https://doi.org/10.4082/kjfm.2018.39.2.96 • Korean J Fam Med 2018;39:96-100

KJFM
Korean Journal of Family Medicine

eISSN: 2092-6715

■ Original Article

Comparison of Patient–Physician Interaction Scores of Clinical Practice Examination between Checklists and Rating Scale

Nam Eun Kim, Hoon Ki Park*, Kyong Min Park, Bong Kyung Seo, Kye Yeung Park, Hwan Sik Hwang

Department of Family Medicine, Hanyang University College of Medicine, Seoul, Korea

Background: The clinical practice examination (CPX) was introduced in 2010, and the Seoul-Gyeonggi CPX Consortium developed the patient-physician interaction (PPI) assessment tool in 2004. Both institutions use rating scales on classified sections of PPI but differ in their scoring of key components. This study investigated the accuracy of standardized patient scores across rating scales by comparing checklist methods and verified the concurrent validity of two comparable PPI rating tools.

Methods: An educational CPX module dyspepsia case was administered to 116 fourth-year medical students at Hanyang University College of Medicine. One experienced standardized patient rated exams using two different PPI scales. She scored checklists composed of 43 items related to the two original PPI scales through video clips of the same students. From these checklists, we calculated Pearson's correlation coefficient.

Results: The correlations of total PPI score between the checklist and rating scale methods were 0.29 for the Korean Medical Licensing Examination (KMLE) tool and 0.30 for the consortium tool. The correlations between the KMLE and consortium tools were 0.74 for checklists and 0.83 for rating scales. In terms of section scores, the consortium tool showed only three significant correlations between the two methods out of seven sections and the KMLE tool showed only two statistically significant correlations out of five sections.

Conclusion: The rating scale and checklist methods exhibited a weak relationship in the PPI assessment, but a high correlation between assessment tools using the same method. However, the current rating scale requires modification by reorganizing key scoring components through factor analysis.

Keywords: Physician-Patient Relations; Medical Education; Educational Measurement; Behavior Rating Scale; Checklist

Received: June 30, 2016, Revised: October 7, 2016, Accepted: October 14, 2016

*Corresponding Author: **Hoon Ki Park** https://orcid.org/0000-0002-8242-0943
Tel: +82-2-2290-8738, Fax: +82-2-2281-7279, E-mail: hoonkp@hanyang.ac.kr

KJFM

Copyright © 2018 The Korean Academy of Family Medicine
This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0) which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

The need to assess the competence of medical students in high-level skill areas, as described by Miller's framework of clinical assessment (pyramid of competence),[1,2] along with the well-established importance of communication skills between doctors and patients,[3] has led to the introduction of clinical skills assessment as part of the Korean Medical Licensing Examination (KMLE), which has been administered by the National Health Personnel Licensing Examination Board since 2009. Ever since standardized patients were introduced as a tool for clinical assessment,[4] their use has been widely accepted as one of the most effective methods in medical education.[1,5,6] The use of standardized patients in the clinical practice examination (CPX) has also allowed for a better evaluation of the patient-physician interaction (PPI), which is important for patient outcomes[7,8] and satisfaction,[9,10] as demonstrated by numerous studies over several decades.

Among the three core elements of CPX (information gathering, clinical reasoning, and PPI), the PPI cannot be easily assessed with unequivocal accuracy when using tools consisting of many sets of section-dependent differentiated instruments.[11,12] While the importance of ongoing attempts to validate the PPI assessment in improving quality health care has been repeatedly emphasized, it is difficult to develop a general measurement tool that reflects the unique nature of nonverbal communication[13] and subjective socioemotional behavior.[14] Arguments over the psychometric properties, reliability, and validity of assessment methods have led to comparative studies of the checklist and rating scale methods.[15,16] Although the checklist method, originally regarded as objective, is now considered "objectified,"[15,16] the rating method is still subject to rater biases and technical issues.[2]

Table 1. Baseline characteristics of study subjects (n=116)

Characteristic	Value
Age (y)	26.4±3.0
Sex	
Male	74 (63.8)
Female	42 (36.2)
Type of school	
College	59 (50.9)
School	57 (49.1)

Values are presented as mean±standard deviation or number (%).

