

BRIEF

The Relationship Between a Multiple Mini-interview and Situational Judgment Test for Admissions

Wendy C. Cox, PharmD,^a Michael Wolcott, PharmD, PhD,^b Faustina Hahn, PharmD,^a
Jacqueline E. McLaughlin, PhD, MS^a

^a The University of North Carolina at Chapel Hill, UNC Eshelman School of Pharmacy, Chapel Hill, NC

^b High Point University, School of Dental Medicine and Oral Health, High Point, NC

Submitted February 7, 2022; accepted May 9, 2022; published March 2023.

Objective. To explore the relationship between a multiple mini-interview (MMI) and situational judgment test (SJT) designed to evaluate nonacademic constructs.

Methods. A 30-question ranked-item SJT was developed to test three constructs also measured by MMIs during a pharmacy school's admissions process. First-year pharmacy students were invited to complete the SJT in fall 2020. One hundred four students took the SJT (82.5% response rate), with 97 (77% of possible participants) having MMI scores from the admissions process. Descriptive statistics and other statistical analyses were used to explore the psychometric properties of the SJT and its relationship to MMI scores.

Results. Seventy-four percent of students identified as female (n=72), and 11.3% identified with an underrepresented racial identity (n=11). The average age, in mean (SD), was 21.8 (2.1) years. Students' mean (SD) scores were 85.5 (3.1) (out of 100 points) on the SJT and 6.1 (1.0) (out of 10 points) on the MMI. Principal components analysis indicated that the SJT lacked construct validity and internal reliability. However, reliability of the entire SJT instrument provided support for using the total SJT score for analysis ($\alpha = .63$). Correlations between total SJT and MMI scores were weak ($r_p < 0.29$).

Conclusion. Results of this study suggest that an SJT may not be a good replacement for the MMI to measure distinct constructs during the admissions process. However, the SJT may provide useful supplemental information during admissions or as part of formative feedback once students are enrolled in a program.

Keywords: situational judgment test, multiple mini-interviews, admissions, assessment

INTRODUCTION

Identifying and evaluating nonacademic constructs in students applying to health sciences programs is a primary goal of many admissions committees. Nonacademic constructs (also called social and behavioral constructs), such as empathy and integrity, are paramount to successful practice as a health care professional. Therefore, programs are exploring ways to accurately assess these constructs in applicants and in students as they progress through programs.¹⁻⁵

A situational judgment test (SJT) is an assessment technique that has gained popularity in health sciences

schools over the past several years as a method to measure social and behavioral aspects of students.^{6,7} An SJT is a written assessment tool where a case or scenario is presented and the test taker must rate the appropriateness of various responses to the scenario, written and validated by subject matter experts. Test takers may be asked to rank order responses from most appropriate to least appropriate or to select the best response. The time to administer an SJT is minimal, and it only requires one person to administer the test to a group; however, an SJT is time-consuming to develop, and there is not one commercially available test in health professions education. There are also different ways to design SJTs, resulting in varying outcomes and difficulty identifying and evaluating one construct of interest.⁶⁻¹⁵

Because of the ease of administration, SJTs may be less resource-intensive than administering multiple mini-interviews (MMIs), used for similar purposes in evaluating nonacademic constructs. Originally, MMIs were developed

Corresponding Author: Wendy C. Cox, PharmD, Associate Professor, Division of Practice Advancement and Clinical Education, Associate Dean for Admissions and Accreditation, UNC Eshelman School of Pharmacy, The University of North Carolina at Chapel Hill, Chapel Hill, NC. Tel: 919-966-9990. Email: wendy_cox@unc.edu

to evaluate nonacademic qualities of applicants to medical residency programs and have gained popularity in health professions education.¹⁶⁻²¹ In these evaluations, test takers are presented with a written case and have a few minutes to formulate a response. Then they enter a room with an evaluator and provide thoughts on the case. The purpose is for the evaluator to be able to identify and evaluate the construct of interest that the MMI is targeting. This method requires a lot of resources to plan and implement.^{16,22,23}

At the University of North Carolina Eshelman School of Pharmacy, we have used MMIs as part of our admissions process to the Doctor of Pharmacy (PharmD) program since 2015 to measure nonacademic constructs.²³⁻²⁵ The MMI has been a helpful tool in our admissions process; however, it is resource intensive, specifically regarding the faculty and staff time required to plan and administer the MMI. Our research team hypothesized that an SJT may be used to accurately identify the constructs of interest measured in our MMI, which could be used in admissions and would use fewer resources. The objective of this study was to explore the relationship between a MMI and an SJT designed to evaluate the same nonacademic constructs.

