

2011-2016년 의학교육평가컨소시엄 임상종합평가의 효과성

이상엽^{1,2} · 이예리^{2,3} · 김미경^{2,4}¹부산대학교 의과대학 의학교육학교실, ²의학교육평가컨소시엄, ³한국외국어대학교 의과대학원협회 의학교육평가사업단, ⁴중앙대학교 의과대학 병리학교실

Effectiveness of Medical Education Assessment Consortium Clinical Knowledge Mock Examination (2011-2016)

Sang Yeoup Lee^{1,2} · Yeli Lee^{2,3} · Mi Kyung Kim^{2,4}¹Department of Medical Education, Pusan National University School of Medicine, Yangsan; ²Medical Education Assessment Consortium; ³Medical Education Assessment Corporation, The Korea Association of Medical Colleges; ⁴Department of Pathology, Chung-Ang University College of Medicine, Seoul, Korea

Good assessment is crucial for feedback on curriculum and to motivate students to learn. This study was conducted to perform item analysis on the Medical Education Assessment Consortium clinical knowledge mock examination (MEAC CKME) (2011-2016) and to evaluate several effects to improve item quality using both classical test theory and item response theory. The estimated difficulty index (P) and discrimination index (D) were calculated according to each course, item type, A (single best answer)/R (extended matching) type, and grading of item quality. The cut-off values used to evaluate P were: >0.8 (easy); 0.6-0.8 (moderate); and <0.6 (difficult). The cut-off value for D was 0.3. The proportion of appropriate items was defined as those with P between 0.25-0.75 and D \geq 0.25. Cronbach α was used to assess the reliability and was compared with those of the Korean Medical Licensing Examination (KMLE). The results showed the recent mean difficulty and decimation index was 0.62 and 0.20 for the first MEAC CKME and 0.71 and 0.19 for the second MEAC CKME, respectively. Higher grade items evaluated by a self-checklist system had better D values than lower grade items and higher grade items gradually increased. The preview and editing process by experts revealed maintained P, decreased recall items, increased appropriate items with better D values, and higher reliability. In conclusion, the MEAC CKME (2011-2016) is deemed appropriate as an assessment to evaluate students' competence and prepare year four medical students for the KMLE. In addition, the self-checklist system for writing good items was useful in improving item quality.

Corresponding author

Sang Yeoup Lee
Department of Medical Education,
Pusan National University School of
Medicine, 49 Busandaehak-ro,
Mulgeum-eup, Yangsan 50612, Korea
Tel: +82-51-510-8124
Fax: +82-51-510-8125
E-mail: saylee@pnu.edu
https://orcid.org/0000-0002-3585-9910

Received: May 16, 2017

1st revised: September 14, 2017

2nd revised: November 24, 2017

Accepted: December 12, 2017

Keywords: Medical education, Medical measurement, Medical student

서 론

평가는 학생과의 대화의 장이기도 하고 학생이 공부하고 싶은 마음이 생기도록 하며, 학생은 평가를 통해 자신이 부족한 부분을 알게 된다. 또 학교는 평가를 통해 취약한 교과과정이 있는지를 살펴볼 수도 있고, 심지어 평가는 과락 여부를 결정하는 기준이 되기도 한다[1]. 의과대학·의학전문대학원(이후 의과대학으로 통칭)의 교육목표는 의학지식을 습득하고 임상술기능력을 길러 실제 진료에서 문제를 해결하고 전인적인 치료를 행하는 의사가 되도록 하는 것이며[2], 이를 위해 학생이 갖추어야 하는 의학지식과 임상능력을 평가하기 위해 필기시험, 구두시험, 임상술기시험 등 다양한 평가방법이 활용된다. 이 중에서도 필기시험의 경우, 좋은 문항을 잘 개발하기만 한다면 Anderson 등[3]이 제시한 교육목표 분류법,

즉 기억(remembering), 이해(understanding), 적용(applying), 분석(analysing), 평가(evaluating), 창의(creating) 중에서 창의를 제외한 나머지는 모두 필기시험으로 측정할 수 있다. 물론 이러한 목적을 이루려면 도입하려는 평가가 측정도구의 양호도지수, 즉 난이도, 변별도, 신뢰도, 타당도 등이 적절한 평가라는 것이 전제가 되어야 한다[4]. 잘 아는 대로 우리나라에서 의사가 되려면 의과대학을 졸업한 후에 반드시 의사국가시험에 합격해야 한다. 의과대학마다 이 최소한의 목표를 달성하기 위해 나름의 교육과정을 운영하고 있고, 그들의 의학과 4학년이 과연 의사국가시험 필기시험을 제대로 준비하고 있는지를 알고 싶어 한다. 따라서 국내 여러 의과대학 의학과 4학년 학생들이 공동으로 평가를 받게 되면 그 속에서 자기 의과대학의 역량을 점검해 볼 수 있기에 많은 의과대학이 이러한 성격의 평가의 시행을 원하는 바이다.

20개 의과대학이 이러한 목적을 가지고 2008년 11월 13일 기본 의학교육평가컨소시엄을 조직하였고, 2011년 5월 13일부터는 의학 교육평가컨소시엄으로 명칭을 변경하여 운영해 왔다. 2013년부터는 전국 41개 의과대학 중 39개 의과대학이, 2015년부터는 40개 의과대학이 참여하여 주로 의학과 1-2학년을 위한 기초종합평가와 의학과 3-4학년을 위한 임상종합평가 시험에 응시하고 있다. 2009년 8월 26-27일 및 9월 9-10일로 나누어 20개 의과대학에서 처음으로 임상종합평가를, 2009년 12월 23일에는 처음으로 기초종합평가를 시행하였다. 임상종합평가는 2010년부터 기존의 연 1회에서 연 2회 시행으로 횟수를 늘려 1차 임상종합평가는 8월 말, 2차 임상종합평가는 12월 초에 시행하고 있고, 기초종합평가는 계속해서 연 1회 시행하고 있다. 기초종합평가와 임상종합평가 중 한 가지만 응시하는 경우가 각각 1개 대학, 2개 대학이 있어 대개 39개 의과대학이 임상종합평가에, 38개 의과대학이 기초종합평가에 응시하고 있다. 의사국가시험 필기시험은 5지 선다형, 즉 단일정답형(A형)과 확장결합형(R형) 문항으로 구성되며 시험과목 수, 교시별 문제 수, 문제 형식, 시험시간 등이 명시된 시험시간표는 매년 조금씩 변경되고 있다. 의학교육평가컨소시엄은 가능한 의사국가시험 필기시험 준비에 도움이 되고자 매해 달라지는 의사국가시험 필기시험의 시험시간표와 구성에 맞추어 매년 자체적으로 출제계획표를 구성하여 출제계획표에 맞추어 모든 참여대학이 참여하여 3배수에 해당하는 문항을 개발하고 수정작업을 거쳐 1차 임상종합평가에 사용될 문항을 먼저 선정하고, 이후 2차 임상종합평가 문항을 선정한다. 나머지 미선정된 문항은 폐기하거나 다음에 사용하기 위해 문항은행에 보관하고 있다.

의학교육평가컨소시엄은 컨소시엄 참여대학 의학생이 보다 좋은 문항으로 시험을 치르도록 하기 위해 그동안 다각도의 노력을 기울여 왔다. 2013년 시험의 질을 전문적으로 관장하는 문항개발관리소 위원회가 구성되었고 2013년 1차 임상종합평가부터는 문항의 질 등급을 나누기 위한 자가점검표를 만들어 활용하기 시작하였다. 2013년 기초종합평가부터는 출제된 문항을 수정하고 선정하기 전에 전문위원들이 문항을 사전에 검토하는 사전검토(preview)과정을 시행하였고, 2014년부터는 해당 과정을 임상종합평가까지 확대하였다. 아울러 2015년 임상종합평가부터는 시험지 인쇄 전에 선정이 완료된 문항을 최종적으로 다시 검토하고 시험지 편집을 재확인하기 위한 편집워크숍을 신설하여 수행하였다. 따라서 저자 등은 첫째, 2011년부터 2016년까지 연도별로 임상종합평가문항을 연도별, 과목별, 유형별, 등급별, A/R형별로 구분한 후 고전검사이론과 문항반응이론에 따른 난이도, 변별도, 신뢰도 등을 분석하고 둘째, 컨소시엄의 문항의 질 관리 노력에 따른 변화를 살펴봄, 셋째, 이를 의사국가시험의 공개된 분석결과와 비교해봄으로써 임상종합평가의 효과성을 평가하고자 하였다.

연구대상 및 방법

1. 연구자료

1차 임상종합평가 문항에 대한 4학년 응답결과는 2011-2016년, 2차 임상종합평가는 2011년 자료가 확보되지 않아 2012-2016년 자료를 분석대상으로 하였다(Table 1). 이때 과목(총론, 각론 및 보건의약관계법규), 응시시기, 문항유형(암기형, 해석형, 문제해결형, 응

Table 1. Basal characteristics of MEAC CKME, 2011-2016

Year	Total items	Courses			MCQ types		Medical school	Year 4 student	Total score	Courses		
		General medicine	Specialized medicine	Medical law	A type ^{a)}	R type ^{b)}				General medicine score	Specialized medicine score	Medical law
1st MEAC CKME												
2011	450 (100.0)	105 (23.3)	325 (72.2)	20 (4.4)	391 (86.9)	59 (13.1)	30	2,252	288.5±41.6	61.3±9.9	215.0±31.1	12.2±3.7
2012	400 (100.0)	80 (20.0)	300 (75.0)	20 (5.0)	348 (87.0)	52 (13.0)	34	2,575	247.5±33.3	52.7±6.7	183.8±26.8	11.0±2.9
2013	400 (100.0)	80 (20.0)	300 (75.0)	20 (5.0)	344 (86.0)	56 (14.0)	38	2,974	232.9±33.6	47.4±7.6	173.3±25.9	12.2±2.8
2014	400 (100.0)	74 (18.5)	306 (76.5)	20 (5.0)	349 (87.3)	51 (12.7)	38	2,956	239.1±49.0	42.6±8.1	187.6±39.8	8.9±3.4
2015	400 (100.0)	80 (20.0)	300 (75.0)	20 (5.0)	340 (85.0)	60 (15.0)	39	3,042	251.1±34.5	52.1±7.3	187.7±26.8	11.3±3.2
2016	360 (100.0)	60 (16.7)	280 (77.7)	20 (5.6)	302 (83.9)	58 (16.1)	39	2,986	222.6±30.2	38.0±4.8	171.6±25.2	13.0±3.1
2nd MEAC CKME												
2012	400 (100.0)	82 (20.5)	298 (74.5)	20 (5.0)	348 (87.0)	52 (13.0)	30	1,982	276.1±30.8	55.3±6.4	207.3±24.9	13.5±2.2
2013	400 (100.0)	80 (20.0)	300 (75.0)	20 (5.0)	344 (86.0)	56 (14.0)	37	2,641	271.5±30.9	51.6±7.0	205.7±23.7	14.2±2.7
2014	400 (100.0)	80 (20.0)	302 (75.5)	18 (4.5)	344 (86.0)	56 (14.0)	38	2,762	284.2±29.8	54.6±6.4	216.3±23.5	13.4±2.3
2015	400 (100.0)	80 (20.0)	300 (75.0)	20 (5.0)	340 (85.0)	60 (15.0)	39	2,885	284.8±28.7	52.8±6.3	218.9±22.5	13.0±2.6
2016	360 (100.0)	60 (16.7)	280 (77.7)	20 (5.6)	302 (83.9)	58 (16.1)	39	2,951	255.5±25.7	38.8±5.1	203.1±20.7	13.7±2.5

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

MEAC CKME, Medical Education Assessment Consortium clinical knowledge mock examination; MCQ, multiple choice question.

