BRIEF

A Five-Minute Situational Judgment Test to Assess Empathy in First-Year Student Pharmacists

Michael D. Wolcott, PharmD, PhD,^a Carly Lupton-Smith,^b Wendy C. Cox, PharmD,^a Jacqueline E. McLaughlin, PhD, MS^a

^a University of North Carolina Eshelman School of Pharmacy, Chapel Hill, North Carolina

^b Johns Hopkins University Whiting School of Engineering, Baltimore, Maryland

Submitted January 19, 2018; accepted May 9, 2018; published August 2019.

Objective. To design, implement, and evaluate the utility of a situational judgment test (SJT) to assess empathy in first-year student pharmacists as part of an end-of-year capstone experience.

Methods. First-year students completed a five-minute SJT in lieu of a multiple mini interview (MMI) during the end-of-year capstone. For each SJT item, students selected the two most appropriate response options from a list of five. Various strategies to score the SJT were compared to evaluate the psychometric properties of the test. Student performance on the SJT was examined in relationship to performance on other measures, (eg, MMI stations, personality assessments, and admissions data).

Results. A total of 135 first-year pharmacy students completed an average of 9.5 items. Scoring keys based on subject matter experts' and student responses demonstrated high reliability. There was a positive, weak relationship between student performance on the SJT and performance on the adaptability station used in the capstone, and an inverse, weak relationship with students' agreeableness scores.

Conclusion. This study suggests that the SJT may be a feasible and efficient assessment strategy in pharmacy education. Additional research is needed to inform SJT design, implementation, and interpretation.

Keywords: situational judgment test, empathy, assessment, capstone, multiple mini interview (MMI)

INTRODUCTION

The assessment of social and behavioral competencies such as integrity, empathy, and teamwork in students remains a formidable challenge in health professions and higher education.^{1,2} Nevertheless, the importance of evaluating these qualities has greatly increased within pharmacy education as these measures are now used to inform student selection for pharmacy school, describe development throughout a curriculum, and predict academic performance.³⁻⁷ The strategies used to accurately capture this information are diverse and have a myriad of advantages and disadvantages.^{8,9} Approaches often require extensive resources and fail to provide reliable and valid data in an efficient manner.

Having accurate measures of these attributes in health professions education is critical because developing these skillsets in students can affect patient outcomes

and professional satisfaction.¹⁰ Greater practitioner empathy, for example, is linked to improved patient adherence to treatment, achievement of treatment goals, fewer malpractice complaints, and a reduced incidence of provider burnout and personal distress.¹¹⁻¹³ Research suggests that empathy training can be integrated into curricula to promote students' development of knowledge and skills in this area.¹⁴ However, instruments that provide reliable and valid data are essential to evaluate the impact of such programs and students' progress over time.

The situational judgment test (SJT) is an emerging assessment strategy in health professions education that overcomes the limitations of other assessment methods. Founded in personnel selection, the SJT was developed to evaluate skills beyond measures of cognitive ability previously used to predict occupational performance.⁸ The examinee is presented with a hypothetical scenario they are likely to encounter in practice. Items are designed to assess the construct of interest (eg, empathy, integrity) and examinees are required to select what they believe are the appropriate response options. The key used to score performance on an SJT can be developed by

Corresponding Author: Jacqueline E. McLaughlin, UNC Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, 329 Beard Hall, Chapel Hill, NC 27599. Tel: 919-966-4557. E-mail: jacqui_mclaughlin@unc.edu

aggregating response data from subject matter experts to create what is known as a “rational key,” or from the most common responses from the group being tested to create an “empirical key.”⁹ A more detailed review of the advantages and disadvantages to various SJT design and implementation approaches in medical education can be found elsewhere.^{15,16} A summary of the SJT compared to other common measurement approaches and their key characteristics is provided in Table 1.

While the SJT is an effective strategy to evaluate social and behavioral competencies in other disciplines, the utility of this approach in pharmacy education is unknown. The purpose of this study was to describe the design and implementation of an SJT intended to assess empathy as part of an end-of-year capstone for first-year Doctor of Pharmacy (PharmD) students. This SJT was uniquely designed to be a rapid five-minute assessment, unlike previously described SJTs that dedicated hours to the examination.¹⁷ Moreover, at the time of this research, there was no evidence in the literature regarding the use of SJTs in pharmacy practice.

