2002, 2003년도 의사국가시험에서 확장결합형 문항의 양호도 평가

한국보건의료인국가시험원¹, 한림대학교 의과대학 기생충학교실² 및 의학교육연구소³

임미경^{1,3} · 허 선^{2,3}

= Abstract =

Usability of Extended-matching Type Items in the Korean Medical Licensing Examinations (2002, 2003)

Mi-Kyoung Yim^{1,3} and Sun Huh^{2,3}, MD, PhD

National Health Personnel Licensing Examination Board¹, Seoul, Korea, Department of Parasitology², College of Medicine and Institute of Medical Education³, Hallym University, Chuncheon Korea

Purpose: In 2002, extended-matching type (R-type) items were introduced to the Korean Medical Licensing Examination. To evaluate the usability of R-type items, the results of the Korean Medical Licensing Examination in 2002 and 2003 were analyzed based on item types and knowledge levels.

Methods: Item parameters, such as difficulty and discrimination indexes, were calculated using the classical test theory. The item parameters were compared across three item types and three knowledge levels.

Results: The values of R-type item parameters were higher than those of A- or K-type items. There was no significant difference in item parameters according to knowledge level, including recall, interpretation, and problem solving. The reliability of R-type items exceeded 0.99. With the R-type, an increasing number in correct answers was associated with a decreasing difficulty index.

Conclusion: The introduction of R-type items is favorable from the perspective of item parameters. However, an increase in the number of correct answers in pick 'n' - type questions results in the items being more difficult to solve.

Key Words: Medical education, Classical test theory, Difficulty index, Discrimination index, Extended-matching type item

교신저자: 허 선, 한림대학교 의과대학 기생충학교실 및 의학교육연구소
강원도 춘천시 옥천동 1

Tel: 033)248-2652, Fax: 033)251-8354, E-mail: shuh@hallym.ac.kr

* 이 논문의 내용은 한국보건의료인국가시험원의 공식 입장이 아님.

** 이 연구는 2004년도 한국의학교육학회 연구장려상으로 이루어졌음.

Table I. Composition of the 2002 and 2003 Korean Medical Licensing Examinations Based on Subjects and Item Types No.(%)

Subjects	Type*	2002				2003			
		A	K	R	Total	A	K	R	Total
		No. items	No. items	No. items	No. items	No. items	No. items	No. items	No. items
General Part		47 (10.44)	49 (10.89)	4 (0.89)	100 (22.22)	62 (12.40)	41 (8.20)	7 (1.40)	110 (22.00)
Special Part		165 (36.67)	147 (32.67)	18 (4.00)	330 (73.33)	209 (41.80)	143 (28.60)	18 (3.60)	370 (74.00)
Medical Law		11 (2.44)	9 (2.00)	0 (0.00)	20 (4.44)	13 (2.60)	7 (1.40)	0 (0.00)	20 (4.00)
Total		223 (49.6)	205 (45.6)	22 (4.9)	450 (100.00)	284 (56.80)	191 (38.20)	25 (5.00)	500 (100.00)

* type: A-one best answer type, K-multiple true-false type, R-extended matching type

서 론

2002년부터 의사국가시험은 확장결합형 (extended matching type, R 형) 문항을 도입하고 있으며, 2003년에는 복수정답형 확장결합형 문항 (Pick 'n'-type item) 도 도입하여 10개 이상의 답가지에서 하나가 아닌 2개 이상의 정답을 선택하는 문항을 출제하기 시작하였다. 이러한 문항은 전체 문항 500 개 중 약 5% 전후에서 출제하지만, 이런 형태의 문항은 앞으로 늘어날 것이다. 이러한 확장결합형 문항 또는 복수정답형 확장결합형 문항이 과연 시험 문항으로서 타당한 난이도 및 분별도의 특성을 가지고 있는지 분석된 결과를 통해 점검하고, 다른 문항 유형과의 비교를 통하여 확장결합형 문항의 특성을 평가해 보는 일이 필요하다.

그러므로, 확장결합형 문항을 다각도로 평가하기 위하여 문항의 유형 (type) 및 지식수준 (knowledge level) 에 따라 대상 문항의 난이도, 분별도 등 문항의 양호도를 검정할 수 있는 지표를 분석하려고 하였다. 이러한 결과가 이미 시행된 확장결합형 문항을 평가하는데 뿐만 아니라, 향후 양질의 문항 개발을 위해서도 유용한 자료가 될 것이다.