^{a)}Single best answer. ^{b)}Extended matching.

용형), 문항의 형식(A형/R형), 문항의 질 등급(A, B, C) 등의 요인을 포함하였다. 응용형은 미국의사국가시험 필기시험의 분류대로 해석형과 문제해결형을 합친 것이다. 문항의 질 등급은 문항개발관리소위원회가 미국의사국가고시 기초의학 및 임상의학 문항개발 지침[5]을 바탕으로 개발한 자가점검표에 따라 나누었고, 문항개발과정에서 보다 좋은 문항을 개발하도록 할 뿐 아니라 개발된 문항을 정리하는 워크숍을 통해 보다 더 좋은 문항이 되도록 안내자 역할을 하였다. 자가점검표는 다음과 같이 9가지 항목으로 구성하였다. (1) 중요하고 흔한 것을 다루거나 혹은 다소 드물더라도 놓치면 위중해지는 것을 다루면 1점, 아니면 0점, (2) 해석 또는 문제해결형이면 1점, 아니면 0점, (3) 첫 문장이 환자가 표현하는 말로 시작하면 1점, 아니면 0점, (4) 신체진찰 내용이 기술되면 1점, 아니면 0점, (5) 문항줄기에 검사실 검사가 제시되어 있으면 1점, 아니면 0점, (6) 문항줄기에 영상검사 자료가 제시되어 있으면 1점, 아니면 0점, (7) 커버 더 옵션 규칙을 지켰으면 1점, 아니면 0점, (8) 답가지가 균질되면 1점, 아니면 0점, (9) 틀린 4가지 답가지가 모두 매력적이면 1점, 아니면 0점을 부여하였다. 이 중에 총점이 7-9점이면 좋은 문항(A), 4-6점이면 보통 문항(B), 1-3점이면 나쁜 문항(C)으로 나누었다.

2. 통계분석

문항분석에는 고전검사이론과 문항반응이론을 모두 사용하였다. 전체적으로 그리고 과목별(총론, 각론 및 보건의약관계법규), 문항 유형별(암기형, 해석형, 문제해결형, 응용형), 문항의 형식별(A형/R

형), 문항의 질 등급별(A, B, C)로 구분해서 고전검사이론에 의한 난이도, 교정난이도(문항추측도가 제거된 난이도), 변별도를 구했고, 문항반응이론을 적용한 것은 3모수 로지스틱모형(TCL 8.3 for Windows; Scriptics Corp., Mountain View, CA, USA)으로 난이도, 변별도 및 추정치(본 논문에서는 제시하지 않음)를 구했다. 이때 TCL 8.3 for Windows (Scriptics Corp.)을 사용하여 개발한 알고리즘을 사용하였다. 문항난이도는 문항의 쉽고 어려운 정도를 나타내는 지수로서 높으면 쉬운 문항이고 낮으면 어려운 문항으로 해석하였다. 문항반응이론에 의한 변별도는 각 문항이 피험자의 능력 수준을 변별할 수 있는 정도를 나타내는 지수로서 높을수록 문항변별도가 높은 문항으로 해석하였다[4]. 고전검사이론에 의한 변별도는 상위 27%와 하위 27%의 집단의 난이도 차이로 구했고 해석은 마찬가지로이다. 한국보건의료인국가시험원에서는 2015년도 79회 의사국가시험 문항분석부터 제시한 총점과의 상관계수로 구하는 변별도와 구분하기 위해 변별도 대신에 분별도라고 부르지만, 여기서는 지난 2011년부터 2016년 임상종합평가의 분석결과와 일관되게 비교하고자 이 값을 사용하였다. 또한 79회, 80회 의사국가시험 문항분석처럼 난이도 0-0.6, 0.6-0.8 및 0.8-1.0으로 구분하여 해당 비율을 구했고, 변별도는 0.3 전후로 구분하여 해당 비율을 구했다. 적정문항의 수와 비율은 고전검사이론의 난이도 0.25-0.75이면서 고전검사이론의 변별도 0.25 이상일 때로 임의로 정하여 산정하였다[6]. 신뢰도는 Cronbach α 값으로 제시하였으며, 참고로 P지역 연합모의고사와 국가고시필기시험의 신뢰도를 함께 제시하였다[7-10].

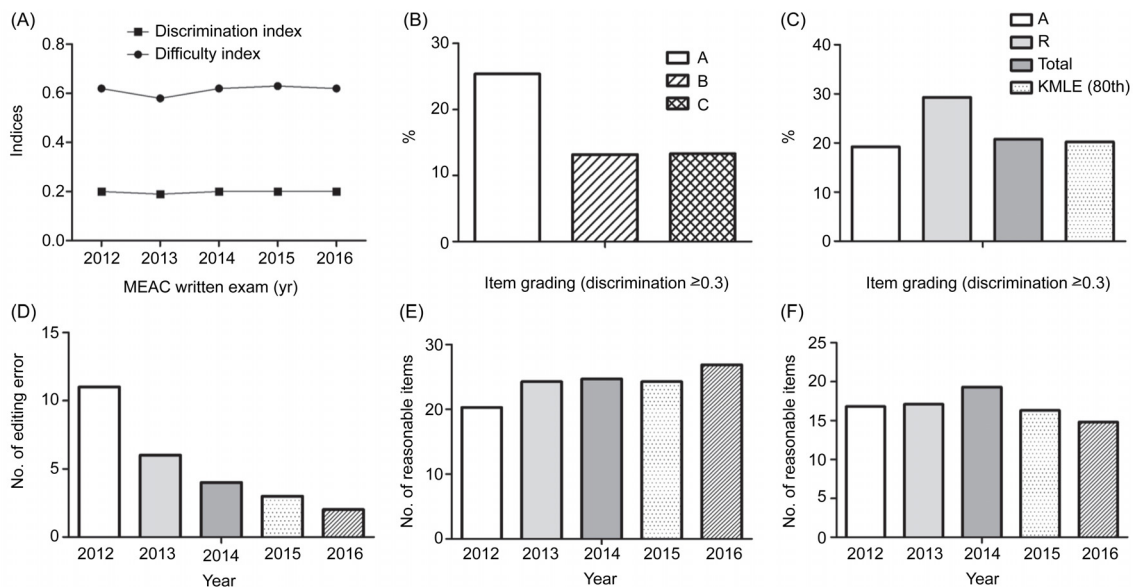


Figure 1. (A) Difficulty index and discrimination index of total exam; (B) discrimination index according to grading of item quality; (C) discrimination index according to A (single best answer)/R (extended matching) types of MEAC clinical knowledge mock exam; (D) number of editing error, (first+second) MEAC clinical knowledge mock exam; (E) percentage of reasonable item, first MEAC clinical knowledge mock exam; and (F) percentage of reasonable item, second MEAC clinical knowledge mock exam, 2012–2016. MEAC, Medical Education Assessment Consortium; KMLE, Korean Medical Licensing Examination.

신뢰도는 시험이 평가하고자 하는 것을 일관성 있게 측정하는지로 시험이 오차 없이 정확하게 측정할 정도를 의미하며, 1에 가까울수록 신뢰도가 높다고 해석하였다.

결 과

1. 난이도

1) 과목별

1차 임상종합평가의 고전검사이론에 의한 난이도는 2011년과 2013년을 제외하고는 0.62-0.63 (100점 만점 점수로 환산하면 62-63점)으로 안정적으로 유지되었고, 의학총론, 의학각론 및 보건의약 관계법규 과목별로도 치우침이 없이 일정한 난이도를 유지하였다 (Figure 1A). 문항반응이론에서는 -0.5부터 0.5까지의 난이도는 중간을 의미하고, 0.5부터 2.0까지는 문항이 어려운 것을 뜻하는데 [11], 문항반응이론에서 임상종합평가는 모두 중간 정도의 난이도를 보였다. 반면에 2차 임상종합평가의 고전검사이론에 의한 난이도는 1차 임상종합평가보다 높은데, 처음에는 0.69, 0.68이었다가 2014년부터는 0.71로 일정하게 나타났다. 과목별로 보면 100점으로 환산했을 때 총론은 1차 임상종합평가에 비해 평균 3.6점 높아진 데 비해 의학각론과 보건의약관계법규의 성적은 각각 10점과 11.8점 높아졌다. 난이도에서 문항을 추측하여 문항의 답을 맞힌 응시자의 비율인 문항추측도를 제거한 난이도가 교정난이도이므로 난이도보다 값이 더 낮게 나타났다(Table 2).

2) 유형별

1차 임상종합평가의 암기형 비율은 2012년 48.4%에서 2016년 11.9%로 무려 1/4로 크게 줄었다. 2차 임상종합평가의 암기형 비율도 2012년 26.5%에서 2016년 10.3%로 절반 이하로 줄었다. 난이도는 지식의 정도를 묻는 유형에 있어서는 1차 임상종합평가에서는 문제해결형 또는 응용형의 난이도지수가 낮게 나타나 의사국가시험과 같은 경향을 보였으나, 2차 임상종합평가에서는 그 차이가 줄어들었다(Table 3).