METHODS

As part of the curricular transformation at the University of North Carolina (UNC) Eshelman School of Pharmacy, a capstone was included at the end of the first

year.¹⁸ The capstone was designed to provide feedback to students about their areas of strengths and opportunities while providing information to the school about the curriculum.⁴ One component of the capstone was a five-station multiple-mini interview (MMI), which was modeled after the MMI used as a component of admissions.³

In the 2017 capstone, the MMI scenario for the empathy station of the capstone MMI was replaced with an empathy SJT to examine the SJT as a reliable and less resource-intensive alternative to the MMI. The design of the SJT was informed by recommendations in the medical education literature and published examples.^{16,17,19} Approximately 25 scenarios were drafted by a member of the research team based on prompts used in admissions, examples in the literature, theoretical descriptors of empathy, and personal experiences in pharmacy practice. The draft questions and responses were reviewed by the research team for clarity and construct-relevance.

Six practicing pharmacists were invited to pilot the instrument. The pharmacists ranked the five response items in order of appropriateness for each scenario. The Kendall coefficient of concordance was used to examine the agreement in ranking responses. A value of 0.6 or greater indicated a high level of agreement among raters.²⁰ Participants were also asked to rate each SJT scenario based on its construct-relevance (ie, how well

Table 1. Comparison of Commonly Used Approaches to Measure Social and Behavioral Competence

	Single Construct Questionnaires	Personality Assessments ⁹	Multiple Mini Interview ⁸	Situational Judgment Tests ¹¹
Content	Variable based on the instrument (eg, Jefferson scale of empathy, emotional intelligence, etc.)	Questions oriented to quantify levels of personality traits (eg, the NEO Big 5, HEXACO, etc.)	Scenarios are oriented to evaluate a construct of interest (eg, adaptability, integrity, empathy)	Scenarios are designed to evaluate a construct of interest (eg, adaptability, integrity, empathy)
Format	Questionnaire/ survey completed by the individual	Questionnaire/ survey completed by the individual	Individuals discuss their response to a scenario with a rater	Examinees select or rank optimal responses to a presented scenario
Scoring Process	Variable; points related to presence of a quality/ attribute	Variable; points related to presence of a quality/ attribute	Rater scores the individual based on observed discussion	Variable; points assigned based on consensus with a key
Advantages	Often short/brief instruments	Substantial research	High validity & reliability	Moderate validity & reliability
	Focuses on a specific construct	Generally stable construct Numerous uses	Can assess multiple constructs at once	Can assess multiple constructs at once
Disadvantages	Potential for faking	Often require commercial licenses	Resource intensive	Potential for faking
	Requires multiple surveys if various skills	May oversimplify personality traits	Potential for faking	Highly variable design and scoring strategies
	Variable quality	Can be lengthy	Influence of rater bias	

it assessed empathy). Items with the highest agreement for measuring empathy were placed at the top of the capstone SJT to ensure questions that were most aligned with the construct were answered by the greatest number of examinees. Items not marked as empathy-related were removed. Nineteen questions were included in the final SJT. A rational key was constructed based on the rankings provided by the subject matter experts.

The SJT was administered through an electronic survey software (Qualtrics, SAP. Walldorf, Baden-Württemberg, Germany) using a computer laptop provided in the MMI station room. A facilitator was present in the room to set-up the survey instrument and prompt student to follow the on-screen directions. The electronic survey instrument was designed to automatically close five minutes after the start of the survey to ensure all students had an equivalent amount of time to respond. Paper copies of the test were available in the event that technical difficulties occurred.

An empirical key was developed using the frequency of each response option selected by the students. The rational and empirical keys were compared to identify items with differences in ranking. Those items were qualitatively inspected to determine whether design features or context may have contributed to differences in the keys. We were interested in determining if these items included context-relevant information that would have resulted in different keys based on dissimilar levels of clinical experience between subject matter experts and students.