대상 및 방법

가. 연도별, 과목별, 지식수준별 문항 구성

2002년도 총 응시자 3,576명 전원 및 2003년도

총 응시자 3,647명 전원의 의사국가시험 응답 결과를 가지고 문항분석을 시행하였다. 현재 의사국가시험은 의학각론, 의학총론, 보건관계법규 세 과목으로 이루어져 있으며 문항들은 객관식 선다형 문항으로 구성되어 있다. 문항유형을 세부적으로 살펴보면 하나의 문항줄기에 따른 5개의 정답지 중에서 하나를 고르는 단일정답형 (one best answer type, A 형) 문항, 가, 나, 다, 라 4개의 보기 가운데 맞는 보기의 조합을 고르는 조합형 (multiple true-false type, K 형) 문항, 그리고 여러 개의 문항이 많게는 20여 개의 답가지를 공유하면서 해당 문항에 맞는 답을 고르는 확장결합형 (R 형) 문항으로 구성되어 있다 (Case and Swanson, 1998). 2002년도 의사국가시험 450문항 중 확장결합형 문항은 22개, 2003년도에는 전체 500 문항 중 25문항으로 매년 5%를 차지하고 있다. 사용된 2개년도 의사국가시험의 문항구성을 과목별, 문항유형별로 살펴보면 Table I 과 같다.

한편 의사국가시험의 문항을 지식수준별로 살펴보면 암기형, 해석형, 문제해결형 문항으로 구성되어 있다 (한국보건의료인국가시험원, 1999a). 지식의 위계는 지식 (knowledge), 이해 (comprehension), 적용 (application), 분석 (analysis), 종합 (synthesis), 평가 (evaluation)로 분류되면서 오른쪽으로 갈수록 고차원적 정신능력을 측정한다 (Bloom, 1956). 이것을 다시 사실의 기억 (recall of facts), 자료의 해석 (interpretation of data), 문제해결 (problem solving)로 분류하면서 지식수준에 따른 문항을 작성한다

Table II. Composition of Extended-matching Type Items Based on the Knowledge Level

Type Subject	2002				2003			
	Recall	Interpretation	Problem -solving	Total	Recall	Interpretation	Problem -solving	Total
	No. items(%)	No. items(%)	No. items(%)	No. items(%)	No. items(%)	No. items(%)	No. items(%)	No. items(%)
General Part	-	2 (9.09)	2 (9.09)	4 (18.18)	-	4 (16.00)	3 (12.00)	7 (28.00)
Special Part	-	14 (63.64)	4 (18.18)	18 (81.82)	2 (8.00)	10 (40.00)	6 (24.00)	18 (72.00)
Total	-	16 (72.73)	6 (27.27)	22 (100.00)	2 (8.00)	14 (56.00)	9 (36.00)	25 (100.00)

(한국보건의료인국가시험원, 1999a).

나. 확장결합형 문항의 채점 방식과 난이도 및 분별도

확장결합형 문항은 주로 어떤 증례에 대하여 여러 전공에서 접근할 수 있는 진단이나 치료법을 모두 고르는 형식의 문항이므로 단순 암기정보보다는 해석형이나 문제해결형 문항을 출제하기에 더 알맞은 형태로서 확장결합형 문항의 지식수준별 구성을 살펴보면 해석형과 문제해결형이 주로 구성하고 있음을 알 수 있다(Table II).

위와 같이 확장결합형 문항은 세 지식수준의 문항을 동일한 비율로 포함하고 있는 것이 아니므로 지식수준에 따른 난이도와 분별도의 차이가 문항유형의 차이에도 영향을 줄 것으로 예측된다.

전체적으로 2002년도 시험은 난이도 평균 0.7150, 분별도 평균 0.2079이며, 2003년도 시험의 난이도 평균 0.6871, 분별도 평균 0.2138로서 2003년도 시험이 더 어려우면서 분별력이 좋은 시험이었다.

한편 2002년도 시험까지는 확장결합형 문항도 하나의 문항출기에 대해서는 하나의 정답지만 선택하도록 하고 있으나, 2003년도부터는 옳은 답을 2개, 3개까지도 고르게 하는 복수정답제를 택하였다. 그러나 복수정답제라 하더라도 부분점수는 인정하지 않고 있다.