3) 등급별 및 A/R별

문항등급별 난이도의 차이는 거의 없어 보이지만, 의사국가시험과 마찬가지로 문항의 형태에 있어서는 A형 문항보다는 R형 문항의 난이도지수가 낮은 것으로 볼 때 학생들은 R형 문항을 보다 더 어려워한다는 것을 알 수 있다(Table 4).

2. 변별도

1) 과목별, 유형별

1차 임상종합평가의 고전검사이론에 의한 변별도는 전체적으로는 0.20이고, 과목별로는 의학각론이 나머지보다 높았다. 문제유형별로는 차이가 없었다. 반면에, 2차 임상종합평가는 전체적으로는 변별도가 0.18이고, 과목별로는 마찬가지로 의학각론이 나머지보다 높았고, 문제유형에 있어서는 문제해결형, 해석형, 암기형 순서로 높았다. 이는 문항반응이론에 의한 변별도 또한 마찬가지였다. 문항반응이론에 따라 3모수 로지스틱모형을 사용한 문항분석에서 문항변별도 지수가 0.34 이하이면 변별력이 거의 없는 문항, 0.3-1.7

Table 2. Difficulty index of MEAC CKME according to courses, 2011-2016

Year	Total			General medicine			Specialized medicine			Medical law		
	P	CP	P (IRT)	P	CP	P (IRT)	P	CP	P (IRT)	P	CP	P (IRT)
1st MEAC CKME												
2011	0.58±0.23	0.55±0.29	-0.47±2.65	0.58±0.25	0.48±0.31	-0.17±2.97	0.66±0.23	0.58±0.29	-0.59±2.57	0.61±0.18	0.51±0.23	-0.02±1.99
2012	0.62±0.24	0.52±0.30	-0.20±2.84	0.66±0.22	0.57±0.27	-0.54±3.11	0.61±0.24	0.52±0.30	-0.15±2.74	0.55±0.23	0.44±0.29	0.42±3.07
2013	0.58±0.24	0.48±0.30	0.04±2.74	0.59±0.25	0.49±0.31	-0.04±2.63	0.58±0.24	0.47±0.30	0.08±2.76	0.61±0.22	0.51±0.28	-0.07±2.90
2014	0.62±0.24	0.53±0.30	-0.19±2.00	0.61±0.24	0.51±0.29	-0.34±2.59	0.64±0.24	0.54±0.30	-0.21±1.84	0.49±0.21	0.36±0.26	0.77±1.64
2015	0.63±0.24	0.53±0.30	-0.37±2.73	0.65±0.22	0.56±0.27	-0.44±2.82	0.63±0.25	0.53±0.31	-0.38±2.70	0.57±0.24	0.46±0.30	-0.04±2.87
2016	0.62±0.24	0.52±0.31	-0.20±2.85	0.63±0.25	0.55±0.32	-0.70±3.54	0.61±0.25	0.52±0.31	-0.10±2.69	0.65±0.17	0.56±0.21	-0.14±2.78
2nd MEAC CKME												
2012	0.69±0.26	0.61±0.32	-0.89±2.65	0.68±0.26	0.59±0.32	-0.79±3.07	0.70±0.26	0.62±0.32	-0.91±2.48	0.68±0.30	0.60±0.37	-0.93±3.39
2013	0.68±0.24	0.60±0.29	-0.95±2.79	0.65±0.27	0.57±0.33	-0.75±3.04	0.69±0.23	0.61±0.29	-0.94±2.71	0.71±0.20	0.64±0.26	-1.85±3.04
2014	0.71±0.24	0.64±0.30	-1.24±2.79	0.68±0.24	0.60±0.30	-0.96±3.15	0.72±0.24	0.65±0.30	-1.30±2.69	0.74±0.21	0.68±0.26	-1.49±2.92
2015	0.71±0.24	0.64±0.29	-1.14±2.87	0.66±0.24	0.58±0.30	-0.36±3.14	0.73±0.23	0.66±0.29	-1.37±2.75	0.65±0.26	0.56±0.32	-0.92±3.16
2016	0.71±0.24	0.64±0.29	-1.28±2.87	0.65±0.25	0.56±0.32	-0.65±3.38	0.73±0.23	0.66±0.29	-1.41±2.74	0.68±0.23	0.60±0.29	-1.45±2.94

Values are presented as mean±standard deviation.

MEAC CKME, Medical Education Assessment Consortium clinical knowledge mock examination; P, difficulty index (by classical test theory); CP, corrected difficulty index (by classical test theory); P (IRT), difficulty index (by item response theory).

Table 3. Difficulty index of MEAC CKME according to item types, 2011–2016

Year	Recall			Interpretation			Problem solving			Application		
	P	CP	P (IRT)	P	CP	P (IRT)	P	CP	P (IRT)	P	CP	P (IRT)
1st MEAC CKME												
2011	0.63±0.25 (218)	0.54±0.31 (218)	-0.34±2.78 (218)	0.66±0.23 (111)	0.58±0.28 (111)	-0.85±2.59 (111)	0.64±0.21 (121)	0.55±0.26 (121)	-0.34±2.43 (121)	0.65±0.22 (232)	0.56±0.27 (232)	-0.58±2.51 (232)
2012	0.60±0.25 (102)	0.51±0.31 (102)	-0.12±3.06 (102)	0.69±0.21 (113)	0.62±0.26 (113)	-0.93±2.75 (113)	0.58±0.24 (178)	0.48±0.30 (178)	0.16±2.59 (178)	0.62±0.23 (298)	0.53±0.29 (298)	-0.22±2.76 (298)
2013	0.59±0.24 (84)	0.49±0.30 (84)	0.23±2.65 (84)	0.60±0.25 (134)	0.50±0.31 (134)	-0.14±2.96 (134)	0.57±0.24 (179)	0.46±0.30 (179)	0.07±2.62 (179)	0.58±0.24 (316)	0.48±0.30 (316)	-0.01±2.76 (316)
2014	0.60±0.21 (30)	0.50±0.26 (30)	-0.18±1.40 (30)	0.67±0.23 (135)	0.59±0.29 (135)	-0.49±1.97 (135)	0.60±0.24 (232)	0.50±0.31 (232)	-0.01±2.06 (232)	0.63±0.24 (367)	0.53±0.30 (367)	-0.19±2.04 (367)
2015	0.60±0.23 (63)	0.51±0.28 (63)	-0.11±2.78 (63)	0.65±0.26 (141)	0.57±0.32 (141)	-0.68±2.72 (141)	0.62±0.23 (196)	0.52±0.29 (196)	-0.24±2.71 (196)	0.63±0.24 (337)	0.54±0.30 (337)	-0.42±2.72 (337)
2016	0.63±0.18 (43)	0.53±0.22 (43)	-0.45±2.53 (43)	0.61±0.25 (134)	0.51±0.31 (134)	0.05±2.86 (134)	0.62±0.26 (183)	0.53±0.32 (183)	-0.33±2.92 (183)	0.62±0.25 (217)	0.52±0.32 (317)	-0.17±2.90 (317)
2nd MEAC CKME												
2012	0.68±0.27 (106)	0.60±0.34 (106)	-0.80±2.88 (106)	0.73±0.24 (127)	0.66±0.30 (127)	-1.27±2.52 (127)	0.67±0.26 (161)	0.59±0.32 (161)	-0.63±2.57 (161)	0.69±0.25 (294)	0.62±0.32 (294)	-0.92±2.57 (294)
2013	0.68±0.24 (82)	0.60±0.31 (82)	-0.83±3.04 (82)	0.70±0.24 (118)	0.62±0.31 (118)	-1.26±2.82 (118)	0.68±0.23 (194)	0.60±0.29 (194)	-0.80±2.68 (194)	0.69±0.23 (315)	0.61±0.29 (315)	-0.98±2.73 (315)
2014	0.67±0.29 (41)	0.58±0.36 (41)	-0.90±3.34 (41)	0.73±0.23 (150)	0.67±0.29 (150)	-1.48±2.69 (150)	0.70±0.23 (209)	0.63±0.29 (209)	-1.14±2.75 (209)	0.71±0.23 (359)	0.64±0.29 (359)	-1.28±2.73 (359)
2015	0.59±0.27 (38)	0.49±0.34 (38)	0.04±3.20 (38)	0.72±0.23 (147)	0.65±0.29 (147)	-1.20±2.93 (147)	0.73±0.23 (215)	0.66±0.28 (215)	-1.31±2.73 (215)	0.72±0.23 (362)	0.66±0.29 (362)	-1.27±2.81 (362)
2016	0.70±0.22 (37)	0.63±0.27 (37)	-1.33±2.69 (37)	0.73±0.24 (137)	0.66±0.30 (137)	-1.42±2.92 (137)	0.71±0.24 (184)	0.63±0.30 (184)	-1.18±2.88 (184)	0.72±0.24 (321)	0.64±0.30 (321)	-1.28±2.89 (321)

Values are presented as mean±standard deviation (number of items). MEAC CKME, Medical Education Assessment Consortium clinical knowledge mock examination; P, difficulty index (by classical test theory); CP, corrected difficulty index (by classical test theory); P (IRT), difficulty index (by item response theory).