METHODS

We developed a 30-question ranked-item SJT to test three constructs also measured by MMIs during the school's admissions process: adaptability, empathy, and integrity (Appendix 1). Prior to the administration of the SJT, the scoring system was based on the responses from eight experts. The experts were faculty and postdoctoral fellows who were practicing pharmacists. The SJT was piloted with postdoctoral fellows to verify the functionality and requests for minor edits to improve the readability.

First-year pharmacy students were invited to complete the 30-item SJT in fall 2020. The SJT was administered via Qualtrics (Qualtrics International Inc) at the end of an orientation session. Participants were instructed to rank the five answer options in order of what they should do from most likely to least likely. Two of the questions from the empathy section were removed due to technical errors within Qualtrics. This led to a total of 28 questions on the SJT (eight empathy questions and 10 adaptability and integrity questions each). The SJT score was paired with the respective MMI score. The MMI data were extracted from the admissions office, which had the scores for each of the MMI stations (1=poor to 10=exceptional). Data from the MMI model at the school have previously been shown to have strong construct validity and high internal consistency.²³ The admissions data also included demographic data such as age, gender, and race/ethnicity.

Data were analyzed using descriptive statistics, non-parametric statistical tests, and psychometric analyses. Specifically, concordance analysis, Cronbach alpha, principal components analysis, correlation, and linear regression were used to explore SJT psychometric properties and the relationship of the SJT to the MMI scores. For principal components analysis, varimax rotation with Kaiser rule (ie, eigenvalue >1) was used to identify and retain factors. Analyses were done within R (The R Foundation for Statistical Computing), SPSS version 26 (IBM Corp), and Excel. This study was approved via expedited review by the University of North Carolina at Chapel Hill Institutional Review Board.

RESULTS

One hundred four students took the SJT (82.5% response rate), with 97 participants having MMI scores from the most recent admission cycle (some had MMIs from an Early Assurance Admissions Program, which were not used in this study because they were over one year old and not identical to the MMIs administered to the rest of the class). Seventy-four percent of students identified as female (n=72), and 11.3% identified with an underrepresented racial identity (n=11), and students' mean (SD) age was 21.8 (2.1) years (Table 1).

Students scored a mean (SD) of 85.5 (3.1) (out of 100 points) on the SJT and 6.1 (1.0) (out of 10 points) on the MMI. The principal components analysis indicated that the SJT lacked construct validity (ie, factored into more than the three constructs intended) and internal reliability

Table 1. Demographic Data of First-Year Pharmacy Students Who Participated in a Study to Explore the Relationship Between a Multiple Mini-Interview and a Situational Judgment Test Designed to Evaluate Nonacademic Constructs

	Students (N=97)
Gender, No. (%)	
Male	25 (25.8)
Female	72 (74.2)
Race, No. (%)	
White	72 (51.1)
Asian/Asian American ^a	25 (25.8)
Underrepresented racial minority ^b	11 (11.3)
Age, mean (SD), years	21.7 (2.1)

^a Includes South Asian, Cambodian, Chinese, Filipino, Japanese, Korean, Malaysian, Pakistani, Vietnamese, and participants with mixed races with the aforementioned groups.

^b Includes Hispanic, Cuban, Mexican, Puerto Rican, South American, Dominican, Native American, Black, and participants with mixed races with the aforementioned groups.

Table 2. Median and Interquartile Ranges of the Situational Judgment Test Using Different Scoring Mechanisms

	Rank ^a	T2 ^b	SB ^c	T2W ^d	B2 ^e	SW ^f	SBB ^g
Empathy	136 (130-140)	12 (11-13)	6 (5-7)	2 (1-2.85)	11 (10-12)	0 (0-0)	6 (5-6)
Integrity	178 (172-182)	16 (15-17)	7 (6-8)	0 (0-1)	18 (16-18.75)	0 (0-0)	8 (7-9)
Adaptability	166 (162-172)	14 (13-15)	6 (5-7)	2 (1-3)	16 (15-17)	0 (0-0)	8 (7-8)
Total SJT	478 (466.5-490)	42 (40-44)	19 (17-20)	4 (3-5)	44.5 (43-47)	0 (0-0)	22 (20-23)

Abbreviations: SJT=situational judgment test.