Because there is minimal consensus concerning SJT scoring methods, various scoring strategies were explored to quantify student performance (eg, distance, square distance).¹⁵ We highlight one scoring strategy in this report based on its feasibility, acceptable psychometric properties, and ease of calculation. The “distance” scoring approach used the distance of the selected response from the correct response, which means that higher point values were assigned to correct responses based on the rational or empirical key. For example, selecting the highest-ranked response option contributed four points to the score, the second highest-ranked options contributed three points, and so on (ie, 4, 3, 2, 1, and 0). Because the Pearson correlations for the “distance” scores and the “square distance” scores were very strong ($r_p = .99$ for “distance, empirical key” and “square distance, empirical key” and $r_p = .99$ for “distance, rational key” and “square distance, rational key”), results were only reported for the “distance” scores. Mean student score and standard deviations (SD) were determined based on each scoring key, and the reliability was evaluated using Cronbach alpha, with values greater than 0.70 considered acceptable.²¹

To further investigate the SJT results, student performance based on the scoring strategies was compared to other methods believed to measure social and behavioral competence. Pairwise comparison of the Pearson correlations of SJT performance with performance on other capstone MMI stations and the admissions empathy MMI station were evaluated. The relationships of SJT performance to scores from the HEXACO personality inventory were also examined: honesty, emotionality, extraversion, agreeableness, conscientiousness, openness, and altruism.²² Finally, Pearson correlations of SJT scores with Pharmacy College Admissions Test (PCAT) percentile ranks and performance on the closed book test also administered during the capstone were explored. Because the PCAT and closed-book test primarily measure clinical knowledge, examining the relationship between these assessments and the SJT can provide evidence of discriminant validity and support the hypothesis that the SJT measured an attribute other than knowledge (ie, a weak correlation).

Correlations less than 0.30 were considered weak, values between 0.30 to 0.70 were moderate, and those greater than 0.70 were strong.²¹ Statistical analyses were conducted using Stata 14.2 for Windows (StataCorp. College Station, TX), with p -values less than .05 denoting significance. Continuous data were represented as mean (SD) and categorical data were represented as number (percent). This study was considered exempt from review by the University of North Carolina Institutional Review Board.

RESULTS

All first-year students ($N=143$) completed the SJT. Eight students incorrectly submitted more than two answers for at least one test question and their data were subsequently removed from the final analyses ($n=135$). Students completed between five and 16 items [mean (SD)=9.5 (2.0)]. Analyses included performance data for the first 12 items because each of these items had at least 20 respondents, which was considered an appropriate sample size for our analyses.

Kendall's coefficient of concordance for each item on the rational key indicated levels of agreement ranging from 0.46 to 0.93 ($p < .05$ for all items). Two items (number 1 and 7) had a coefficient below 0.6, and overall agreement across the items was achieved. Eight of the items had differences in the ranking of items between the rational and empirical keys (Table 2). These items included context-specific information that may have influenced the decision-making process for a clinical expert in comparison to a student. For example, items with equivalent rankings included scenarios that were broad

Table 2. Rational (Subject Matter Expert) and Empirical (Student) Keys for an Empathy-Based Situational Judgement Test

Item	Rational Key	Empirical Key ^a
1	C, B, D, A, E	C, d, b, A, E
2	B, D, E, A, C	B, D, E, A, C
3	E, D, B, A, C	d, e, a, b, C
4	B, A, C, E, D	B, A, C, E, D
5	B, A, E, D, C	B, A, E, D, C
6	A, D, B, E, C	A, D, c, b, e
7	B, A, C, D, E	B, A, d, c, E
8	B, D, E, A, C	e, D, b, A, C
9	A, B, D, E, C	A, B, D, E, C
10	E, C, D, A, B	E, d, c, A, B
11	B, D, E, C, A	B, a, c, d, e
12	C, D, E, A, B	d, c, E, b, a

^a Lower-case letters indicate a rank difference between the two keys

and related to general aspects of empathy, such as asking more questions to better understand a person's perspective, dealing with inaccurate information, or handling someone who is upset. Items with dissimilar rankings were more context-specific, such as allowing a patient to have a particular meal, handling treatment disagreements, and managing medical errors.

Overall, scoring methods with the rational and empirical keys were shown to be reliable, with Cronbach alphas of 0.75 to 0.73, respectively. Based on the rational key, SJT scores ranged from 30 to 95 with an average of 56.6 (SD=12.4) and from 31 to 82 with an average of 58 (SD=12) based on the empirical key.