따라서 확장결합형 문항의 신뢰도 및 문항분석은 2개년도 자료를 모두 사용하지만 복수정답제에 대

한 분석은 2003년도 자료만이 대상이 된다. 그리고 문항유형별 및 지식수준별 비교도 다른 문항유형과의 비교를 통해 확장결합형 문항의 양호도를 검증하고자 하는 것으로서 연도별 비교가 목적이 아니므로 문항난이도와 분별도 지수가 좋은 2003년도 자료를 대상으로 하였다. 보건의약관계법규는 전체 시험에서 차지하는 비율이 작을 뿐만 아니라 확장결합형 문항을 포함하고 있지 않으므로 이후 비교 분석의 대상은 주로 의학총론과 의학각론이다.

다. 분석 방법

첫째, 고전검사이론에 따라 문항유형 및 지식 수준에 따른 문항지수를 산출하고 분산분석을 통해 문항유형간, 지식수준간 문항지수 평균의 차이를 검증하였다. 난이도(item difficulty index)는 정답률(score)로, 분별도(item discrimination index)는 문항-총점상관관계(item total correlation)로 계산하였다(황인홍, 2000).

둘째, 확장결합형 문항의 내적일관성 신뢰도를 Cronbach alpha 값으로 표시하고, 과목별 확장결합형 문항에 대한 난이도 및 분별도의 평균을 구하여 연도별 차이를 비교하였다.

셋째, 확장결합형 문항에서 정답수에 따른 난이도 및 분별도의 차이를 분산분석(Analysis of variance=ANOVA)으로 관찰하였다.

DBSTAT 4.0(김수녕과 허선, 2002)을 이용하여 통계 처리하였다.

Table III. Mean Difference in Discrimination and Difficulty Indexes Based on the Item Types for the 2003 Korean Medical Licensing Examination

	Item type*				F value	P value
	A	K	R	Total		
	(N=271)	(N=184)	(N=25)	(N=480)		
Difficulty	0.72±0.23	0.65±0.23	0.61±0.27	0.69 ±0.24	7.763	0.000
Discrimination	0.22±0.09	0.20±0.10	0.28± 0.11	0.21± 0.11	9.092	0.000

*type: A-one best answer type, K-multiple true-false type, R-extended matching type

Table IV. Mean Difference in Discrimination and Difficulty Indexes Based on the Knowledge Level for the 2003 Korean Medical Licensing Examination

	Item type				F value	P value
	Recall	Interpretation	Problem-solving	Total		
	(N=131)	(N=212)	(N=137)	(N=480)		
Difficulty	0.70±0.24	0.69±0.25	0.68±0.23	0.69±0.24	0.231	0.794
Discrimination	0.20±.094	0.21±.096	0.23±0.10	0.21±.098	3.167	0.043

결 과

가. 문항유형에 따른 문항분석

난이도는 문항유형을 독립변수로 한 분산분석 결과 문항유형에 따라 차이가 있었다 ($p<0.05$). 확장결합형<조합형<단일정답형의 순서로 난이도 값이 나타나는데 이 난이도는 정답률이므로 정답률이 높은 단일정답형이 가장 쉽고 확장결합형이 가장 어려웠음을 나타낸다 (Table III). 어떤 유형 사이의 차이가 유의한 것인지 Scheffe 검증결과에서는 단일정답형과 조합형의 차이만 유의하게 나타났다 ($p<0.05$).

분별도 역시 분산분석 결과 문항유형에 따라 차이가 있었다 ($p<0.05$). 조합형<단일정답형<확장결합형의 순서로 확장결합형의 분별도가 가장 높았으며 (Table III), 이러한 유형 사이 차이가 유의한 것인지 검증하는 Scheffe 검증결과 단일정답형과 확장결합형, 조합형과 확장결합형 사이 유의한 차이가 있었다 ($p<0.05$).

나. 지식수준에 따른 문항분석

난이도는 지식수준에 따라 평균에 차이가 없었다 ($p>0.05$). 평균 정답률은 0.687, 암기형 정답률은 0.698, 해석형 정답률 0.685, 문제해결형 정답률은 0.678 이었다 (Table IV). 분별도는 지식수준을 독립변수로 한 분산분석 결과 평균에 차이가 있었다 ($p<0.05$). 암기형<관단형<문제해결형의 순서로 문제해결형의 분별도가 가장 높았다 (Table IV).

