Predictive Value of the Korean Academy of Family Medicine In-Training Examination for Certifying Examination

Jung-Jin Cho*, Ji-Yong Kim[1]

Department of Family Medicine, Hallym University Sacred Heart Hospital, Hallym University College of Medicine, Anyang, [1]Kangbuk Samsung Hospital Seocho Medical Clinic, Seoul, Korea

Background: In-training examination (ITE) is a cognitive examination similar to the written test, but it is different from the Clinical Practice Examination of the Korean Academy of Family Medicine (KAFM) Certification Examination (CE). The objective of this is to estimate the positive predictive value of the KAFM-ITE for identifying residents at risk for poor performance on the three types of KAFM-CE.

Methods: 372 residents who completed the KAFM-CE in 2011 were included. We compared the mean KAFM-CE scores with ITE experience. We evaluated the correlation and the positive predictive value (PPV) of ITE for the multiple choice question (MCQ) scores of 1st written test & 2nd slide examination, the total clinical practice examination scores, and the total sum of 2nd test.

Results: 275 out of 372 residents completed ITE. Those who completed ITE had significantly higher MCQ scores of 1st written test than those who did not. The correlation of ITE scores with 1st written MCQ (0.627) was found to be the highest among the other kinds of CE. The PPV of the ITE score for 1st written MCQ scores was 0.672. The PPV of the ITE score ranged from 0.376 to 0.502.

Conclusion: The score of the KAFM ITE has acceptable positive predictive value that could be used as a part of comprehensive evaluation system for residents in cognitive field.

Keywords: In-Training Examination; Certification Examination; Multiple Choice Question; Clinical Practice Examination; Positive Predictive Value

Korean Journal of Family Medicine

INTRODUCTION

Based on the time-point of assessment, clinical assessment is divided into diagnostic, formative, and summative evaluation.[1,2] The in-training examination (ITE) that is conducted for residents during training courses could be used as a diagnostic evaluation if it is conducted at the time when starting a resident training course, but since it is carried out during a resident training course, it is a representative formative evaluation. Namely, while ITE measures the clinical performance capacities of individual residents to tell how much of the educational goal was achieved during a training course, the results of ITE could also provide both residents and educators with feedback to improve the medical examination

performance capacities of individual residents. Meanwhile, the Korean Academy of Family Medicine-Certification Examination (KAFM-CE) is a summative evaluation since it is conducted after completing a training course.

ITE scores are often used as individual resident performance indicators. Especially, it has to be able to predict which residents may fail at CE. Although studies have been carried out on the predictability of ITE for CE, studies mainly discussed the predictability for a written CE, which is a similar cognitive assessment exam.[3-9] It is hard to find the studies about the predictability of KAFM-ITE for KAFM-CE.

Miller[10] divided clinical assessment steps into cognitive field and behavior field. Clinical practice examination (CPX) using standardized patients (SPs) is widely used in medical education to assess the behavior field (skill & attitude, performance and action), which cannot be measured by a written examination. An assessment of health professionals at the level of 'shows how' or simulation is a higher level of assessment according to Miller's pyramid than an assessment of 'know' or 'how to know'.

Being a higher level of assessment than a cognitive assessment, the CPX was introduced in KAFM-CE as one of the 2nd examination in 2009. So the KAFM-CE takes the following 3 types of examinations that include the 1st written examination, 2nd slide examination and 2nd CPX.

The 1st KAFM-CE is similar to KAFM-ITE as both evaluate resident's cognitive field level. Although the 2nd slide examination uses the slide presentation of figures and pictures in its test, it is in effect a cognitive assessment just like KAFM-ITE. Namely, the KAFM-ITE is similar to the 1st written and 2nd slide examinations of the KAFM-CE, but different from the CPX by testing a behavior field level. Therefore, the present study aimed to evaluate the positive predictive value of the KAFM-ITE for the risk of low score group at the KAFM-CE and to know the correlation of 3 types of the KAFM-CEs to KAFM-ITE score of the third-year residents. In order to compare the educative experience of ITE on subsequent CE, the mean scores of 3 types of the KAFM-CEs were compared according to the ITE experiences.