The SJT scores had a weak linear relationship with composite PCAT scores ($r=0.24$ to 0.28 , all $p<.05$) and no linear relationship to the closed book examination ($r=0.00$ to 0.03 , all $p>.05$; Table 3). The relationship to other MMI stations was variable. The most significant was a positive, weak relationship of SJT scores with scores on the adaptability station ($r=0.23$ to 0.24 , all $p<.01$). Of the personality attributes assessed by the HEXACO instrument, agreeableness had a weak, inverse relationship with performance on the SJT ($r=-0.21$ to -0.16). All other correlations with the HEXACO instrument were not significant for both keys ($r=-0.15$ to 0.11 , all $p>.05$). There were also negligible, positive correlations between SJT scores and student scores on the admissions MMI empathy station ($r=0.11$ to 0.12 , all $p>.05$).

DISCUSSION

The SJT is an emerging assessment methodology in health professions education that may provide valuable insight for evaluating students' social and behavioral

Table 3. Comparison of Pharmacy Students' Performance on an Empathy-based Situational Judgement Test With Other Measures of Competency, Cognition, and Personality

Method	Rational Scoring Key	Empirical Scoring Key
Admissions MMI: Empathy	0.11	0.11
Capstone MMI: Collaboration	0.05	0.05
Capstone MMI: Critical Thinking	0.11	0.10
Capstone MMI: Adaptability	0.24 ^b	0.24 ^b
PCAT Percentile Rank	0.28 ^b	0.24 ^b
Capstone: Closed Book Test	0.02	0.00
HEXACO: Agreeableness ^c	-0.18 ^a	-0.21 ^a

Abbreviations: MMI=multiple mini-interview; PCAT=pharmacy college admissions test; HEXACO= Honesty-Humility (H), Emotionality (E), Extraversion (X), Agreeableness (A), Conscientiousness (C), and Openness to Experience (O)

^a $p<.05$

^b $p<.01$

^c All other HEXACO constructs (ie, honesty, emotionality, extraversion, conscientiousness, openness, and altruism) demonstrated weak to negligible correlations with each scoring key ($0.02\leq r_p\leq 0.15$)

competency.^{8,16,17} The SJT offers several benefits compared to other strategies such as the MMI in that it is efficient and less resource-intensive yet still produces valid and reliable results when designed effectively.^{16,17,19,23,24} In this pilot study, we explored the properties of a rapid, targeted SJT to assess empathy of first-year student pharmacists during an end-of-year capstone. Our work emphasizes three key findings: a rapid assessment using an SJT can produce reliable results, the context of the scenario may be important to consider when designing SJT items, and the relationship of the SJT to other measures of competency requires further exploration.

A highlight of our study was the ability to design and implement a short, targeted SJT that produced reliable results as observed by the high internal consistency. Typically, SJTs address multiple constructs at once, which requires a substantial amount of time for development and testing (often one to two hours).¹⁶ Our approach of focusing on one construct for an abbreviated amount of time demonstrates the flexibility of the SJT as an assessment strategy that does not compromise reliability. In this case, the SJT development and administration required fewer resources than executing the MMI. For example, the SJT did not require facilitator training or scoring

throughout the testing process; all that was necessary was the space, computers, and time to develop and pilot the test items. Furthermore, previous studies showed substantial variation among scoring methods¹⁵; however, our design approach generated reliably ranked students in a consistent manner across a variety of scoring strategies. This is an important finding given the unique structure of our assessment compared to design approaches used in previously published studies. Additional research on SJTs should aim to identify and further develop those design features (eg, question format, response options, scoring strategies) that consistently produce reliable data with minimally complex scoring mechanisms.

An additional feature of the study was exploring differences between students and subject matter experts in response option ranking as a proxy for understanding the role of context in item design. The extent to which an SJT is truly “situational” requires further investigation.²³ In our study, examining differences in the rational and empirical keys was an attempt to gain additional insight into SJT design. Differences between students and subject matter experts in their ranking of response options was likely attributable to differences in levels of expertise based on clinical experience. Questions in which experts and student differed in their response rankings were more context-dependent compared to questions that did not differ in ranking. Those items assessed general attributes of empathy that would be expected from typical socialization and development rather than from training as a healthcare provider. Future research should explore the impact of context-specific factors in SJT design by comparing how examinees with varying levels of experience respond to items. Techniques such as differential item functioning (DIF) could be used to quantify differences between these groups and elucidate when context-specific information may or may not be a desirable design feature.