다. 확장결합형 문항의 신뢰도 및 문항분석

확장결합형 문항간 상관관계는 매우 높은 편으로 내적일관성 신뢰도 (Cronbach- α)는 모두 0.99를 넘는 양호한 수준이었으며, 각 문항이 제거되었을 때의 신뢰도 값을 보았을 때에도 전체 신뢰도를 감소시키는 문항은 없었다. 또한 난이도는 2002년도에 더 쉬운 문항이었으며, 반면 분별도는 2003년도에 더 높았다 ($p<0.05$, Table V).

각 문항의 난이도 및 분별도 값을 구체적으로 살펴보면 Table VI과 같다.

Table V. Item Analysis and Reliability of Extended-Matching Type Items of the 2002 and 2003 Korean Medical Licensing Examinations Using the Classical Theory

Year	2002		2003	
No. examinees	3,576		3,647	
Subjects	General part	Special Part	General part	Special Part
No. items	4	18	7	18
Mean Difficulty index	0.650	0.859	0.761	0.544
Mean Discrimination index	0.235	0.233	0.269	0.285
Cronbach-alpha	0.9975	0.9997	0.9938	0.9975

Table VI. Difficulty and Discrimination Indexes of Extended-Matching Type Items in 2002, 2003

Subject	Item No.	2002		Item No.	2003	
		Difficulty	Discrimination		Difficulty	Discrimination
General Part	1	0.38	0.27	1	0.44	0.29
	2	0.96	0.24	2	0.87	0.21
	3	0.90	0.21	3	0.95	0.27
	4	0.36	0.22	4	0.84	0.37
				*5	0.74	0.19
				*6	0.98	0.16
				7	0.51	0.39
	Mean	0.65	0.24	Mean	0.76	0.27
Special Part	1	0.91	0.31	*1	0.07	0.12
	2	0.98	0.21	2	0.51	0.25
	3	0.92	0.32	3	0.55	0.29
	4	0.52	0.37	*4	0.20	0.18
	5	0.95	0.29	5	0.76	0.15
	6	0.92	0.21	*6	0.17	-0.05
	*7	0.42	0.11	7	0.56	0.41
	8	0.79	0.18	8	0.75	0.38
	*9	1.00	0.14	9	0.60	0.37
	10	0.99	0.15	10	0.38	0.26
	11	0.90	0.34	11	0.32	0.36
	12	0.62	0.31	12	0.58	0.49
	13	0.86	0.34	13	0.91	0.35
	14	0.96	0.23	14	0.99	0.25
	15	0.95	0.22	15	0.83	0.34
	16	0.94	0.21	16	0.83	0.31
	17	0.97	0.23	17	0.41	0.35
	18	0.87	0.35	18	0.38	0.32
	Mean	0.86	0.25	Mean	0.54	0.29

* Items with discrimination index less than 0.20.

Table VII. Mean Difference in Discrimination and Difficulty Indexes Based on the Number of Correct Answers in Extended-Matching Type Items for the 2003 Korean Medical Licensing Examination

	No. of correct answers				F value	P value
	1 (N=12)	2 (N=9)	3 (N=4)	Total (N=480)		
Difficulty	0.80±0.17	0.43±0.19	0.41±0.25	0.61±0.27	12.796	0.000
Discrimination	0.27±0.08	0.31±0.11	0.24±0.11	0.28±0.1	0.564	0.577

라. 확장결합형 문항의 정답수에 따른 문항지수 변화

2003년도 확장결합형은 정답을 답지에서 하나만 고르는 것이 아니라 2개, 3개까지 고르게 할 수 있도록 한 데에 2002년도와 차이가 있다. 따라서 이와 같이 정답개수를 증가시키는 것에 따라서 난이도와 분별도 값이 영향을 받는지를 알아보았다. Table VII에서와 같이 난이도에서는 집단간 유의한 차이가 있었고 ($p<0.05$), 분별도에는 정답개수에 따른 차이가 유의하지 않았다 ($p>0.05$). 난이도는 골라야 할 정답개수에 따라 큰 차이가 관찰되었는데, 정답을 1개 고를 때 정답률이 0.80이나 되는데 비하여 2개 고를 때 0.43으로 낮아졌고, 3개를 고를 때, 0.41로 떨어졌다.