Table 4. Difficulty index of MEAC CKME according to grading of item quality and A/R type, 2011–2016

Year	Grading of item quality									A/R type					
	A			B			C			Type A ^{a)}			Type R ^{b)}		
	P	CP	P (IRT)	P	CP	P (IRT)	P	CP	P (IRT)	P	CP	P (IRT)	P	CP	P (IRT)
1st MEAC CKME															
2011	0.65±0.22 (227)	0.56±0.27 (227)	-0.49±2.51 (227)	0.62±0.25 (171)	0.52±0.32 (171)	-0.30±2.93 (171)	0.70±0.24 (51)	0.63±0.30 (51)	-0.88±2.21 (51)	0.65±0.23 (391)	0.56±0.28 (391)	-0.47±2.63 (391)	0.61±0.29 (59)	0.51±0.36 (59)	-0.43±2.80 (59)
2012	0.61±0.24 (341)	0.52±0.30 (341)	-0.13±2.87 (341)	0.65±0.21 (57)	0.56±0.26 (57)	-0.51±2.56 (57)	0.62±0.17 (2)	0.52±0.21 (2)	-2.58±4.58 (2)	0.63±0.23 (348)	0.54±0.28 (348)	-0.31±2.83 (348)	0.52±0.29 (52)	0.40±0.36 (52)	0.58±2.77 (52)
2013	0.57±0.24 (241)	0.46±0.31 (241)	0.18±2.70 (241)	0.61±0.23 (138)	0.51±0.29 (138)	-0.11±2.71 (138)	0.64±0.28 (17)	0.55±0.35 (17)	-1.18±3.20 (17)	0.60±0.22 (344)	0.50±0.28 (344)	-0.02±2.71 (344)	0.47±0.31 (56)	0.34±0.39 (56)	0.46±2.92 (56)
2014	0.63±0.24 (278)	0.53±0.30 (278)	-0.20±2.01 (278)	0.62±0.23 (104)	0.53±0.29 (104)	-0.07±1.86 (104)	0.60±0.30 (13)	0.51±0.38 (13)	-0.91±2.75 (13)	0.63±0.23 (349)	0.54±0.29 (349)	-0.16±1.78 (349)	0.56±0.27 (51)	0.45±0.34 (51)	-0.41±3.16 (51)
2015	0.64±0.24 (227)	0.55±0.31 (227)	-0.44±2.53 (227)	0.61±0.23 (157)	0.51±0.29 (157)	-0.10±2.89 (157)	0.70±0.28 (16)	0.62±0.35 (16)	-2.12±3.19 (16)	0.64±0.24 (340)	0.55±0.29 (340)	-0.43±2.78 (340)	0.56±0.26 (60)	0.45±0.33 (60)	-0.03±2.41 (60)
2016	0.61±0.25 (224)	0.52±0.32 (224)	-0.13±2.82 (224)	0.63±0.23 (121)	0.54±0.29 (121)	-0.44±2.91 (121)	0.57±0.18 (15)	0.47±0.22 (15)	0.61±2.95 (15)	0.64±0.23 (302)	0.55±0.29 (302)	-0.39±2.84 (302)	0.51±0.29 (58)	0.39±0.36 (58)	0.78±2.75 (58)
2nd MEAC CKME															
2012	0.70±0.25 (337)	0.62±0.31 (337)	-0.49±2.51 (227)	0.65±0.30 (59)	0.57±0.37 (59)	-0.43±3.09 (59)	0.67±0.28 (4)	0.59±0.35 (4)	-1.59±3.24 (4)	0.71±0.24 (348)	0.64±0.30 (348)	-1.02±2.58 (348)	0.54±0.31 (52)	0.43±0.38 (52)	0.04±2.94 (52)
2013	0.69±0.24 (245)	0.61±0.30 (245)	-0.90±2.80 (245)	0.69±0.23 (132)	0.62±0.28 (132)	-1.12±2.63 (132)	0.59±0.27 (16)	0.49±0.34 (16)	-0.53±3.69 (16)	0.69±0.23 (341)	0.61±0.28 (341)	-0.99±2.76 (341)	0.66±0.28 (56)	0.57±0.35 (56)	-0.66±3.02 (56)
2014	0.71±0.24 (271)	0.64±0.30 (271)	-1.25±2.64 (271)	0.72±0.22 (106)	0.65±0.28 (106)	-1.47±3.03 (106)	0.65±0.21 (20)	0.56±0.26 (20)	-0.04±2.99 (20)	0.71±0.24 (344)	0.64±0.29 (344)	-1.25±2.84 (344)	0.70±0.24 (56)	0.63±0.31 (56)	-1.20±2.48 (56)
2015	0.75±0.23 (210)	0.69±0.28 (210)	-1.62±2.54 (210)	0.67±0.24 (171)	0.59±0.30 (171)	-0.64±3.15 (171)	0.63±0.25 (18)	0.54±0.31 (18)	-0.24±2.91 (18)	0.72±0.23 (340)	0.65±0.28 (340)	-1.20±2.93 (340)	0.66±0.27 (60)	0.57±0.34 (60)	-0.80±2.52 (60)
2016	0.73±0.23 (211)	0.66±0.29 (211)	-1.44±2.86 (211)	0.69±0.24 (136)	0.62±0.30 (136)	-1.10±2.94 (136)	0.66±0.18 (11)	0.58±0.23 (11)	-0.65±2.12 (11)	0.72±0.23 (302)	0.65±0.29 (302)	-1.31±2.89 (302)	0.68±0.27 (58)	0.60±0.34 (58)	-1.17±2.79 (58)

Values are presented as mean±standard deviation (number of items). MEAC CKME, Medical Education Assessment Consortium clinical knowledge mock examination; P, difficulty index (by classical test theory); CP, corrected difficulty index (by classical test theory); P (IRT), difficulty index (by item response theory).
^{a)}Single best answer. ^{b)}Extended matching.

이면 변별력이 어느 정도 있는 문항, 1.7 이상이면 변별력이 매우 높은 문항으로 임의로 분류하였을 때, 전체문항의 변별도는 1차, 2차 임상종합평가 모두 약 0.5였다. 연도별로도 유사한 경향이었다 (Table 5).

2) 등급별

A등급의 비율이 1차 임상종합평가의 경우 2011년 50.4%였으나, 2016년에는 62.2%로 증가하였고, 반면에 C등급의 문항은 2011년 11.3%에서 2016년에는 4.1%로 감소하였다(Table 6). 등급별로는 고전검사이론과 문항반응이론 모두 A등급의 변별도가 B등급이나

Table 5. Discrimination index of MEAC CKME according to course and item type, 2011–2016

Year	Total		Courses								Item types					
			General principles		Each system		Health law		Recall		Interpretation		Problem solving		Application of knowledge	
	D	D (IRT)	D	D (IRT)	D	D (IRT)	D	D (IRT)	D	D (IRT)	D	D (IRT)	D	D (IRT)	D	D (IRT)
1st MEAC CKME																
2011	0.21±0.12 (450)	0.56±0.28 (450)	0.19±0.13 (105)	0.51±0.29 (105)	0.22±0.12 (325)	0.58±0.27 (325)	0.20±0.08 (20)	0.46±0.1 (8 (20))	0.22±0.13 (218)	0.59±0.28 (218)	0.21±0.12 (111)	0.52±0.27 (111)	0.21±0.11 (121)	0.53±0.27 (121)	0.21±0.12 (232)	0.53±0.27 (232)
2012	0.20±0.11 (400)	0.50±0.26 (400)	0.17±0.11 (80)	0.45±0.30 (80)	0.21±0.11 (300)	0.52±0.25 (300)	0.13±0.08 (20)	0.40±0.2 (2 (20))	0.17±0.11 (102)	0.46±0.24 (102)	0.18±0.11 (113)	0.47±0.30 (113)	0.22±0.11 (178)	0.54±0.24 (178)	0.20±0.11 (298)	0.51±0.27 (298)
2013	0.19±0.11 (400)	0.51±0.28 (400)	0.19±0.09 (80)	0.51±0.25 (80)	0.20±0.12 (300)	0.51±0.29 (300)	0.15±0.09 (20)	0.41±0.2 (4 (20))	0.20±0.10 (84)	0.52±0.28 (84)	0.18±0.11 (134)	0.46±0.26 (134)	0.21±0.12 (179)	0.54±0.29 (179)	0.19±0.12 (316)	0.50±0.28 (316)
2014	0.20±0.11 (400)	0.97±0.56 (400)	0.18±0.11 (74)	0.85±0.59 (74)	0.21±0.11 (306)	1.02±0.55 (306)	0.19±0.08 (20)	0.78±0.4 (4 (20))	0.21±0.10 (30)	0.94±0.53 (30)	0.20±0.10 (135)	1.01±0.54 (135)	0.20±0.11 (232)	0.96±0.57 (232)	0.20±0.11 (367)	0.98±0.56 (367)
2015	0.20±0.11 (400)	0.51±0.27 (400)	0.19±0.11 (80)	0.46±0.28 (80)	0.21±0.11 (300)	0.52±0.26 (300)	0.18±0.10 (20)	0.53±0.2 (3 (20))	0.19±0.10 (63)	0.47±0.25 (63)	0.21±0.11 (141)	0.54±0.28 (141)	0.21±0.11 (196)	0.49±0.26 (196)	0.21±0.11 (337)	0.51±0.27 (337)
2016	0.20±0.11 (360)	0.49±0.25 (360)	0.15±0.10 (60)	0.41±0.26 (60)	0.21±0.11 (280)	0.51±0.24 (280)	0.18±0.10 (20)	0.39±0.2 (2 (20))	0.20±0.11 (43)	0.47±0.28 (43)	0.20±0.11 (134)	0.49±0.25 (134)	0.20±0.11 (183)	0.48±0.24 (183)	0.20±0.11 (317)	0.49±0.24 (317)
2nd MEAC CKME																
2012	0.21±0.09 (400)	0.57±0.27 (400)	0.17±0.09 (82)	0.47±0.26 (82)	0.22±0.09 (298)	0.60±0.26 (298)	0.14±0.06 (20)	0.51±0.2 (6 (20))	0.19±0.09 (106)	0.58±0.26 (106)	0.21±0.09 (127)	0.59±0.28 (127)	0.21±0.09 (161)	0.55±0.26 (161)	0.21±0.09 (294)	0.56±0.27 (294)
2013	0.19±0.10 (400)	0.49±0.24 (400)	0.19±0.11 (80)	0.50±0.25 (80)	0.19±0.10 (300)	0.49±0.24 (20)	0.17±0.08 (20)	0.37±0.1 (5 (20))	0.18±0.11 (82)	0.50±0.29 (82)	0.19±0.10 (118)	0.49±0.26 (118)	0.20±0.09 (194)	0.48±0.21 (194)	0.19±0.09 (315)	0.48±0.23 (315)
2014	0.19±0.10 (400)	0.50±0.26 (400)	0.17±0.10 (80)	0.44±0.26 (80)	0.20±0.10 (302)	0.53±0.26 (302)	0.18±0.09 (18)	0.40±0.1 (8 (18))	0.17±0.09 (41)	0.53±0.31 (41)	0.19±0.10 (150)	0.50±0.24 (150)	0.20±0.10 (209)	0.50±0.27 (209)	0.19±0.10 (359)	0.50±0.25 (359)
2015	0.19±0.10 (400)	0.48±0.23 (400)	0.16±0.09 (80)	0.50±0.26 (80)	0.19±0.10 (300)	0.48±0.22 (300)	0.16±0.07 (20)	0.40±0.2 (0 (20))	0.17±0.10 (38)	0.46±0.26 (38)	0.19±0.10 (147)	0.48±0.22 (147)	0.19±0.10 (215)	0.48±0.23 (215)	0.19±0.10 (362)	0.48±0.22 (362)
2016	0.19±0.10 (360)	0.46±0.24 (360)	0.17±0.10 (60)	0.44±0.26 (60)	0.19±0.10 (280)	0.48±0.23 (280)	0.15±0.06 (20)	0.32±0.1 (5 (20))	0.18±0.07 (37)	0.37±0.14 (37)	0.18±0.10 (137)	0.46±0.24 (137)	0.19±0.11 (184)	0.48±0.25 (184)	0.19±0.10 (321)	0.48±0.24 (321)

Values are presented as mean±standard deviation (number of items). MEAC CKME, Medical Education Assessment Consortium clinical knowledge mock examination; D, discrimination index (by classical test theory); D (IRT), discrimination index (by item response theory); application, interpretation+problem solving.