^a Rank refers to the method of giving partial credit based on ranking compared to key (score=0-20 per item).

^b In the T2 method, students identified the first- and/or second-ranked item; no specific order (score=0-2 per item).

^c SB refers to the single best option method; students got the first-ranked item as first ranked (score=0-1 per item).

^d In the T2W method, students listed a fourth- or fifth-ranked item in their top one or two rank INCORRECTLY (score=0-2 per item).

^e In the B2 method, students listed a fourth- or fifth-ranked item in their four or five rank CORRECTLY (score=0-2 per item).

^f In the SW method, students listed a fifth-ranked item in their top one rank INCORRECTLY (score=0-1 per item).

^g In the SBB method, students listed a fifth-ranked item in their five rank CORRECTLY (score=0-1 per item).

(ie, $\alpha < .4$ for each construct). Multiple principal components analyses were conducted after excluding items with low concordance ($W < 0.6$), yet the SJT scores continued to factor into more than the three constructs intended. However, the reliability of the entire SJT instrument provided support for using the total SJT score for exploratory analysis ($\alpha = .63$). Correlations between total SJT and MMI scores were weak ($r_p < 0.29$). Correlations of various scoring combinations (ie, top choice, top two choices, bottom two choices) were also conducted and resulted in weak to negligible correlations (Spearman rho range = -0.12 to 0.10) (Table 2).

When analyzed by demographic groups, notable differences were found by race and gender identity for the total SJT score. Female-identifying participants scored higher on SJT items than male-identifying participants by 2.67 points ($p < .001$) when controlling for all other variables in the model. Additionally, having an underrepresented racial identity (eg, Black, Latinx, Native American) was associated with a 1.83-point decrease in SJT score ($p = .03$), controlling for all other variables in the model (Table 3).

DISCUSSION

More health sciences programs are using SJTs in their admissions processes. It is important to understand how this assessment approach can be optimized in admissions and whether it is measuring the intended constructs of interest. In this study, the total MMI score was associated with overall performance on the SJT, although the correlation was weak. This suggests the assessments used in this study may be measuring similar constructs; however, they may be accomplishing that in a different way or providing different insights that require further exploration. For example, there may be an impact of providing potential response options in SJTs that greatly differ from the responses generated by participants during an MMI, which are then interpreted and evaluated by a rater. This research had differing results from other SJT publications in pharmacy and health professions education; a notable difference is that this SJT used a ranking response selection, whereas others often use a technique where examinees rate each response in terms of appropriateness (1 = inappropriate response to 5 = highly appropriate response).

Table 3. Linear Regression Results for Predicting Situational Judgment Test Score From Admissions Data and Student Demographics

	Model 1 (All SJT items) ^a			Model 2 (SJT refined) ^b		
	B (SE)	Beta	p value	B (SE)	Beta	p value
Intercept	79.65 (3.55)		<.001	78.90 (3.65)		<.001
Female	2.67 (0.62)	.39	<.001	2.17 (.64)	.32	.001
URM	-1.83 (0.85)	-0.20	.034	-1.40 (.89)	-.15	.12
Age	-.02 (0.13)	-.01	.90	.00 (.14)	.00	.98
MMI scores	.56 (.27)	.19	.04	.70 (.28)	.24	.01
Application review score	.14 (.11)	.11	.23	.12 (.11)	.10	.32

Abbreviations: SJT=situational judgment test; URM=underrepresented racial minority; MMI= multiple mini-interview.

^a Model 1 has 25 survey items, after three items were dropped (items with $p < .002$). $R^2 = .29$.

^b Model 2 has 23 survey items, after five items were dropped (items with $p < .01$). $R^2 = .23$.

This may have consequences to consider in future design of SJTs; however, assessment methods with bespoke designs, such as MMIs and SJTs, generate data unique to the program, and results should be interpreted accordingly.