고 찰

확장결합형 문항의 가장 큰 장점은 의학 안의 여러 다른 전공 분야에서 같은 주제아래 동질성이 있는 같은 내용을 놓고 여러 관점에서 동시에 평가할 수 있다는 점이다. 또한 답가지 수가 많으므로 우연에 의해 정답을 맞출 확률이 줄어들고, 높은 지식수준을 측정하는 문항 개발이 용이하므로 문항의 분별력과 신뢰도를 높일 수 있다고 기대한다(한국보건 의료인국가시험원, 1999b). 의사 국가시험 과목이 과거 내과, 외과, 소아과 등 전공 영역별 과목으로 과목을 나누던 것에서 2002년도 시험부터 현재와 같이 통합한 것은 각 전공에 국한하는 단편적인 문항에서 벗어나 통합적인 시각에서 문제를 해결할

수 있는 능력을 측정하자는 노력의 일환이다. 특히 확장결합형 문항은 모든 전공에서 나올 수 있는 병명 등을 답가지 상자로 나열하고 있기 때문에 과목을 통합하는 것은 확장결합형 문항을 도입하기에 적절한 환경을 만들어준 것으로 볼 수 있다. 이러한 과목통합과 확장결합형 문항의 도입을 통해 국가시험의 질적 향상을 지향하려고 노력하고 있다.

이러한 노력이 어느 정도 성과를 나타냈는지 보기 위한 문항분석 결과를 종합하면, 2003년도 의사 국가시험 응답을 고전검사이론으로 문항분석을 시행한 결과 확장결합형이 분별도가 높고, 정답률이 낮은 어려운 문항으로서 다른 두 문항유형보다 통합적이고 고차원적 문제를 내기 위한 확장결합형 문항의 취지에 부합함을 알 수 있다. 난이도는 조합형이 단일정답형보다 더 낮아서, 단일정답형이 가장 쉽고, 그 다음 조합형, 그리고 확장결합형이 가장 어렵다는 것을 알 수 있었다. 반면 분별도는 확장결합형이 가장 높고, 다음이 단일정답형이고, 조합형이 가장 낮은 것으로 나타났다. 조합형은 단일정답형보다 어려운 문항이지만 분별력은 가장 떨어졌다. 2003년도 시험을 대상으로 지식수준에 따라 난이도는 차이가 없었는데, 이와 같은 결과는 일반적으로 암기형 문항보다는 해석형이나 문제해결형이 확연히 더 어려울 것이라는 기대에 충분히 부합하는 수치는 아니었다. 따라서 확장결합형 문항이 해석형과 문제해결형 문항을 많이 포함하고 있지만 지식수준에 의한 영향보다는 확장결합형이라는 문항유형의 특성에 의해 분별도와 난이도가 높게 나타나는 것으로 볼 수 있다.

확장결합형 문항의 신뢰도가 높았고 전체 신뢰도를 감소시키는 문항은 없었다(Table V). 분별도 지수로 문항을 평가하는 절대적 기준은 없으나 검사도구의 신뢰도와 관련하여 설정한 Ebel (1965)의 기준은 0.40 이상이면 분별력이 높은 문항, 0.3 이상이면 좋은 문항, 0.2 미만인 문항은 분별력이 낮은 문항으로 평가한다. 이러한 기준으로 볼 때 분별도가 0.2 미만인 문항을 “*”로 표시한 Table VI에서 보면 의사국가시험 확장결합형 문항의 분별도는 2002년도에 2개 문항, 2003년도에 5개 문항이 0.2 미만을 나타내었고 다른 문항들은 0.2~0.5 사이로 양호하였다. 문항 난이도 역시 절대적 평가기준은 없으나 0.30 미만이면 매우 어려운 문항, 0.30~0.80 미만이면 적절한 문항, 0.80 이상이면 매우 쉬운 문항이라고 평가한다(성태제, 1996). Cangelosi (1990)은 0.25 이하를 어려운 문항, 0.25~0.77을 적절한 문항, 0.75 이상을 쉬운 문항으로 분류한다. 이러한 기준에 따르면 의사국가시험의 확장결합형에서 어려운 문항은 없었고 2002년도 확장결합형 문항은 매우 쉬웠으며 2003년도 난이도는 적절한 것으로 평가한다. 확장결합형 문항이 2002년도 쉬운 것은 처음 시도하는 것이므로 쉬운 문제로 출제된 것이 원인일 수 있다. 2002년도 의학각론 9번의 경우 또한 2003년도 의학각론 1번의 경우처럼 지나치게 어렵거나 지나치게 쉬운 문항 모두 문항의 분별력을 떨어뜨려 평가의 제 기능을 하지 못한다. 특이한 점은 난이도 차이는 정답개수에 따라 현격한 차이가 난 반면, 분별도 지수는 큰 차이가 나지 않았다는 점으로, 정답의 개수가 늘어남에 따라 전체적으로 어려워졌을 뿐, 분별력의 증가 효과는 없었다. 그리고 분별도는 비록 통계의 유의성은 부족하나 평균값에서 정답수가 2개 일 때 가장 높고, 그 다음 1개, 3개 순서를 나타내고 있으므로 현재 부분점수를 인정하지 않고 있는 한 정답을 3개까지 고르게 하는 것은 분별력을 떨어뜨리고 있다. 부분점수 없이 3개를 모두 맞추어야 정답으로 인정되는 상황은 문항의 정답률 뿐만 아니라 분별력까지 현저히 떨어뜨리는 것으로 보인다. 정답 문항의 수를 늘이는 복수정답형 확장결합형 문항의 확대는 재검토되거나 좀 더 양질의