METHODS

1. Subjects and Methods

The present study subjected 372 residents, who took KAFM-CE on January, 2011.

1) Progression of KAFM-ITE

The KAFM-ITE with 125 multiple choice questions (MCQs) was conducted on September, 2010 to the third-year residents. The examination lasted 2 hours. Among the 372 examinees who took the KAFM-CE, 275 examinees took this ITE examination. The score of 0.4 was assigned to each question to calculate the total score, then the total score was converted into 100 full score scale, which was expressed as 'ITE-MCQ'.

2) Progression of 1st written examination in KAFM-CE

The KAFM-CE 1st written examination with 250 MCQs was conducted on January, 2011. Total examinees were 372, and the examination lasted 4 hours. The score of 0.4 was assigned to each question to calculate the total score, which was expressed as '1st written-MCQ'.

3) Progression of 2nd examination in KAFM-CE

For those who passed the 1st examination, the 2nd slide examination consisted of 50 MCQs by using figures and pictures and the CPX examination that was comprised by 2 cases with SPs were conducted on January, 2011. The examinees of the slide and CPX examinations were 267 residents.

The total 50 minutes slide examination was progressed by assigning 1 minute for each question by showing a slide that the necessary figures or pictures for the question for 30 seconds under the turned off light. The examination score of 2 was assigned for each question to calculate the total score, then the total score was converted into 100 full score scale, which was expressed as '2nd slide MCQ'.

Examinees were divided into AM and PM groups for CPX examination. Total 4 cases were developed to operate different cases for the AM and PM groups. Each examinee visits two examination rooms continuously. The total 25 minutes examination was progressed by assigning 10 minute for each case. 12 SPs were duplicated for each case. A single SP received the medical examination from 16 examinees. In addition, the faculty

doctors were assigned to each room as evaluators not to interfere by the interests of attending schools and training hospitals of examinees. For each item of the scoring checklists, the score of 1 was assigned to calculate the total score. The total score of an examinee was converted into 100 full score scale, which was expressed as '2nd CPX total'.

The domain of CPX consisted of history taking, physical examination, patient education (information sharing), patient-doctor relationship, global rating of the faculty doctor and SP's satisfaction. The evaluation for history taking and patient education were scored by following the checklists of the developed dichotomy by scoring 1 for conducted, 0 for not conducted. The physical examination item was scored 1 for appropriately conducted, 0.5 for not appropriately conducted, and 0 for not conducted. The global rating of the faculty doctor and the SPs were scored according to the 4 scores Likert scale. The score for the each domains of CPX was used in the analysis by converting the score into 100 full scales.

Since the 'total sum of 2nd test' reflected 70% of the 2nd slide examination and 30% of the CPX of the 2011 KAFM-CE, it is the 100 full scale converted score that was converted by reflecting the ratio.

2. Statistical Analysis

Gender distribution of residents who applied the KAFM-ITE was analyzed by using chi square test, and the student T-test was used for the comparison of age and the mean score of the KAFM-CE according to KAFM-ITE experience. The relationship of the scores of KAFM-ITE and KAFM-CE were measured by using the Pearson's correlation coefficient (r). The correlation coefficients beyond ± 0.5 are typically regarded as strong, whereas coefficients between zero to ± 0.5 are usually regarded as weak.[11] By using the two-sided test, the P-value less than 0.01 was considered to have a statistical significance.

As with diagnostic tests used in clinical practice, we could use KAFM-ITE scores to identify residents who are at the risk of failing the KAFM-CE. A resident who failed at the KAFM-CE will be scored below an arbitrary cut point of the KAFM-ITE (true positive) or obtained a score above the arbitrary cut point of the KFAM-ITE (false negative). A resident who passes the KAFM-CE will have a score below the arbitrary cut point of the KAFM-ITE (false positive) or obtain a score above the arbitrary cut point of the KAFM-ITE (true negative). These four outcomes can be calculated from the data on the positive predictive value (PPV) of the KAFM-ITE.