Table 6. Discrimination index of MEAC CKME according to grading of item quality, 2011–2016

Year	A		B		C	
	D	D (IRT)	D	D (IRT)	D	D (IRT)
1st MEAC CKME						
2011	0.22±0.12 (227)	0.53±0.26 (227)	0.20±0.12 (171)	0.56±0.29 (171)	0.27±0.11 (51)	0.68±0.23 (51)
2012	0.19±0.11 (341)	0.51±0.27 (341)	0.20±0.10 (57)	0.45±0.22 (57)	0.16±0.16 (2)	0.31±0.35 (2)
2013	0.20±0.12 (241)	0.54±0.30 (241)	0.19±0.10 (138)	0.46±0.24 (138)	0.21±0.12 (17)	0.52±0.25 (17)
2014	0.21±0.11 (278)	1.03±0.57 (278)	0.18±0.09 (104)	0.86±0.49 (104)	0.11±0.08 (13)	0.72±0.51 (13)
2015	0.22±0.11 (227)	0.54±0.26 (227)	0.19±0.11 (157)	0.46±0.26 (157)	0.13±0.10 (16)	0.49±0.40 (16)
2016	0.21±0.11 (224)	0.51±0.25 (224)	0.19±0.10 (121)	0.45±0.24 (121)	0.19±0.10 (15)	0.36±0.21 (15)
2nd MEAC CKME						
2012	0.21±0.09 (337)	0.57±0.27 (337)	0.20±0.11 (59)	0.58±0.24 (59)	0.18±0.05 (4)	0.35±0.20 (4)
2013	0.19±0.10 (245)	0.49±0.23 (245)	0.20±0.10 (132)	0.51±0.27 (132)	0.15±0.08 (16)	0.39±0.22 (16)
2014	0.20±0.10 (271)	0.52±0.25 (271)	0.17±0.10 (106)	0.46±0.27 (106)	0.19±0.11 (20)	0.50±0.30 (20)
2015	0.20±0.09 (210)	0.49±0.21 (210)	0.17±0.11 (171)	0.47±0.25 (171)	0.17±0.08 (18)	0.46±0.23 (18)
2016	0.19±0.10 (211)	0.48±0.24 (211)	0.18±0.10 (136)	0.45±0.24 (136)	0.20±0.08 (11)	0.37±0.14 (11)

Values are presented as mean±standard deviation (number of items). MEAC CKME, Medical Education Assessment Consortium clinical knowledge mock examination; D, discrimination index (by classical test theory); D (IRT), discrimination index (by item response theory); application, interpretation+problem solving.

C등급보다 높았다(Figure 1B).

3) A/R형별

고전검사이론과 문항반응이론 모두 A형 문항보다는 R형 문항의 변별도가 더 높았다. 이는 의사국가시험 필기시험도 마찬가지였다(Figure 1C). 하지만 80회 의사국가시험의 A형(0.17)과 R형(0.25) 문항의 변별도 차이가 0.08인데 비해 같은 해의 1차 임상종합평가의 A형(0.20)과 R형(0.22) 문항의 변별도 차이는 0.02로 크지 않았다. 이는 같은 해 2차 임상종합평가도 유사한 경향을 보였다(Table 7).

3. 신뢰도

신뢰도는 1차 임상종합평가의 경우 2012년부터 2016년까지 0.933-0.972로 매우 높았고, 이는 의사국가시험의 신뢰도와 유사하

거나 약간 더 높았고, P지역의 연합모의고사보다 높았다. 반면에 2차 임상종합평가는 높은 편이지만, 1차보다는 상대적으로 다소 낮았다(Table 8).

4. 난이도와 변별도 구간별 비율

1) 연도별

1차 임상종합평가의 2012-2016년 난이도 0.6-0.8의 비율은 평균 28.5% (25.4%-31.5%)이고, 0.8-1.0의 비율은 평균 28.0% (21.7%-30.8%)로 유사한 반면에, 2차는 난이도 0.8-1.0에 속하는 비율이 1차에 비해 약 2배 증가하였는데, 응시자들의 역량이 높아져서 그런 것인지 아니면 문항의 질 유지 측면에서 미흡한 것인지에 대해 검토가 필요하다. 참고로 제80회 의사국가시험은 각각 24.0%와 50.8%

Table 7. Discrimination index of MEAC CKME according to grading of A/R types, 2011-2016

Year	1st MEAC CKME				2nd MEAC CKME			
	Type A ^{a)}		Type R ^{b)}		Type A ^{a)}		Type R ^{b)}	
	D	D (IRT)	D	D (IRT)	D	D (IRT)	D	D (IRT)
2011	0.21 ± 0.12	0.55 ± 0.28	0.25 ± 0.12	0.60 ± 0.25				
2012	0.19 ± 0.11	0.48 ± 0.26	0.21 ± 0.11	0.59 ± 0.28	0.21 ± 0.09	0.57 ± 0.27	0.20 ± 0.11	0.58 ± 0.27
2013	0.19 ± 0.11	0.50 ± 0.28	0.21 ± 0.10	0.57 ± 0.29	0.19 ± 0.09	0.48 ± 0.24	0.20 ± 0.11	0.52 ± 0.23
2014	0.20 ± 0.11	0.95 ± 0.52	0.22 ± 0.11	1.12 ± 0.77	0.19 ± 0.10	0.51 ± 0.27	0.20 ± 0.09	0.47 ± 0.20
2015	0.20 ± 0.11	0.50 ± 0.27	0.23 ± 0.12	0.51 ± 0.27	0.18 ± 0.10	0.47 ± 0.23	0.23 ± 0.09	0.53 ± 0.20
2016	0.20 ± 0.11	0.47 ± 0.25	0.22 ± 0.12	0.58 ± 0.24	0.18 ± 0.10	0.46 ± 0.25	0.21 ± 0.09	0.47 ± 0.19

Values are presented as mean ± standard deviation.

MEAC CKME, Medical Education Assessment Consortium clinical knowledge mock examination; D, discrimination index (by classical test theory); D (IRT), discrimination index (by item response theory); application, interpretation + problem solving.

^{a)}Single best answer. ^{b)}Extended matching.

Table 8. Reliability (Cronbach α) of MEAC CKME, 2011-2016, compared to a local consortium exam and Korean Medical Licensing Examination

Year (total no. of items)	1st MEAC CKME	2nd MEAC CKME	P local consortium exam	Korean Medical Licensing Examination
2011 (450)	0.948	-	-	Null (76th)
2012 (400)	0.937	0.940	0.937	Null (77th)
2013 (400)	0.937	0.935	0.927	0.940 (78th)
2014 (400)	0.971	0.935	0.935	0.940 (79th)
2015 (400)	0.943	0.930	0.932	0.942 (80th)
2016 (360)	0.933	0.922	0.926	Null (81th)

MEAC CKME, Medical Education Assessment Consortium clinical knowledge mock examination.

Table 9. Subgroup rates of difficulty index and discrimination index of MEAC CKME, 2011-2016

Year	P ^{a)}			D ^{a)}	
	0-0.6	0.6-0.8	0.8-1.0	<0.3	≥0.3
1st MEAC CKME					
2011	169 (37.6)	146 (32.4)	135 (30.0)	326 (72.4)	124 (27.6)
2012	175 (43.8)	102 (25.4)	123 (30.8)	325 (81.3)	75 (19.2)
2013	195 (48.8)	118 (29.5)	87 (21.7)	325 (81.3)	75 (19.2)
2014	164 (41.3)	118 (29.7)	115 (29.0)	317 (79.8)	80 (20.2)
2015	155 (38.8)	126 (31.5)	119 (29.7)	306 (76.5)	94 (23.5)
2016	150 (41.6)	106 (26.5)	104 (28.9)	285 (79.2)	75 (20.8)

(Continued to the next page)

Table 9. Continued

Year	P ^{a)}			D ^{a)}	
	0-0.6	0.6-0.8	0.8-1.0	<0.3	≥0.3
2nd MEAC CKME					
2012	130 (32.5)	82 (20.5)	188 (47.0)	345 (86.2)	55 (13.8)
2013	129 (32.5)	106 (26.7)	162 (40.8)	342 (86.1)	55 (13.9)
2014	112 (28.0)	100 (25.0)	188 (47.0)	339 (84.7)	61 (15.3)
2015	109 (27.3)	108 (27.0)	183 (45.7)	344 (86.0)	56 (14.0)
2016	92 (25.7)	97 (27.1)	169 (47.2)	306 (85.5)	52 (14.5)
Korean Medical Licensing Examination					
80th (400)	101 (25.2)	96 (24.0)	203 (50.8)	319 (79.8)	81 (20.2)

Values are presented as number of items (%).

MEAC CKME, Medical Education Assessment Consortium clinical knowledge mock examination; P, difficulty index (by classical test theory); D, discrimination index (by classical test theory).

^{a)}Reasonable item: P (0.25-0.75) and D (≥0.25).