Currently, the PPI assessment tools of both the KMLE and Seoul-Gyeonggi CPX Consortium, which was developed independently in 2004, use different rating scale methods. Moreover, each tool contains different sections and different key components. Standardized patients may apply rating scales either holistically or analytically followed by synthesizing gradient meanings of major rhetoric. This study was conducted to investigate the accuracy of standardized patient scores across rating scales by comparing checklist methods and to verify the concurrent validity of two comparable PPI rating tools.

METHODS

Videotaped standardized patients-based CPX cases of 'dyspepsia' from the fourth-year medical student assessment administered in 2015 at Hanyang University College of Medicine, a member of the Seoul-Gyeonggi CPX Consortium, were used for this study. The study subjects included 113 fourth-year medical students and 3 holdover students, including both medical college (50.9%) and medical school students (49.1%). The study subjects had a mean age of 26.4 and consisted of 63.8% male students (Table 1).

The current KMLE-PPI rating scale method is scored by summing the scores on a 4-point Likert scale for each section, which consists of several key components. The checklist was newly produced in our study by breaking down the key components of each section by the rating scale method. The KMLE checklist consists of 25 key components and five sections scored by the rating scale method: five items for PPI1, six for PPI2, four for PPI3, five for PPI4, and five for PPI5. Likewise, the consortium checklist consists of 34 components based on seven sections scored using a 6-point Likert rating scale: three items for PPI1, seven items for PPI2, four items for PPI3, seven items for PPI4, five items for PPI5, three items for PPI6, and five items for PPI7. The integrated checklist contains 43 components, including 16 that overlap. All checklists consist of affirmative statements with yes/no answers. "Rating scale" refers to scoring on a 4- or 6-point Likert scale, and "checklist scoring" refers to summing the scores for each checklist item into one composite section score.

The PPI assessment was completed by standardized patients with over 15 years of field experience who participated in the CPX case simulation. Scoring the PPI using both the KMLE and consortium rating scale methods was performed immediately after the student-stan-

Table 2. Correlation between the total PPI score of the checklist method and the rating scale method

PPI		Checklist				Rating scale			
		KMLE		Consortium		KMLE		Consortium	
		r	P-value	r	P-value	r	P-value	r	P-value
Checklist	KMLE	1.00							
	Consortium	**0.74**	<0.0001	1.00					
Rating scale	KMLE	**0.29**	0.002	0.31	0.001	1.00			
	Consortium	0.26	0.005	**0.30**	0.001	**0.83**		1.00	<0.0001

Bold type is considered statistically significant.
PPI, patient-physician interaction; KMLE, Korean Medical Licensing Examination.

dardized patients encounter. The same standardized patients re-evaluated the PPI using the checklist method after watching the video recordings. Repeated focus playback of the video was allowed. The PPI was assessed in accordance with the Likert rating scale and checklist methods.

First, we studied the correlation between the total PPI score of the checklist method and the rating scale method for each institutional tool. The correlation between the scores obtained with the same method for the two tools was also analyzed (Table 2). As the consortium test has sevn PPI sections (Table 3) and the KMLE has five PPI sections (Table 4), we further studied the relationship between the scores for each section obtained by the checklist and rating scale methods within the same organization. Although factor analysis was outside the scope of this study, we further analyzed a section from each method: the PPI6 skill (information sharing) of the consortium and the PPI4 skill (explanation) of the KMLE. In doing so, we examined the correlations of the evaluation items with the superordinate section, "provided information in easily understood statements."

We analyzed the correlations between the KMLE and consortium PPI rating scores for the same scoring method. Within the scores for each institutional tool, we further analyzed the correlation between the scores obtained by the rating scale and the checklist. All correlations in this study were analyzed using Pearson's correlation coefficient. As a rule of thumb, correlation coefficients between 0.25 and 0.50 are believed to indicate a fair degree of relationship, correlation coefficients between 0.50 and 0.75 are believed to indicate a moderate to good relationship, and correlation coefficients greater than 0.75 are believed to indicate a very good to excellent relationship.[17] A P-value <0.05 was considered statistically significant. All statistical analyses were performed using SAS ver. 9.4 (SAS Institute Inc., Cary, NC, USA).

