L, Kerrin M, Irish B, Wilkinson M. Evaluating cognitive ability, knowledge tests and situational judgement tests for postgraduate selection.

Med Educ. 2012;46(4):399-408. DOI PubMed GoogleScholar
4. Patterson F, Ashworth V, Zibarras L, Coan P, Kerrin M, O'Neill P. Evaluations of situational judgement tests to assess non-academic attributes in selection. Med Educ. 2012;46(9):850-68. DOI PubMed GoogleScholar
5. McDaniel MA, Psotka J, Legree PJ, Yost AP, Weekley JA. Toward an understanding of situational judgment item validity and group differences. Journal of Applied Psychology. 2011;96(2):327-36. DOI GoogleScholar
6. Griethuijsen, RA, van Eijck MW, Haste H, den Brok PJ, Skinner NC, Mansour N, Gencer AS, BouJaoude S. Global patterns in students' views of science and interest in science. Res Sci Educ. 2015;45(4):581-603. DOI GoogleScholar
7. Catano VM, Brochu A, Lamerson CD. Assessing the reliability of situational judgment tests used in high-stakes situations. International Journal of Selection and Assessment. 2012;20(3):333-46. DOI GoogleScholar
8. Cabrera MA, Nguyen NT. Situational judgment tests: a review of practice and constructs assessed. International Journal of Selection and Assessment. 2001;9(1-2):103-13. DOI GoogleScholar
9. Ployhart RE, Weekley JA, Holtz BC, Kemp C. Web-based and paper-and-pencil testing of applicants in a proctored setting: are personality, biodata, and situational judgment tests comparable? Personnel Psychology. 2003;56(3):733-52. DOI GoogleScholar
10. Lievens F, Patterson F, Corstjens J, Martin S, Nicholson S. Widening access in selection using situational judgement tests: evidence from the UKCAT. Med Educ. 2016;50(6):624-36. DOI PubMed GoogleScholar
11. Schripsema NR, van Trigt AM, Borleffs JC, Cohen-Schotanus J. Impact of vocational interests, previous academic experience, gender and age on situational judgement test performance. Adv in Health Sci Educ Theory Pract. 2017;22(2):521-32. DOI PubMed GoogleScholar