Table 10. Subgroup rates of difficulty index and discrimination index of MEAC CKME according to item types, 2011-2016

Year	Recall					Interpretation					Problem solving					Application				
	P		D			P		D			P		D			P		D		
	0-0.6	0.6-0.8	0.8-1.0	<0.3	≥0.3	0-0.6	0.6-0.8	0.8-1.0	<0.3	≥0.3	0-0.6	0.6-0.8	0.8-1.0	<0.3	≥0.3	0-0.6	0.6-0.8	0.8-1.0	<0.3	≥0.3
1st MEAC CKME																				
2011	84/218 (38.5)	64/218 (29.4)	70/218 (32.1)	149/218 (68.3)	69/218 (31.7)	38/111 (34.2)	41/111 (36.9)	32/111 (28.9)	85/111 (76.6)	26/111 (23.4)	47/121 (38.8)	41/121 (33.9)	33/121 (27.3)	92/121 (76.0)	29/121 (24.0)	85/232 (36.7)	82/232 (35.3)	65/232 (28.0)	177/232 (76.3)	55/232 (23.7)
2012	50/102 (49.0)	20/102 (19.6)	32/102 (31.4)	89/102 (87.3)	13/102 (12.7)	33/113 (29.2)	33/113 (29.2)	47/113 (41.6)	94/113 (83.2)	19/113 (16.8)	87/178 (48.9)	49/178 (27.5)	42/178 (23.6)	135/178 (75.8)	43/178 (24.2)	125/298 (42.0)	82/298 (27.5)	91/298 (30.5)	236/298 (79.2)	62/298 (20.8)
2013	36/84 (42.8)	34/84 (40.5)	14/84 (16.7)	70/84 (83.3)	14/84 (16.7)	57/134 (42.6)	42/134 (31.3)	35/134 (26.1)	113/134 (84.3)	21/134 (15.7)	99/179 (55.3)	42/179 (23.5)	38/179 (21.2)	139/179 (77.7)	40/179 (22.3)	159/316 (50.3)	84/316 (26.6)	73/316 (23.1)	255/316 (80.7)	61/316 (19.3)
2014	13/30 (43.3)	9/30 (30.0)	8/30 (26.7)	25/30 (83.3)	5/30 (16.7)	45/135 (31.3)	43/135 (31.9)	47/135 (34.8)	108/135 (80.0)	27/135 (20.0)	106/232 (45.7)	66/232 (28.4)	60/232 (25.9)	184/232 (79.3)	48/232 (20.7)	151/367 (41.1)	109/367 (29.7)	107/367 (29.2)	292/367 (79.6)	75/367 (20.4)
2015	27/63 (42.9)	23/63 (36.5)	13/63 (20.6)	54/63 (85.7)	9/63 (14.3)	46/141 (32.6)	46/141 (32.6)	49/141 (34.8)	101/141 (71.6)	40/141 (28.4)	82/196 (41.8)	57/196 (29.1)	57/196 (29.1)	151/196 (77.0)	45/196 (23.0)	128/337 (38.0)	103/337 (30.6)	106/337 (31.4)	252/337 (74.8)	85/337 (25.2)
2016	17/43 (39.5)	18/43 (41.9)	8/43 (18.6)	34/43 (79.1)	9/43 (20.9)	59/134 (44.0)	37/134 (27.6)	38/134 (28.4)	104/134 (77.6)	30/134 (22.4)	74/183 (40.4)	51/183 (27.9)	58/183 (31.7)	147/183 (80.3)	36/183 (19.7)	133/317 (42.0)	88/317 (27.7)	96/317 (30.3)	251/317 (79.2)	66/317 (20.8)
2nd MEAC CKME																				
2012	38/106 (35.8)	18/106 (17.0)	50/106 (47.2)	94/106 (88.7)	12/106 (11.3)	36/127 (28.3)	26/127 (20.5)	65/127 (51.2)	110/127 (86.6)	17/127 (13.4)	53/161 (32.9)	37/161 (23.0)	71/161 (44.1)	135/161 (83.9)	26/161 (16.1)	92/294 (31.3)	64/294 (21.8)	138/294 (46.9)	251/294 (85.4)	43/294 (14.6)
2013	29/82 (35.3)	18/82 (22.0)	35/82 (42.7)	73/82 (89.0)	9/82 (11.0)	34/118 (28.8)	28/118 (23.7)	56/118 (47.5)	98/118 (83.1)	20/118 (16.9)	66/194 (34.0)	58/194 (29.9)	70/194 (36.1)	169/194 (87.1)	25/194 (12.9)	100/315 (31.8)	88/315 (27.9)	127/315 (40.3)	269/315 (85.4)	46/315 (14.6)
2014	15/41 (36.6)	8/41 (19.5)	18/41 (43.9)	38/41 (92.7)	3/41 (7.3)	32/150 (21.3)	41/150 (27.3)	77/150 (51.4)	129/150 (86.0)	21/150 (14.0)	65/209 (31.1)	51/209 (24.4)	93/209 (44.5)	172/209 (82.3)	37/209 (17.7)	97/359 (27.0)	92/359 (25.6)	170/359 (47.4)	301/359 (83.8)	58/359 (16.2)
2015	15/38 (39.5)	13/38 (34.2)	10/38 (26.3)	34/38 (89.5)	4/38 (10.5)	40/147 (27.2)	33/147 (22.5)	74/147 (50.3)	124/147 (84.4)	23/147 (15.6)	54/215 (25.1)	62/215 (28.8)	99/215 (46.1)	186/215 (86.5)	29/215 (13.5)	94/362 (26.0)	95/362 (26.2)	173/362 (47.8)	310/362 (85.6)	52/362 (14.4)
2016	10/37 (27.0)	12/37 (32.4)	15/37 (40.6)	35/37 (94.6)	2/37 (5.4)	34/137 (24.8)	32/137 (23.4)	71/137 (51.8)	120/137 (87.6)	17/137 (12.4)	48/184 (26.1)	53/184 (28.8)	83/184 (45.1)	151/184 (82.1)	33/184 (17.9)	82/321 (25.5)	85/321 (26.5)	154/321 (48.0)	271/321 (84.4)	50/321 (15.6)

Values are presented as number of each items/total items (%).

MEAC CKME, Medical Education Assessment Consortium clinical knowledge mock examination; P, difficulty index (by classical test theory); D, discrimination index (by classical test theory); application, interpretation+problem solving.

로 쉬운 문항의 비율이 매우 높았다. 변별도 0.3 이상의 비율은 약 20%로 의사국가시험과 유사하였다(Table 9). 적정문항의 정의를 고전검사이론 난이도(P) 25-75이면서 고전검사이론 변별도(D) 0.25 이상일 때로 임의로 정했을 때 1차 임상종합평가에서 적정문항의 비율은 약 25%였으며 2012년부터 그 비율이 조금씩 증가하는 추세를 보였다(Figure 1E). 반면에 2차 임상종합평가는 적정문항의 비율이 1차보다 낮았다(Figure 1F).

2) 문항유형별

2012-2016년 1차 임상종합평가에 비해 2차 임상종합평가에서 암기형, 해석형 및 문제해결형 모두 난이도 0.8-1.0의 쉬운 문항의 비율이 더 높았고, 해석형과 문제해결형을 합친 응용형의 경우 난이도 0.8-1.0의 쉬운 문항의 비율 또한 1차 임상종합평가보다 2차 임상종합평가에서 약 20% 더 많았다. 변별도 0.3 이상의 비율은 암기형보다 해설형이나 문제해결형이 더 높은 경향을 보였다(Table 10).

3) 등급별

문항개발과정에서 자가점검표를 활용한 문항등급에 있어서 A등급의 비율은 서서히 증가하여 2016년 1차 임상종합평가에서 62.2%로 최고로 높았고, 반면에 C등급의 문항은 4.1%로 가장 낮았다.

1차, 2차 임상종합평가 모두 등급별로 난이도 분포는 고르게 나타났으나, 변별도 0.3 이상인 비율은 A문항이 B, C문항보다 의미 있게 높았다. 가장 최근 2016년 1차 임상종합평가의 경우 변별도 0.3 이상인 비율은 A문항 중에서는 25.4%인 반면에 B, C문항 중에서는

Table 11. Subgroup rates of difficulty index and discrimination index of MEAC CKME according to grading of item quality, 2011–2016

Year	A					B					C				
	P			D		P			D		P			D	
	0-0.6	0.6-0.8	0.8-1.0	<0.3	≥0.3	0-0.6	0.6-0.8	0.8-1.0	<0.3	≥0.3	0-0.6	0.6-0.8	0.8-1.0	<0.3	≥0.3
1st MEAC CKME															
2011	85/227 (37.4)	78/227 (34.4)	64/227 (28.2)	167/227 (73.6)	60/227 (26.4)	69/171 (40.4)	52/171 (30.4)	50/171 (29.2)	132/171 (77.2)	39/171 (22.8)	15/51 (29.4)	16/51 (31.4)	20/51 (39.2)	27/51 (52.9)	24/51 (47.1)
2012	152/341 (44.6)	83/341 (24.3)	106/341 (31.1)	275/341 (80.6)	66/341 (19.4)	22/57 (38.6)	18/57 (31.6)	17/57 (29.8)	48/57 (84.2)	9/57 (15.8)	1/2 (50.0)	1/2 (50.0)	0/2 (0)	2/2 (100)	0/2 (0)
2013	124/241 (51.4)	72/241 (29.9)	45/241 (18.7)	183/241 (75.9)	58/241 (24.1)	62/138 (44.9)	42/138 (30.4)	34/138 (24.7)	125/138 (90.6)	13/138 (9.4)	6/17 (35.3)	3/17 (17.6)	8/17 (47.1)	13/17 (76.5)	4/17 (23.5)
2014	113/278 (40.6)	80/278 (28.8)	85/278 (30.6)	211/278 (75.9)	67/278 (24.1)	44/104 (42.3)	35/104 (33.7)	25/104 (24.0)	91/104 (87.5)	13/104 (12.5)	6/13 (46.1)	2/13 (15.4)	5/13 (38.5)	13/13 (100)	0/13 (0)
2015	83/227 (36.6)	75/227 (33.0)	69/227 (30.4)	165/227 (72.7)	62/227 (27.3)	68/157 (43.3)	46/157 (29.3)	43/157 (27.4)	126/157 (80.3)	31/157 (19.7)	4/16 (25.0)	5/16 (31.3)	7/16 (43.7)	15/16 (93.7)	1/16 (6.3)
2016	95/224 (42.4)	58/224 (25.9)	71/224 (31.7)	167/224 (74.6)	57/224 (25.4)	46/121 (38.0)	44/121 (36.4)	31/121 (25.6)	105/121 (86.8)	16/121 (13.2)	9/15 (60.0)	4/15 (26.7)	2/15 (13.3)	13/15 (86.7)	2/15 (13.3)
2nd MEAC CKME															
2012	107/337 (31.7)	71/337 (21.1)	159/337 (47.2)	292/337 (86.6)	45/337 (13.4)	21/59 (35.6)	11/59 (18.6)	27/59 (45.8)	49/59 (83.1)	10/59 (16.9)	2/4 (50.0)	0/4 (0)	2/4 (50.0)	4/4 (100)	0/4 (0)
2013	82/245 (33.5)	59/245 (24.1)	104/245 (42.4)	212/245 (86.5)	33/245 (13.5)	38/132 (28.8)	40/132 (30.3)	54/132 (40.9)	110/132 (83.3)	22/132 (16.7)	7/16 (43.7)	6/16 (37.5)	3/16 (18.8)	16/16 (100)	0/16 (0)
2014	73/271 (26.9)	68/271 (25.1)	130/271 (48.0)	221/271 (81.5)	50/271 (18.5)	29/106 (27.4)	29/106 (27.4)	48/106 (45.2)	97/106 (91.5)	9/106 (8.5)	9/20 (45.0)	3/20 (15.0)	8/20 (40.0)	18/20 (90.0)	2/20 (10.0)
2015	43/210 (20.5)	57/210 (27.1)	110/210 (52.4)	176/210 (83.8)	34/210 (16.2)	59/171 (34.5)	46/171 (26.9)	66/171 (38.6)	151/171 (88.3)	20/171 (11.7)	7/18 (38.9)	5/18 (27.8)	6/18 (33.3)	16/18 (88.9)	2/18 (11.1)
2016	48/211 (22.7)	55/211 (26.1)	108/211 (51.2)	173/211 (82.0)	38/211 (18.0)	40/136 (29.4)	38/136 (27.9)	5/136 (42.7)	123/136 (90.4)	13/136 (9.6)	4/11 (36.4)	4/11 (36.4)	3/11 (27.2)	10/11 (90.9)	1/11 (9.1)