Another critical issue in SJT research is understanding what is actually being measured by the assessment (eg, construct validity). The relationship of SJT performance to other student variables (eg, PCAT, HEXACO) is a first step in addressing this issue, as we have done within our presented work. Overall, the lack of strong linear relationships between the SJT and other capstone stations and personality assessments signifies our SJT measured a unique attribute of students that was not measured using these other instruments. It was a positive finding that there were low correlations with knowledge tests, such as the PCAT, as the goal of our SJT was to measure skills other than cognitive ability.^{16,23,24} Similarly, the absence of a strong relationship to personality attributes was a positive finding. The SJTs are intended to measure

social and behavioral competencies that are ideally not associated extensively with personality traits;¹⁵⁻¹⁷ therefore, our SJT was successful at discerning a construct with minimal overlap of other qualities. More research is needed to evaluate the validity of SJTs and explicate what is being measured; for example, studies are needed to connect SJT performance with observed behaviors in real-life scenarios to determine whether knowledge of an appropriate response translates to informed behaviors.

The data presented in this study reflect the first attempt to integrate the SJT into pharmacy education. As such, there are several limitations to our study. Notably, the first implementation of an assessment strategy limits the ability to evaluate its predictive validity, which is an area where SJTs would be highly valuable. Future research should include how SJT performance relates to subsequent performance in experiential settings or other areas requiring social and behavioral skill sets such as residency or job sites. Moreover, student and faculty perceptions about the utility of the SJT was beyond the scope of this study, yet these perceptions could provide valuable insight into its acceptability. Understanding the perceived utility of the SJT in pharmacy education would be a notable contribution to the literature.

Questions regarding the use of the SJT in pharmacy and health professions education remain. Our pilot study focused on using the SJT to assess empathy; however, the authors believe that SJTs may be a feasible approach to assess various social and behavioral constructs. For example, can integrity, adaptability, collaboration, and other skills all be as adequately measured using an SJT, or are certain other constructs better suited for this assessment strategy? From a design perspective, are there specific elements that should or should not be included in test items to optimize the reliability and validity of the data collected? In addition, pharmacy educators must consider how assessments of empathy and similar constructs fit within the curriculum. If empathy can be improved,¹⁴ how could data from SJTs be used to inform and improve the empathy curriculum? The experiential curriculum or courses in which empathy is emphasized through reflection or role-play, for example, may be well suited for both developing and assessing this skill. We believe SJTs hold promise for use in assessment in pharmacy education and encourage researchers to explore these and other questions about these instruments.

CONCLUSION

The SJT is an emerging strategy in the assessment of social and behavioral competencies such as empathy when integrated within a health professions curriculum.

The findings of this study suggest the SJT can be a feasible and efficient instrument for assessing students. When developing and implementing SJTs, educators should thoroughly evaluate design features, pilot test items, and develop a comprehensive plan for implementation to ensure resources are available. As this is the first example of implementing the SJT in pharmacy education, we hope it guides future studies that explore SJTs as a method for longitudinal and comprehensive assessment of social and behavioral competencies in pharmacy education.

ACKNOWLEDGMENTS

We thank the faculty, staff, and First Year Capstone Committee at the UNC Eshelman School of Pharmacy for their dedication and contributions to the capstone, MMI, and SJT. Specifically, we would like to thank Heidi Anksorus, Patrick Brown, Miranda Law, Chelsea Renfro, Phil Rodgers, Tom Angelo, and Jackie Zeeman for their assistance in refining the SJT. We also thank Adam Persky for his collaboration and help with the HEXACO personality assessment.