RESULTS

The results of the analysis are shown in Table 2. The PPI scores according to the KMLE checklist and rating scale methods showed a statistically significant correlation coefficient of 0.29, while the PPI scores according to the consortium checklist and rating scale methods showed a statistically significant correlation coefficient of 0.30 (Table 2). The PPI score obtained by the two organizations' checklists showed a statistically highly significant correlation coefficient of 0.74, while the PPI scores obtained by the two organizations' rating scales showed a statistically highly significant correlation coefficient of 0.83 (Table 2). In conclusion, scores obtained by different methods within the same organization's assessment tool had relatively low correlations, but scores obtained by the same method were highly correlated regardless of the assessment tool used.

Correlations of the PPI scores for each section obtained by the consortium rating scale and checklist methods were only statistically significant in three of the seven section: PPI1 (comfortable relationship, r=0.24), PPI2 (listening attentively and questioning, r=0.33), and PPI7

Table 3. Correlation between the category PPI score of the checklist method and the rating scale method of the "Seoul-Gyeonggi Clinical Practice Examination Consortium" assessment tool

Rating scale	Consortium checklists													
	PPI1		PPI2		PPI3		PPI4		PPI5		PPI6		PPI7	
	r	P-value	r	P-value	r	P-value	r	P-value	r	P-value	r	P-value	r	P-value
PPI1	**0.24**	0.001	-0.04	0.65	0.05	0.61	0.14	0.14	0.32	0.001	-0.14	0.13	0.15	0.10
PPI2	0.01	0.93	**0.33**	0.0002	0.02	0.79	0.03	0.77	0.06	0.53	-0.08	0.37	0.09	0.34
PPI3	0.18	0.06	0.03	0.75	**0.23**	0.12	0.26	0.004	0.21	0.02	0.01	0.93	0.28	0.002
PPI4	0.38	<0.0001	-0.10	0.27	0.07	0.44	**0.18**	0.05	0.30	0.001	0.02	0.84	0.17	0.06
PPI5	0.22	0.02	-0.12	0.19	0.09	0.33	0.11	0.23	**0.14**	0.14	-0.19	0.04	0.32	0.0004
PPI6	0.08	0.41	0.06	0.51	0.16	0.08	0.13	0.17	0.17	0.07	**-0.05**	0.56	0.12	0.18
PPI7	0.29	0.002	-0.05	0.55	0.20	0.03	0.23	0.01	0.22	0.01	-0.10	0.30	**0.25**	0.01

Bold type is considered statistically significant.
PPI, patient-physician interaction.

Table 4. Correlation between the category PPI score of the checklist method and the rating scale method of the KMLE assessment tool

Rating scale	KMLE clinical skills assessment checklists									
	PPI1		PPI2		PPI3		PPI4		PPI5	
	r	P-value	r	P-value	r	P-value	r	P-value	r	P-value
PPI1	**0.14**	0.13	**0.12**	0.20	**0.04**	0.70	**0.03**	0.74	**0.17**	0.07
PPI2	-0.02	0.82	**0.12**	0.20	0.10	0.28	0.03	0.72	0.47	<0.0001
PPI3	0.01	0.96	0.12	0.19	**0.18**	0.05	0.01	0.92	0.17	0.07
PPI4	0.04	0.69	-0.05	0.58	-0.05	0.61	**0.27**	0.004	0.10	0.31
PPI5	0.03	0.74	0.13	0.18	0.16	0.09	0.04	0.65	**0.37**	<0.0001

Bold type is considered statistically significant.
PPI, patient-physician interaction; KMLE, Korean Medical Licensing Examination.

Table 5. Correlations of checklist items with mean checklist scores and global ratings of information sharing

Item	Gave information in easily understood statements			
	Mean checklist scores		Global ratings	
	Consortium	KMLE CSA	Consortium	KMLE CSA
Gave information concisely and clearly in assimilable chunks	NA	0.48 (<0.01)	NA	-0.01 (0.91)
Checked occasionally patient's understanding of information given	NA	0.49 (<0.01)	NA	0.30 (<0.01)
Allowed me to ask questions	NA	0.68 (<0.01)	NA	0.23 (0.01)
Did not use medical jargon	0.91 (<0.01)	0.33 (<0.01)	-0.14 (0.14)	0.03 (0.76)
Thoroughly explained when jargon was unavoidably used	0.74 (<0.01)	0.44 (<0.01)	0.02 (0.83)	0.09 (0.34)
Thoroughly explained the physical findings, etiology, management plan, and/or laboratory results	0.20 (0.03)	NA	0.25 (<0.01)	NA

Numbers in parentheses are P-value of Pearson correlation coefficients.
KMLE, Korean Medical Licensing Examination; CSA, clinical skills assessment; NA, not applicable.