Values are presented as number of each items/total items (%).

MEAC CKME, Medical Education Assessment Consortium clinical knowledge mock examination; P, difficulty index (by classical test theory); D, discrimination index (by classical test theory).

Table 12. Subgroup rates of difficulty index and discrimination index of MEAC CKME according to A/R types, 2011–2016

Year	Type A ^{a)}					Type R ^{b)}				
	P			D		P			D	
	0-0.6	0.6-0.8	0.8-1.0	<0.3	≥0.3	0-0.6	0.6-0.8	0.8-1.0	<0.3	≥0.3
1st MEAC CKME										
2011	143/391 (36.6)	131/391 (33.5)	117/391 (29.9)	296/391 (75.7)	95/391 (24.3)	26/59 (44.1)	15/59 (25.4)	18/59 (30.5)	30/59 (50.8)	29/59 (49.2)
2012	146/348 (42.0)	92/348 (26.4)	110/348 (31.6)	285/348 (81.9)	63/348 (18.1)	29/52 (55.8)	10/52 (19.2)	13/52 (25.0)	40/52 (76.9)	12/52 (23.1)
2013	160/344 (46.5)	109/344 (31.7)	75/344 (21.8)	285/344 (82.8)	59/344 (17.2)	35/56 (62.5)	9/56 (16.1)	12/56 (21.4)	40/56 (71.4)	16/56 (28.6)
2014	141/348 (40.5)	105/348 (30.2)	102/348 (29.3)	280/348 (80.5)	6/348 (19.5)	23/49 (47.0)	13/49 (26.5)	13/49 (26.5)	37/49 (75.5)	12/49 (24.5)
2015	128/340 (37.6)	104/340 (30.6)	108/340 (31.8)	263/340 (77.4)	77/340 (22.6)	27/60 (45.0)	22/60 (36.7)	11/60 (18.3)	43/60 (71.7)	17/60 (28.3)
2016	112/302 (37.1)	100/302 (33.1)	90/302 (29.8)	244/302 (80.8)	58/302 (19.2)	38/58 (65.5)	6/58 (10.3)	14/58 (24.2)	41/58 (70.7)	17/58 (29.3)
2nd MEAC CKME										
2012	103/348 (29.6)	71/348 (20.4)	174/348 (50.0)	301/348 (86.5)	47/348 (13.5)	27/52 (51.9)	11/52 (21.2)	14/52 (26.9)	44/52 (84.6)	8/52 (15.4)
2013	109/341 (32.0)	92/341 (27.0)	140/341 (41.0)	296/341 (86.8)	45/341 (13.2)	20/56 (35.7)	14/56 (25.0)	22/56 (39.3)	46/56 (82.1)	10/56 (17.9)
2014	98/344 (28.5)	84/344 (24.4)	162/344 (47.1)	289/344 (84.0)	55/344 (16.0)	14/56 (25.0)	16/56 (28.6)	26/56 (46.4)	50/56 (89.3)	6/56 (10.7)
2015	88/340 (25.9)	94/340 (27.6)	158/340 (46.5)	301/340 (88.5)	39/340 (11.5)	21/60 (35.0)	14/60 (23.3)	25/60 (41.7)	43/60 (71.7)	17/60 (28.3)
2016	73/301 (24.3)	88/301 (29.2)	140/301 (46.5)	263/301 (87.4)	38/301 (12.6)	19/57 (33.3)	9/57 (15.8)	29/57 (50.9)	43/57 (75.4)	14/57 (24.6)

Values are presented as number of each items/total items (%).

MEAC CKME, Medical Education Assessment Consortium clinical knowledge mock examination; P, difficulty index (by classical test theory); D, discrimination index (by classical test theory).

^{a)}Single best answer. ^{b)}Extended matching.

각각 13.2%와 13.3%였고, 2016년 2차 임상종합평가의 경우에도 변별도 0.3 이상인 비율은 A문항 중에서는 18.0%인 반면에 B, C문항 중에서는 각각 9.6%와 9.1%였다(Table 11).

4) A/R형별

난이도 0-0.6의 어려운 문항의 비율이 A형 문항에서보다는 R형 문항에서 매우 높았고, 0.3 이상의 변별도를 가진 비율이 A형 문항에서보다는 R형 문항에서 많게는 약 10%까지 높았다. 이는 1차 임상종합평가가 2차보다 더 뚜렷하게 나타났다(Table 12).

고찰

의학교육평가컨소시엄 임상종합평가는 연 2회 시행되는데, 의학교육평가컨소시엄에서는 출제계획표에 따른 1회 시험문항의 3배수를 각 참여 의과대학별로 분배하여 사전에 출제하도록 하고 있다. 이를 위해 의학교육평가컨소시엄에서는 각 의과대학마다 그 해의 출제과 출제할 교수를 대상으로 좋은 문항을 출제하도록 출제워크숍을 시행하고 있다. 출제된 문항은 수정되고 수정된 문항은 출제계획표에 따라 1차, 2차 임상종합평가별로 분배된다. 2014년부터는 문항개발관리소위원회 내 선정된 전문검토위원이 문항정리워크숍을 시행하기 이전에 사전검토하는 과정을 시행하고 있다. 2015년부터는 편집워크숍을 추가적으로 시행하여 전문검토위원이 최종 인쇄 전에 출제된 문항을 최종적으로 검토하고 편집과정의 오류를 줄이고자 노력하고 있다. 그 결과 이미지 누락, 표기오류 등의 오류는 크게 줄었다(Figure 1D).

임상종합평가의 난이도는 컨소시엄 시행 당시 의사국가시험보다 난이도지수를 낮게 하여 학생들이 부족한 것을 인지하고 더 열심히 학업에 임하도록 하기 위하여는 의도대로 본 연구의 분석결과 연도별 차이가 없이 평균 점수 약 60점이 되도록 잘 시행이 되었다. 반면에, 의사국가시험 필기시험의 전체 난이도지수는 76회부터 80회까지 0.74-0.77로 일정한 편이다. 적정 난이도 0.6-0.8 구간의 비율은 77회, 78회, 79회, 80회 의사국가시험 필기시험의 경우 최근 들어 증가하는 추세이기는 하지만[7-10], 각각 19.8%, 21.4%, 23.5%, 24.0%에 불과한 반면, 임상종합평가는 약 30%로 그보다 높다. 또 의사국가시험에 비해 임상종합평가의 난이도가 의학총론, 의학각론 및 보건의약관계법규 과목별로 차이가 없었다. 그 예로 2016년 1차 임상종합평가의 경우 의학총론, 의학각론 및 보건의약관계법규의 난이도지수가 각각 0.63, 0.61 및 0.65로 비슷한 반면에 80회 의사국가시험의 난이도지수는 각각 0.70, 0.74, 0.88로 과목마다 다르게 나타났다. 의사국가시험과 달리 의학교육평가컨소시엄에서는 자가점검표를 이용하여 문항의 질 등급을 매기는데, 이유는 알 수 없지만, 예상과 달리 문항의 질 등급별로는 난이도에 있어

차이가 없었다. 문항의 형태에 있어서는 의사국가시험과 마찬가지로 A형 문항보다는 R형 문항의 난이도지수가 낮았다. 전체 R형 문항 중에서 난이도 0-0.6으로 어려운 문항에 속하는 비율이 A형 문항 중에서 보다 높았다. 문항유형에 있어서는 1차 임상종합평가가 2차보다 문제해결형 또는 응용형의 난이도지수가 낮게 나타나 의사국가시험과 같은 경향을 보였고, 1차 임상종합평가에 비해 2차가 난이도 0.8-1.0에 속하는 문항의 비율이 약 2배 증가하였는데, 이는 응시자들의 문제해결능력 등의 역량이 향상되었기 때문에 그렇다고 볼 수도 있겠으나, 2차 임상종합평가의 문항의 질 유지 혹은 난이도 조정이 미흡했을 가능성도 있어 추후 이 부분에 대한 개선노력이 필요하다고 여겨진다. 그동안 1차, 2차 임상종합평가 모두 암기형 비율은 크게 줄었다. 특히 2016년 1차 임상종합평가의 경우 11.9%로 2011년의 48.4%보다는 1/4로 크게 줄었으나, 여전히 높은 보건약물관계법규 문항 탓에 의사국가시험보다는 아직 높아 앞으로 암기형 문항을 더욱 줄이기 위한 노력이 필요하다. 하지만 매년 일정한 난이도를 유지하고 있고, 전체뿐 아니라 과목별로 난이도의 차이가 없고, 의사국가시험에 비해 난이도지수 0.6-0.8 구간에 속하는 문항의 비율이 높다는 점이 의학교육평가컨소시엄 임상종합평가 시험의 강점이라고 하겠다[7-10].