An important demonstration from this research is the complexity in designing an SJT with high reliability and construct validity. The SJT created in this study failed to load into three distinct factors and demonstrated low internal consistency for each construct. This illustrates there are many factors that can influence participant response selections, which then influence performance and reliability. The construct-driven approach that concentrates on a theoretical focus on what is to be measured is often described as the optimal approach to SJT design; however, this process can be resource and time intensive and often results in a limited number of items being generated.^{10,11} We also illustrated when measuring multiple constructs that it can be difficult to create SJT items that readily distinguish one construct from another, which has been demonstrated with other assessment approaches like MMIs.²⁵ The time required to develop and design SJTs can be significant. It is estimated that it took 30-40 hours to develop this SJT plus the additional time required to pilot the test with experts and refine it. In our experience, it does not take as much time to develop MMI scenarios (approximately 10 hours) but takes much longer to administer them to candidates each interview day. The time spent in SJT development is upfront, whereas most of the time required for MMIs is in administering them. Design is a particular challenge that others must be aware of in this process, as it can limit the utility of the finalized instrument.

Another insight from this research was the subgroup analysis, which suggests that there may be issues in SJT fairness based on key demographic characteristics. Fairness is a critical aspect of admissions practices to ensure equitable access for all candidates. An advantage of SJTs has been evidence of enhanced fairness in scoring practices; in other words, groups are not disadvantaged based on their gender, racial identity, and other demographic aspects.²⁶ However, the SJT in this study had significant differences in performance based on gender and racial identity, which suggests there may be fairness concerns, and this may be influenced by item design. Further research would be needed to determine whether the instrument accurately identified actual differences or to identify which items were biased in some way. Overall score differences ranged between one to three points between the subgroups, which may not have practical significance if it was determined not to impact admissions decisions. It is also important to consider that SJTs often focus on non-academic attributes (empathy, adaptability, integrity, etc), which may be emphasized more depending on cultural

background and gender identity. For example, women are often taught more explicitly about empathy and other attributes as part of societal expectations, which may influence their performance compared to men. A student's SJT performance should be considered as highly contextual when interpreting results, and clarity is needed about whether the results may be affected by cultural or societal factors.

Part of this exploration included evaluating multiple scoring strategies (ie, single-best selection, single-worst selection, etc) and their correlation with MMI scores to determine whether the SJT had a value beyond identifying those with the highest standing on the construct. In other words, SJTs are often used to distinguish top-performing candidates and those more desirable for admissions. We considered whether the SJT may have value in identifying learners who instead would be *at risk* of not successfully completing the program rather than identifying those who are the optimal fit. For example, we investigated the correlation between MMI score and those who selected the worst option as the best (ie, their first-ranked response was a fifth-ranked response on the key). In this research, we did not identify any other patterns of scoring that may improve the correlation. However, it also illustrates that SJT performance and psychometrics may be highly dependent on scoring practices, which has been demonstrated in previous work in medical education.⁹

There are several limitations to this study. First, this study was limited to one institution and had a small sample size, especially those with underrepresented racial identities. Selection bias was present, as the pool only included students accepted to our PharmD program taking the SJT. Additionally, there was more than six months between the administration of the MMI and the SJT, which could have affected the results. The scoring system also had limitations, with the SJT scored in a ranked system versus a single-item choice. There could also be bias with the SJT key development, as different combinations of experts (hospital, community, academic pharmacists) could potentially influence the key. Furthermore, people perceive conflicts and situations differently in certain contexts, which could have affected the key and the responses. Lastly, the key may be oriented to what people believe is right as opposed to what would be the best response to a situation or conflict.

CONCLUSION

The results of this study demonstrated that the SJT lacked construct validity, and the correlation between the SJT and MMI scores were weak. Given these results, an SJT may not be a good replacement for the MMI to measure distinct constructs during the admissions process;

however, it may provide useful information in addition to the MMI during admissions or as part of formative feedback once students are enrolled in a program. Future research should explore the aspects of SJT design (eg, item development) and subsequent impact of using the SJT as a formative and longitudinal assessment strategy in the health professions.