(building relationship, r=0.25) (Table 3). Likewise, correlations of each section's PPI score obtained by the KMLE rating scale and checklist methods were only statistically significant in two of the five sections: PPI4 (explanation, r=0.27) and PPI5 (achieving relationship, r=0.37) (Table 4).

Further examination of the PPI6 section (information sharing) of the consortium assessment tool found that its three evaluation items showed the following correlations with the superordinate section "provided information in easily understood statements": item 1 ("examinee did not use medical jargon"), r=0.91; item 2 ("examinee thoroughly explained when jargon was unavoidably used"), r=0.74; and item 3 ("examinee thoroughly explained the physical findings, etiology, management plan, and/or laboratory results"), r=0.20 (Table 5). The equivalent PPI4 section (explanation) of the KMLE assessment tool showed correlation coefficients of r=-0.01, r=0.30, r=0.23, and r=0.03 with "examinee gave information concisely and clearly in assimilable chunks" (subset 1), "examinee occasionally checked patient's understanding of information given" (subset 2), "examinee checked if there were any questions" (subset 3), and "examinee did not use medical jargon" (subset 4), respectively (Table 5).

DISCUSSION

Our study investigated the interrelationship of the PPI scores obtained using the assessment tools of the KMLE and the Seoul-Gyeonggi CPX Consortium. The PPI scores obtained by both the rating scale and checklist methods exhibited a good relationship within a given method. This suggests that the total PPI score obtained by summing the scores in each subordinate section using rating scale methods is consistent across assessment tools. Moreover, the total PPI scores obtained by summing the scores of subset items checked 'yes' using the checklist method show a moderately strong relationship between the two organizations' assessment tools.

During PPI assessment, a standardized patient marks each section's overall performance on a 6-point Likert scale based on each section's evaluation component items. Therefore, treating each section's evaluation components as checklists can be considered an analytical approach to assessment. In conventional assessment training, standardized patients are taught to assess each component and then comprehensively rate each section using the rating scale. Hence, the preliminary step in the checklist method is the quantitative measurement of performance on each section. The PPI scores obtained by the consortium checklist and rating scale methods show a weak relationship, as do the PPI scores obtained by the KMLE checklist and rating scale methods. This indicates that when standardized patients assess PPIs with the rating scale method, the most important part is the comprehensive understanding of the global performance of the examinee in a specific section of the PPI, not the evaluation component items.

This result agrees with a longstanding argument emphasizing that subjective ratings have a higher reliability than objective checklists.[2,15,16] Both organizations' assessment tools use a 4- or 6-point Likert scale. The results of this study suggest that current standardized patients properly evaluate the PPI components of examinees using the rating scale method in a comprehensive manner. There have been concerns that standardized patients do not realize the precise intention of the rating scale method, as they are preoccupied with checking each performance item.[18] Nevertheless, these results suggest that standardized patients adapt easily to the rating scale method, and this highlights the need to improve the validity of the rating scale method rather than replacing it with the checklist method.

While the total PPI scores of the consortium's rating scale and checklist methods show significant correlation, there is a lack of correlation of the PPI scores obtained by the two methods within the sections, with only three out of seven being significant. We can infer from this that the standardized patient understands each PPI section's abstract concept and globally assesses the examinee rather than analytically assessing detailed items. Conversely, the standardized patient might have misunderstood the current classification of sections and items. Considering the results for the relationship between each item's score and the total score for one specific criterion, for example the section examining "gave information in easily understood statements," if we exclude the weak relationship in item 3, the correlations between the total score for the superordinate section of the consortium test and the scores for item 1 and item 2 display an excellent relationship, with a correlation coefficient of 0.99. This implies a possible structural error whereby key concepts were misplaced in non-relevant sections in the

current rating scale.