의사국가시험 필기시험의 전체 변별도지수는 76회부터 80회까지 0.17-0.18로 일정하게 유지되고 있는데, 특히 1차 임상종합평가의 변별도는 약 0.20으로 이보다 높은 편이었다. 특히 의사국가시험의 변별도 0.3 이상의 비율은 18.8%, 15.0%, 30.8%, 20.0%로 일정하지 않았던 반면에, 1차 임상종합평가의 변별도 0.3 이상의 비율은 약 20%로 일정한 편이다. 과목별로는 의사국가시험과 임상종합평가 모두 의학각론의 변별도가 가장 높았다. 그 예로 2016년 1차 임상종합평가의 전체 변별도는 0.20이었고, 과목별로는 의학총론이 0.15, 의학각론이 0.21, 보건의약관계법규가 0.18이었던 반면에, 같은 해에 시행되었던 80회 의사국가시험의 전체 변별도는 0.18이고, 과목별로는 의학총론이 0.16, 의학각론이 0.19, 보건의약관계법규는 0.10으로 의사국가시험의 과목 간 변별도 차이가 0.09로서 임상종합평가의 0.6보다 컸다. 문제유형에 있어서는 2016년 1차 임상종합평가의 경우 문제유형별로는 차이가 없이 모두 0.20으로 동일한 반면에, 80회 의사국가시험은 문제해결형(0.19), 해석형(0.17), 암기형(0.13) 순서로 차이가 있었다. 임상종합평가의 경우 문항반응이론에 의한 변별도는 약 0.5로 적정 수준이었고, 과목별, 유형별 특징은 고전검사이론에 의한 변별도와 마찬가지로였다. 하지만 고전검사이론의 변별도 0.3 이상의 비율은 암기형 문항보다 해결형 문항이나 문제해결형 문항에서 다소 더 높았고, 2차 임상종합평가는 1차와 달리 문제유형별로 변별도의 차이를 보였다. 문항의 형태에 있어서 2014년 이후가 그 이전보다 A형과 R형 문항 모두 변별도가 더 높아졌다. A형과 R형 문항의 변별도를 비교하면, 1, 2차 임상종합평가 모두 의사국가시험 필기시험과 마찬가지로 고전검사이론과 문항

반응이론 모두 A형 문항보다는 R형 문항의 변별도가 더 높았고, 0.3 이상의 변별도를 가진 비율을 보더라도 A형 문항보다는 R형 문항에서 많게는 약 10%까지 훨씬 높았다. 하지만 A형과 R형 문항의 변별도 차이가 의사국가시험에 비해 임상종합평가가 작았는데, 이는 임상종합평가의 A형 문항의 변별도가 의사국가시험보다 상대적으로 높고, R형의 변별도가 상대적으로 낮기 때문이었다[12]. 2013년 1차 임상종합평가부터는 문항의 질 등급을 나누기 위한 자가점검표를 활용하였는데, 이는 앞서 등급 간에 차이가 없었던 난이도와 달리 변별력을 높이기 위한 전략으로는 유용한 것으로 입증되었다. 고전검사이론과 문항반응이론 모두 A등급의 변별도가 B등급이나 C등급보다 높아 문항의 질이 높을수록 변별력이 높은 문항인 것을 알 수 있다.

임의로 적정문항의 정의를 고전검사이론 난이도(P) 25-75이면서 고전검사이론 변별도(D) 0.25 이상일 때로 정했을 때, 1차 임상종합평가의 적정문항의 비율이 2012년부터 점점 증가하여 2016년 약 25%였으나(Figure 1E), 2차 임상종합평가의 적정문항의 비율이 1차보다 낮았다(Figure 1F). 이는 최초 3배수 출제된 문항 중에서 좋은 문항을 1차 임상종합평가에 우선적으로 배치를 했기 때문일 수 있지만, 향후 전체적으로 보다 좋은 문항을 많이 개발하여 2차 임상종합평가의 경우에도 변별력이 높은 보다 좋은 문항이 출제되도록 노력을 기울여야 할 것이다.

신뢰도는 0.9가 최대이고 그보다 낮으면 신뢰도를 높이기 위한 노력을 기울여야 하고, 그 이상이면 문항의 수가 충분히 많기 때문에 문항의 수를 보다 줄일 것을 추천하고는 있지만, 신뢰도가 0.9 이상이면 국가시험 혹은 전국적 시험으로는 매우 신뢰할 만한 시험으로 간주한다[13]. 본 연구에서 1차 임상종합평가의 신뢰도는 0.933-0.972로 매우 높았고, 이는 같은 해 실시되었던 의사국가시험이나 P지역의 연합모의고사와 유사하거나 약간 더 높았다. 반면에 2차 임상종합평가 역시 신뢰도는 0.9 이상으로 높은 편이지만, 1차에 비해서는 낮아 더 좋은 문항을 개발할 필요가 있었다.

연구의 제한점으로는 아쉽게도 문항의 질 평가에 있어 중요한 요소인 답가지의 매력도까지는 제시하지 못한 것이다. 마찬가지로 자가점검표 9번 항목(틀린 4가지 답가지가 모두 매력적이면 1점, 아니면 0점)과 답가지 매력도와 비교 또한 시행하지 못했다. 대신에 문항추측도를 제거한 난이도, 즉 교정난이도를 제시하였다. 또 문항반응이론으로 문항을 분석할 때 난이도와 변별도만 제시하였고, 아쉽게도 문항정보함수를 이용한 문항들의 측정오차에 대한 정보와 검사정보함수를 통한 시험의 성격까지는 제시하지 못하였다. 이는 후속연구를 통해 제시하도록 하겠다.

요약하면, 결론적으로 고전검사이론과 문항반응이론에 의한 다각도의 분석결과 의학교육평가컨소시엄에서 수행한 임상종합평가는 전국 의과대학 의학과 4학년 학생들이 자신의 역량을 점검하고 의사국가시험을 준비하기 위한 평가로 도움이 된다고 여겨진다. 비

록 여기서 자료를 제시하지는 않았지만, 2017년 3월에 시행한 설문조사에서 임상종합평가에 참여하는 38개 대학 중에 31개 대학이 임상종합평가시험의 질에 대해 90.3%가 우수하다고 응답한 바 있다. 2013년부터 활용하기 시작한 자가점검표는 측정의 양호도 측면에서 우수한 좋은 문항을 개발하는 데 매우 유용한 것으로 나타났다. 그 결과 갈수록 A등급 문항의 비율이 가장 높아져서 고무적이다. 출제할 교수에게 사전에 출제워크숍을 통해 문항개발지침을 교육하고, 출제가 되면 전문검토위원이 수정이 필요한 사항이 있는지를 사전에 검토하고, 출제된 교수들이 문항정리워크숍을 통해 전문영역별로 다같이 모여 함께 수정하고, 선정하여 교시별로 배분하고, 이후 다시 전문검토위원이 편집워크숍을 통해 임상종합평가 시험지를 인쇄하기 전에 선정이 완료된 문항을 최종적으로 검토하고, 시험지 구성과 편집을 살펴보는 등 일련의 노력들을 통해 난이도를 일정하게 유지하고, 갈수록 암기형 문항을 줄여가고, 보다 변별력이 좋은 문항의 비율을 높임으로써 결국은 적정문항의 비율을 높이고 높은 신뢰도를 확보하게 되었다. 82회 의사국가시험 필기시험부터는 의사국가시험 평가목표가 105개 의사직무상황으로 바뀌어 출제되는 첫 해이다. 이제 의학교육평가컨소시엄은 2017년부터 한국의과대학·의학전문대학원협회 산하 의학교육평가사업단으로 전환되어 41개 의과대학이 모두 참여하게 되었다. 이를 통해 전국의 의과대학 및 의학전문대학원 학생들이 의사국가시험 필기시험에 충분히 대비할 수 있도록 이러한 노력들은 계속되어야 할 것이다.

감사의 글

이 연구는 부산대학교병원 의학연구소 학술연구비의 지원을 받아 연구되었다(2014-19). 또한 분석과정에 헌신적으로 도움을 준 부산대학교 의과대학 의학교육지원실의 이선희, 최수연 선생님에게 깊은 감사를 드린다.

저자 기여

이상엽: 연구설계, 통계, 초고작성, 연구전체 총괄; 이에리: 연구 실행, 결과분석; 김미경: 연구설계-실행-분석 총괄

REFERENCES

1. Baik SH. The new horizon for evaluations in medical education in Korea. *J Educ Eval Health Prof.* 2005;2(1):7-22.
2. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med.* 1990;65(9 Suppl):S63-7.
3. Anderson LW, Krathwohl DR, Airasian PW, Cruikshank KA, Mayer RE, Pintrich PR, et al. A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives. Boston (MA): Pearson/Allyn and Bacon; 2001.

4. Baik SH. Major reforms and issues of the medical licensing examination systems in Korea. *Korean Med Educ Rev.* 2013;15(3):125-31.
5. Case SM, Swanson DB. *Constructing written test questions for the basic and clinical sciences.* 3rd ed. Philadelphia (PA): National Board of Medical Examiners; 2002.
6. Reeder FF, Joos AE. *Fire service instructor: principles and practice.* 2nd ed. Burlington (MA): Jones & Bartlett Learning; 2013.
7. Korea Health Personnel Licensing Examination Institute. 77th Medical licensing examination item analysis report, 2013. Seoul: Korea Health Personnel Licensing Examination Institute; 2015.
8. Korea Health Personnel Licensing Examination Institute. 78th Medical licensing examination item analysis report, 2014. Seoul: Korea Health Personnel Licensing Examination Institute; 2015.
9. Korea Health Personnel Licensing Examination Institute. 79th Medical licensing examination item analysis report, 2015. Seoul: Korea Health Personnel Licensing Examination Institute; 2015.
10. Korea Health Personnel Licensing Examination Institute. 80th Medical licensing examination item analysis report, 2016. Seoul: Korea Health Personnel Licensing Examination Institute; 2016.
11. Lim HS, Lee YM, Ahn DS, Lee JY, Im H. Item analysis of clinical performance examination using item response theory and classical test theory. *Korean J Med Educ.* 2007;19(3):185-95.
12. Levine HG, McGuire CH, Nattress Jr LW. The validity of multiple choice achievement tests as measures of competence in medicine. *Am Educ Res J.* 1970;7(1):69-82.
13. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ.* 2011;2:53-5.