REFERENCES

1. Niessen ASM, Meijer RR. On the use of broadened admission criteria in higher education. *Perspect Psychol Sci.* 2017;12(3):436-448.
2. Bardes CL, Best PC, Kremer SJ, Dienstag JL. Perspective: medical school admissions and non-cognitive testing: some open questions. *Acad Med.* 2009;84:1360-1363.
3. Cox WC, McLaughlin JE, Singer D, Lewis M, Dinkins MM. Development and assessment of the multiple mini-interview in a school of pharmacy admissions model. *Am J Pharm Educ.* 2015;79(4):Article 53.
4. Persky AM, Greene JM, Anksorus H, Fuller KA, McLaughlin JE. Developing an innovative, comprehensive first-year capstone to assess and inform student learning and curriculum effectiveness. *Am J Pharm Educ.* 2018;83(5):Article 6730.
5. Cowart K, Dell K, Rodriguez-Snapp N, Petrelli HM. An examination of correlations between MMI scores and pharmacy school GPA. *Am J Pharm Educ.* 2016;80(6):Article 98.
6. Heldenbrand SD, Flowers SK, Bordelon BJ, et al. Multiple mini-interview performance predicts academic difficulty in the PharmD curriculum. *Am J Pharm Educ.* 2016;80(2):Article 27.
7. Cameron AJ, MacKeigan LD, Mitsakakis N, Pugsley JA. Multiple mini-interview predictive validity for performance on pharmacy licensing examination. *Med Educ.* 2017;51:379-389.
8. Patterson F, Knight A, Dowell J, Nicholson S, Cousans F, Cleland J. How effective are selection methods in medical education? A systematic review. *Med Educ.* 2016;50:36-60.
9. Hojat M, Erdmann JB, Gonnella JS. Personality assessments and outcomes in medical education and the practice of medicine: AMEE Guide No. 79. *Med Teach.* 2013;35:e1267-e1301.
10. Decety J, Fotopoulou A. Why empathy has a beneficial impact on others in medicine: unifying theories. *Front Behav Neurosci.* 2015;8:457.
11. Halpern J. What is clinical empathy? *J Gen Intern Med.* 2003;18(8):670-674.
12. Brazeau CMLR, Shroeder R, Rovi S, Boyd L. Relationships between medical student burnout, empathy, and professionalism climate. *Acad Med.* 2011;86(10 Suppl):33-36.
13. Hojat M, Louis DZ, Markham FW, Wender R, Rabinowitz C, Gonnella JS. Physicians' empathy and clinical outcomes for diabetic patients. *Acad Med.* 2011;86(3):359-364.
14. Smith KE, Norman GJ, Decety J. The complexity of empathy during medical school training: evidence for positive change. *Med Educ.* 2017;51:1146-1159.
15. De Leng WE, Stegers-Jager KM, Husbands A, Dowell JS, Born MP, Themmen APN. 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.
16. Patterson F, Zibarras L, Ashworth V. Situational judgement tests in medical education and training: research, theory and practice: AMEE Guide No. 100. *Med Teach.* 2015;38(1):3-17.
17. McDaniel MA, Hartman NS, Whetzel DL, Grubb WL. Situational judgment tests, response instructions, and validity: a meta-analysis. *Per Psychol.* 2007;60:63-91.
18. Roth MT, Mumper RJ, Singleton SF, et al. A renaissance in pharmacy education at the University of North Carolina at Chapel Hill. *N C Med J.* 2014;75(1):48-52.
19. Patterson F, Ashworth V, Zibarras L, Coan P, Kerrin M, O'Neill P. Evaluations of situational judgment tests to assess non-academic attributes in selection. *Med Educ.* 2012;46:850-868.
20. Siegel S, Castellan Jr. NJ. *Nonparametric Statistics for the Behavioral Sciences.* 2nd ed. New York, NY: McGraw-Hill; 1998.
21. Hinkle DE, Wiersma W, Jurs SG. *Applied Statistics for the Behavioral Sciences.* 5th ed. Boston, MA: Houghton Mifflin; 2002.
22. Lee K, Ashton MC. Psychometric properties of the HEXACO personality inventory. *Multivar Behav Res.* 2004;39:329-358.
23. Cousans F, Patterson F, Edwards H, Walker K, McLachlan JC, Good D. Evaluating the complementary roles of an SJT and academic assessment for entry into clinical practice. *Adv in Health Sci Educ.* 2017;22:401-413.
24. Lievens F, Sackett PR. The validity of interpersonal skills assessment via situational judgments tests for predicting academic success and job performance. *Journal of Applied Psychology.* 2012;97(2):460-468.