Similarly, the lack of correlation for three sections of the KMLE scale again suggests the standardized patient emphasizes a comprehensive approach. Items 1 and 4 show a weak relationship with the PPI 4 section, "provided information in easily understood statements." Consequently, use of the rating scale in both current assessment tools needs to be reassessed regarding the validity of their sections and items. If we are to maintain the current assessment sections, systematic training will be necessary for the standardized patient to fully understand and accurately grade each assessment factor.

Our study has a few limitations. First, as the data were from a single standardized patient who performed assessment at one station, it is difficult to generalize them to other settings. Second, the rating scale method was applied immediately after the encounter, whereas the checklist method was performed while viewing the recorded video, which might have introduced error. Moreover, it may be advisable to take more time during the checklist method for more precise analysis, considering that there are 43 items to be considered.

This study partially verified the accuracy of standardized patients' scoring of PPIs using rating scales in comparison with checklist methods, but with a controversial level of standards required. Further studies that develop a more accurate instrument for measuring competency of communication skills are needed. In conclusion, the rating scale and analytic checklist methods are weakly related in PPI assessment, and the current rating scale method may require improvement by re-examining the construct of the assessment criteria.

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

REFERENCES

1. Miller GE. The assessment of clinical skills/competence/performance. Acad Med 1990;65(9 Suppl):S63-7.
2. Wass V, van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. Lancet 2001;357:945-9.
3. Makoul G, Schofield T. Communication teaching and assessment in medical education: an international consensus statement. Netherlands Institute of Primary Health Care. Patient Educ Couns 1999;37:191-5.
4. Barrows HS, Abrahamson S. The programmed patient: a technique for appraising student performance in clinical neurology. J Med Educ 1964;39:802-5.
5. Howley LD. Performance assessment in medical education: where we've been and where we're going. Eval Health Prof 2004;27:285-303.
6. Barrows HS. An overview of the uses of standardized patients for teaching and evaluating clinical skills. Acad Med 1993;68:443-51.
7. Kaplan SH, Greenfield S, Ware JE Jr. Assessing the effects of physician-patient interactions on the outcomes of chronic disease. Med Care 1989;27(3 Suppl):S110-27.
8. Stewart MA. Effective physician-patient communication and health outcomes: a review. CMAJ 1995;152:1423-33.
9. Bartlett EE, Grayson M, Barker R, Levine DM, Golden A, Libber S. The effects of physician communications skills on patient satisfaction; recall, and adherence. J Chronic Dis 1984;37:755-64.
10. Comstock LM, Hooper EM, Goodwin JM, Goodwin JS. Physician behaviors that correlate with patient satisfaction. J Med Educ 1982;57:105-12.
11. Boon H, Stewart M. Patient-physician communication assessment instruments: 1986 to 1996 in review. Patient Educ Couns 1998;35:161-76.
12. Epstein RM, Franks P, Fiscella K, Shields CG, Meldrum SC, Kravitz RL, et al. Measuring patient-centered communication in patient-physician consultations: theoretical and practical issues. Soc Sci Med 2005;61:1516-28.
13. Gallagher TJ, Hartung PJ, Gerzina H, Gregory SW Jr, Merolla D. Further analysis of a doctor-patient nonverbal communication instrument. Patient Educ Couns 2005;57:262-71.
14. Wasserman RC, Inui TS. Systematic analysis of clinician-patient interactions: a critique of recent approaches with suggestions for future research. Med Care 1983;21:279-93.
15. Cohen DS, Colliver JA, Marcy MS, Fried ED, Swartz MH. Psychometric properties of a standardized-patient checklist and rating-scale form used to assess interpersonal and communication skills. Acad Med 1996;71(1 Suppl):S87-9.
16. Cohen DS, Colliver JA, Robbs RS, Swartz MH. A large-scale study of the reliabilities of checklist scores and ratings of interpersonal and communication skills evaluated on a standardized-patient examination. Adv Health Sci Educ Theory Pract 1996;1:209-13.
17. Dawson B, Trapp RG. Basic & clinical biostatistics. 4th ed. New York (NY): Lange Medical Books/McGraw-Hill; 2004.
18. Lee YH, Park JH, Ko JK, Yoo HB. The change of CPX scores according to repeated CPXs. Korean J Med Educ 2011;23:193-202.