But if we use the real data of pass or fail on KAFM-CE, the PPV depends on the level of difficulty of KAFM-CE. To estimate the PPV of the KAFM-ITE, we assumed that the ITE scores that placed a resident in the lower 7% of the distribution when compared to other residents in the same year of training indicated an at-risk resident on that scale. Selection of the seventh percentile is arbitrary, but this percentile rank approximates the percent of examinees that fail the KAFM-CE. This method used in a study for the American Board of Family Practice.[6]

RESULTS

1. Comparison of Basic Characteristics and Mean Scores on KAFM−CE by KAFM−ITE Experience

The applicants of the 1st written MCQ test were 372, and applicants of the 2nd slide examination & CPX were 367 in KAFM-CE. Among the 372 applicants, 277 completed the KAFM-ITE. There was no significant difference in the distribution of age and sex according to KAFM-ITE experience. The mean MCQ score of 1st written test in the group was 80.4 ± 5.4 that was significantly higher than the mean MCQ score of 77.3 ± 7.4 without applying KAFM-ITE. But no significant difference was observed for the mean MCQ scores of the 2nd slide examination, the total CPX, and the total sum of 2nd test according to the ITE experiences (Table 1).

2. Correlation Between KAFM−ITE and KAFM−CE Scores

The correlation coefficient between the ITE-MCQ score and 1st written MCQ scores was 0.627 by showing a strong positive correlation. The correlation coefficient between the ITE-MCQ score and the total sum of 2nd test scores was 0.502, the correlation coefficient of the 2nd slide MCQ score to the ITE score was 0.429, and the correlation coefficient of the 2nd CPX total score to the ITE score was 0.376.

Correlation coefficients of each domains of CPX to ITE-MCQ resulted to show weak positive relationship to patient-

Table 1. Baseline characteristics of attendees in Certifying Examination for Family Medicine specialty.

		In-training examination (−) (n = 96)	In-training examination (+) (n = 276)	P-value
Sex	Male	65	166	0.338
	Female	34	110	
Age		33.9 ± 3.4	33.1 ± 3.1	0.372
1st written MCQ (n = 372)		77.3 ± 7.4	80.4 ± 5.4	<0.001
Total sum of 2nd test (n = 367)		67.8 ± 6.1	71.1 ± 6.5	0.711
2nd slide MCQ (n = 367)		72.5 ± 6.3	75.9 ± 7.8	0.706
2nd CPX total (n = 367)		57.0 ± 9.2	59.8 ± 8.2	0.063

Values are presented as number or mean ± standard deviation. P-value by chi square test or student T-test.
1st written MCQ: the score of 1st written multiple choice question examination, total sum of 2nd test: the score of sum (0.7 × 2nd slide MCQ score + 0.3 × 2nd CPX score), 2nd slide MCQ: the score of 2nd slide multiple choice question examination, 2nd CPX total: total score of clinical practice examination.

Table 2. Correlation coefficient of in-training examination MCQ score with 1st written MCQ scores, 2nd total score, 2nd slide MCQ scores, & 2nd CPX scores in Certifying Examination for Family Medicine specialty.

	Correlation coefficient	P-value
1st written MCQ	0.627	<0.001
Total sum of 2nd test	0.502	<0.001
2nd slide MCQ	0.429	<0.001
2nd CPX total	0.376	<0.001
History taking	0.258	<0.001
Physical examination	0.260	<0.001
Information sharing	0.125	0.039
Patient & physician interaction	0.318	<0.001
Global rating of faculty doctor	0.382	<0.001
Global rating of standardized patient	0.229	<0.001

P-value by Pearson's correlation.
1st written MCQ: the score of 1st written multiple choice question examination, total sum of 2nd test: the score of sum (0.7 × 2nd slide MCQ score + 0.3 × 2nd CPX score), 2nd slide MCQ: the score of 2nd slide multiple choice question examination, 2nd CPX total: total score of clinical practice examination.

physician relationship (r = 0.318), global rating of faculty (r = 0.382) for history taking (r = 0.251), physical examination (r = 0.262), patient education (r = 0.125), and satisfaction of SP (r = 0.229) (Table 2).