REFERENCES

1. Lane I. Professional competencies in health sciences education: from multiple intelligences to the clinic floor. *Adv Health Sci Educ Theory Pract.* 2010;15(1):129-146. doi:10.1007/s10459-009-9172-4
2. Choi A, Flowers S, Heldenbrand S. Becoming more holistic: a literature review of nonacademic factors in the admissions process of colleges and schools of pharmacy and other health professions. *Curr Pharm Teach Learn.* 2018;10(10):1429-1437. doi:10.1016/j.cptl.2018.07.013
3. Berwick D, Finkelstein J. Preparing medical students for the continual improvement of health and health care: Abraham Flexner and the new "public interest." *Acad Med.* 2010;85:S56-S65. doi:10.1097/ACM.0b013e3181ead779
4. Warm E, Kinnear B, Lance S, Schauer D, Brenner J. What behaviors define a good physician? Assessing and communicating about noncognitive skills. *Acad Med.* 2022;97(2):193-199. doi:10.1097/ACM.0000000000004215
5. Speedie M, Baldwin J, Carter R, Raehl C, Yanchick V, Maine L. Cultivating 'habits of mind' in the scholarship pharmacy clinician; report of the 2011-2012 Argus Commission. *Am J Pharm Educ.* 2012;76(6):S3. doi:10.5688/ajpe766S3
6. Patterson F, Zibarras L, Ashworth V. Situational judgement tests in medical education and training: research, theory and practice: AMEE Guide No. 100. *Med Teach.* 2016;38(1):3-17. doi:10.3109/0142159X.2015.1072619
7. Reed B, Smith K, Robinson J, Haines S, Farland M. Situational judgment tests: an introduction for clinician educators. *J Am Coll Clin Pharm.* 2022;5(1):67-74. doi:10.1002/jac5.1571
8. 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):850-868. doi:10.1111/j.1365-2923.2012.04336.x
9. Webster E, Paton L, Crampton P, Tiffin P. Situational judgement tests validity for selection: a systematic review and meta analysis. *Med Educ.* 2020;54(10):888-902. doi:10.1111/medu.14201
10. Tiffin P, Paton L, O'Mara D, MacCann C, Lang J, Lievens F. Situational judgement tests for selection: traditional vs construct-driven approaches. *Med Educ.* 2020;54(2):105-115. doi:10.1111/medu.14011
11. Guenole N, Chernyshenko O, Weekly J. On designing construct driven situational judgment tests: some preliminary recommendations. *Int J Test.* 2017;17(3):234-252. doi:10.1080/15305058.2017.1297817
12. de Leng W, Stegers-Jager K, Husbands A, Dowell J, Born M, Themmen A. Scoring method of a situational judgment test: influence on internal consistency reliability, adverse impact and correlation with personality? *Adv Health Sci Educ Theory Pract.* 2017;22(2):243-265. doi:10.1007/s10459-016-9720-7
13. Wolcott M, Lobczowski N, Zeeman J, McLaughlin J. Situational judgment test validity: an exploratory model of the participant response process using cognitive and think-aloud interviews. *BMC Med Educ.* 2020;20:506. doi:10.1186/s12909-020-02410-z
14. Wolcott M, Lobczowski N, Zeeman J, McLaughlin J. Exploring the role of item scenario features on situational judgment test response selections. *Am J Pharm Educ.* 2021;85(6):Article 8546. doi:10.5688/ajpe8546
15. de Leng W, Stegers-Jager K, Born M, Themmen A. Faking on a situational judgment test in a medical school selection setting: effect of different scoring methods. *Int J Sel Assess.* 2019;27(3):235-248. doi:10.1111/ijsa.12251
16. Eva K, Rosenfeld J, Reiter H, Norman G. An admissions OSCE: the multiple mini-interview. *Med Educ.* 2004;38(3):314-326. doi:10.1046/j.1365-2923.2004.01776.x
17. Pau A, Jeevaratnam K, Chen Y, Fall A, Khoo C, Nadarajah V. The multiple mini-interview (MMI) for student selection in health professions training - a systematic review. *Med Teach.* 2013;35(12):1027-1041. doi:10.3109/0142159X.2013.829912
18. Hopson L, Burkhardt J, Stansfield R, Vohra T, Turner-Lawrence D, Losman E. The multiple mini-interview for emergency medicine resident selection. *J Emerg Med.* 2014;46(4):537-543. doi:10.1016/j.jemermed.2013.08.119
19. Stowe C, Castleberry A, O'Brien C, Flowers S, Warmack T, Gardner S. Development and implementation of the multiple mini-interview in pharmacy admissions. *Curr Pharm Teach Learn.* 2014; 6(6):849-855. doi:10.1016/j.cptl.2014.07.007
20. Hecker K, Violato C. A generalizability analysis of a veterinary school multiple mini-interview: effect of number of interviewers, type of interviewers, and number of stations. *Teach Learn Med.* 2011;23(4):331-336. doi:10.1080/10401334.2011.611769
21. Cameron A, MacKeigan L. Development and pilot testing of a multiple mini-interview for admission to a pharmacy degree program. *Am J Pharm Educ.* 2012;76(1):Article 10. doi:10.5688/ajpe76110
22. Rosenfeld J, Reitter H, Trin K, Eva K. A cost efficiency comparison between the multiple mini-interview and traditional admissions interviews. *Adv Health Sci Educ Theory Pract.* 2008;13(1):43-58. doi:10.1007/s10459-006-9029-z
23. Cox W, McLaughlin J, Singer D, Lewis M, Dinkins M. Development and assessment of the multiple mini-interview in a school of pharmacy admissions model. *Am J Pharm Educ.* 2015;79(4):Article 53. doi:10.5688/ajpe79453
24. Singer D, McLaughlin J, Cox W. The multiple mini-interview as an admission tool for a PharmD program satellite campus. *Am J Pharm Educ.* 2016;80(7):Article 121. doi:10/5688/ajpe807121
25. McLaughlin J, Singer D, Cox W. Candidate evaluation using targeted construct assessment in the multiple mini-interview: a multi-faceted Rasch Model analysis. *Teach Learn Med.* 2017;29(1):68-74. doi:10.1080/10401334.2016.1205997
26. Juster F, Baum R, Xou C, et al. Addressing the diversity-validity dilemma using situation judgment tests. *Acad Med.* 2019;94(8):1197-1203. doi:10.1097/ACM.00000000000002769