문항으로 개선되어야 할 필요가 있다.

의사국가시험에서 확장결합형 문항은 2002년도부터 도입되었으므로 대규모 시험에서 확장결합형 문항의 시행 결과를 대상으로 문항분석을 하는 것은 첫 시도로서 본 연구와 동일 과정을 수행한 관련 선례를 찾기는 어렵다. 이와 같거나 비슷한 종류의 결과를 과거에 보고한 내용은 적어도 국내 의학학술지 논문을 검색하는 코리아메드(<http://koreamed.org>) 및 전세계 의학학술지 논문을 검색하는 퍼브메드(<http://pubmed.org>)에서 찾을 수가 없다. 그러나 검사도구의 제작 및 평가 과정에서 문항의 난이도, 분별도, 신뢰도를 분석하는 것은 일반적인 연구방법이다.

검사도구가 신뢰성을 가지려면 이와 같은 문항분석 및 평가가 지속적으로 이루어져야 할 것이다. 앞으로 확장결합형 문항의 비율이 증가할 것이고 더 고차원 정신능력을 측정하기 위한 문항으로서 의의를 가지려면 확장결합형 문항이 다른 유형보다 적절한 난이도와 높은 분별도를 유지하는 것이 필요하다. 정답 개수를 증가시키는 것만이 난이도 및 분별도와 같은 문항 지수를 향상시키는 작업이 아니므로 적절한 난이도와 높은 분별도로 문항을 만드는 최적의 조건을 찾아내는 작업도 후속 연구로서 제안한다. 또한 앞으로 이번에 사용한 고전검사이론 뿐만 아니라 문항반응이론에 따라 분석하였을 때 두 방법에 따른 결과를 보고 문항반응이론 적용의 유용성을 찾는 작업도 필요하다.

참 고 문 헌

- 김수녕, 허선(2002). 국산통계플그림으로 우리 나라 의학 학술지에서 사용하는 통계를 해결할 수 있을까? *한국의학교육* 14(1), 111-117.
- 성태제(1996). *문항제작 및 분석의 이론과 실제*. 서울: 학지사.
- 한국보건의료인국가시험원(1999a). *지식수준별 MC Q 작성법*. 서울. <http://www.kuksiwon.or.kr> 에서 구할 수 있음.
- 한국보건의료인국가시험원(1999b). *확장결합형(R형)*

- 시험문항 만들기 - 기초이론에서 작성까지. 서울.
<http://www.kuksiwon.or.kr> 에서 구할 수 있음.
- 황인홍(2000) 분별도 지표로서 문항·총점 상관계수의 활용. *한국의학교육* 12(1), 45-51.
- Bloom BS(1956). *The Taxonomy of Educational Objectives*, Hamdbbook I, Cognitive domain. New York, U.S.A.: David Mckay.
- Cangelosi JS(1990). *Designing tests for evaluating student achievement*. New York, U.S.A.: Academic Press.
- Case SM, Swanson DB(1998). *Constructing written test questions from the basic and clinical sciences (2nd ed.)*. Philadelphia, U.S.A.: National Board of Medical Examiners. Available from <http://www.nbme.org>.
- Ebel RL(1965). *Measuring Educational Achievement*. Englewood Cliffs, New Jersey, U.S.A.: Prentice-Hall.