Table 3. Positive predictive value of in-training examination MCQ score for 1st written MCQ scores, 2nd total score, 2nd slide MCQ scores, & 2nd CPX scores in Certifying Examination for Family Medicine specialty.

	Above 7% or not
1st written MCQ	0.69
Total sum of 2nd test	0.16
2nd slide MCQ	0.27
2nd CPX total	0.15

1st written MCQ: the score of 1st written multiple choice question examination, total sum of 2nd test: the score of sum (0.7 × 2nd slide MCQ score + 0.3 × 2nd CPX score), 2nd slide MCQ: the score of 2nd slide multiple choice question examination, 2nd CPX total: total score of clinical practice examination.

3. The Positive Predictive Value of KAFM−ITE Score for the KAFM−CE Score

The PPV of ITE-MCQ score for the 1st written MCQ score was 0.69. The PPV of the total sum of 2nd test for the lower score group was 0.16. The PPV of ITE-MCQ score for the lower score group on the 2nd slide MCQ score was 0.27, and PPV of ITE-MCQ for the lower score group on the 2nd CPX total score was 0.15 (Table 3).

DISCUSSION

In- training examination that is conducted during the resident training program is the formative evaluation that can evaluate how much of the educational goal was achieved by an education program. For the resident training, the examination results could be used variously.

To compare the educative experience of KAFM-ITE on subsequent CE, the mean scores of 3 types of the KAFM-CEs were compared. The 1st written MCQ score of the group applying KAFM-ITE was significantly higher than the group not applying ITE. The correlation coefficient of the 2nd slide MCQ score to the ITE-MCQ score was 0.429 by showing high correlation of the score, but did not show difference to the ITE application. Also, no significant difference was observed by the ITE application in CPX score. The 1st written MCQ test revealed a similarity to the ITE test, but different from the 2nd slide MCQ. Therefore, the learning effects of applying previous ITE test for subsequent CE could be observed only in the same format and same level test field. Of course, relatively qualified residents may apply the ITE test, but no score difference was observed at the other CE. Therefore, it is proper to be interpreted as the educative experience of the previous test.

The ITE test could be used to differentiate high risk residents who can be failed at the CE test. Among the various applicable indexes that could be used for that purpose, the present study applied the positive predictive value and the correlation coefficient.

The results of reviewing the correlation between ITE-MCQ score and each type of the CE test scores resulted to show a highest association with the same cognitive field level test; 1st written MCQ score vs. ITE-MCQ. However, differently formatted cognitive field test of the 2nd slide MCQ test and the total sum of the 2nd test reflecting 70% of 2nd slide MCQ showed about a similar coefficient of 0.5. The result is very similar to the result that reported the correlation coefficient of 0.627 between the 4th year Laboratory Medicine resident ITE test and CE written test scores.[12]

The 2nd CPX total score assesses the behavior fields; its correlation coefficient was relatively low and weak in 0.376, which means a different level of CPX compared to the ITE in terms of Miller's pyramid.[10] Among the domains of CPX,

the lowest correlation coefficient was observed in information sharing. The correlation shows a similar result to other studies that reviewed the correlation between cognitive written test and CPX test scores. In a study of subjecting medical school students, the written test score was highly correlated to the physical examination, which was followed by patient education, history taking, but clinical etiquette resulted to show no association.[13] In the study of analyzing the CE score for neurology specialist, no association was observed in the patient education or physical examination to written test.[14]

Rather the correlation coefficient, the positive predictive value is more proper in identifying the high risk residents who could be failed at the CE. Namely, confirming how well the low score group of ITE can predict high risk of residents who could be failed at the CE is important in utilizing the ITE results. Previously, several studies reported the usefulness of ITE score in predicting CE score. In foreign articles, several studies reported about the usefulness of ITE score in predicting CE score.[3-9] In Korea, the study for otolaryngology specialist was reported that the continuously low scored residents in ITE resulted with low score in CE, and the high scored residents in ITE generally resulted with high CE score.[15] In our study, the PPV for the 1st written test score was 0.69, which is the similar result to the PPV of 0.72 in the study on the American Board of Family Practice.[4]

The PPV of the differently formatted same cognitive field test of the 2nd slide MCQ score was found to be 0.27. The results of the low PPV even though the test is in the same cognitive level could be interpreted in various aspects. Seeing that no difference was observed in the slide test score by the ITE application, the absence of learning effect and the right skewed distribution of 2nd slide MCQ score could be suggested as the explanation.