Appendix 1. Examples of Situational Judgment Test (SJT) Questions From Each Category

SJT Directions: Rank each of the following response options based on how you SHOULD respond to the scenario. Use 1 to indicate the *MOST* appropriate response and 5 to indicate the *LEAST* appropriate response. There can be no ties or duplicates.

Empathy

You were asked by the physician to speak with a patient's family about the upcoming chemotherapy treatment for their 8-year old son. When you start talking about the negative side effects of the drug treatment, the mother becomes visibly upset and asks you to "stop talking about this."

- 4 Tell that patient's mother it is hospital policy to review all of the necessary information before beginning chemotherapy and you are required to finish
 - 3 Tell the physician the family refused to complete the education and became upset
 - 5 Conclude the session and document education has been complete
 - 2 Request to schedule a different time to continue discussing the medication when the family would be more comfortable
 - 1 Ask the mother about her concerns with the medication
-

Integrity

A good friend is applying for a job where you work. The hiring manager asks you about the person and shares with you their resume. You notice your friend has lied about several things on their resume, including work experience that would make them more qualified for the job than other applicants who would likely get the job over your friend.

- 3 Tell the hiring manager about the discrepancies on the resume
 - 1 Approach your friend about the discrepancies on the resume
 - 4 Advise the manager not to consider your friend at this time
 - 5 Tell your manager that your friend is the best choice for the position
 - 2 Tell your manager that you would like to not be involved in the decision
-

Adaptability

You recently accepted a promotion at work. You had always wanted to work with your new supervisor, who had an excellent reputation as a mentor, and were excited about the promotion. However, a week into your new role, your supervisor left the company and you now have a new supervisor - someone you do not know at all.

- 1 Meet with the new supervisor to learn more about their expectations
 - 4 Talk to someone in Human Resources to express your concerns
 - 3 Continue in your role doing the job you were hired to do
 - 2 Talk with a friend in the company who knows the manager to learn about their style
 - 5 Go back to your previous position
-