The PPV of the 2nd CPX total score and the total sum of the 2nd test was very low by resulting 0.15. Since there are limited studies that predicted the other level test as behavioral field, the result is hard to be compared. However, since the result indicated that the ITE alone is difficult to predict the lower score of CPX, there has to be an effort to identify high risk residents in behavior field. In the other aspect, since the difference between the PPV of the total sum of the 2nd test and the 2nd CPX total score was very small, it indicated the high CPX score variability even though its lower 30% reflection ratio at the total sum of the 2nd test, and it also reversely showed that the CPX test score has a large effect in

determining the low score group.

Contrary to the fact that most of previous studies mainly evaluated the predictability of ITE for written CE, the present study is unique to evaluate the predictive values for slide examination and CPX along with the PPV for the written test. In addition, the present study analyzes the learning effect of applying ITE.

REFERENCES

1. Bloom BS, Hastings JT, Madaus GF. Handbook on formative and summative evaluation of student learning. New York: McGraw-Hill; 1971.

2. Seong TJ. Current education evalution. Seoul: Hakjisa; 2005.

3. Althouse LA, McGuinness GA. The in-training examination: an analysis of its predictive value on performance on the general pediatrics certification examination. J Pediatr 2008;153:425-8.

4. Babbott SF, Beasley BW, Hinchey KT, Blotzer JW, Holmboe ES. The predictive validity of the internal medicine in-training examination. Am J Med 2007;120:735-40.

5. Klein GR, Austin MS, Randolph S, Sharkey PF, Hilibrand AS. Passing the Boards: can USMLE and Orthopaedic in-Training Examination scores predict passage of the ABOS Part-I examination? J Bone Joint Surg Am 2004;86:1092-5.

6. Replogle WH, Johnson WD. Assessing the predictive value of the American Board of Family Practice In-training Examination. Fam Med 2004;36:185-8.

7. Kearney RA, Sullivan P, Skakun E. Performance on ABA-ASA in-training examination predicts success for RCPSC certification. American Board of Anesthesiology-American Society of Anesthesiologists. Royal College of Physicians and Surgeons of Canada. Can J Anaesth 2000;47:914-8.

8. Grossman RS, Fincher RM, Layne RD, Seelig CB, Berkowitz LR, Levine MA. Validity of the in-training examination for predicting American Board of Internal Medicine certifying examination scores. J Gen Intern Med 1992;7:63-7.

9. Leigh TM, Johnson TP, Pisacano NJ. Predictive validity of the American Board of Family Practice In-Training Examination. Acad Med 1990;65:454-7.

10. Miller GE. The assessment of clinical skills/competence/performance. Acad Med 1990;65(9 Suppl):S63-7.

11. Glaser AN. High-yield TM biostatistics. Baltimore: Lippincott Williamd & Wilkins; 2001.

12. Huh J, Kim J, Park J, Kim H. Correlation of in-training examination score with the residency program or the score of the board examination of laboratory medicine. Korean J Lab Med 2006;26:227-31.

13. Kim SW, Lim SY, Mun GH, Shin JS. Introduction of OSCE (Objective Structured Clinical Examination) to board licensure examination of plastic surgeons. Korean J Med Educ 2006;18:153-60.

14. Kim BJ, Sung JJ, Park HK, Seo DW, Chung CS, Yoon BW. Clinical performance examination utilizing standardized patients in board examination: based on the board examination of Korean Neurological Association for three years. Korean J Med Educ 2011;23:127-35.

15. Jeon SY, Chung MH, Kim KH, Hong WP. Validity of the intraining examinations and the board examinations: an experience in the Korean Society of Otolaryngology. Korean J Med Educ 1997;9:151-7.