

문항반응이론을 적용한 1급 응급구조사 국가시험 대비 모의시험의 적정성 연구

이정은¹ · 문준동^{2*} · 김아정³

¹동남보건대학 응급구조과 교수

²국립공주대학교 응급구조학과 교수

³경일대학교 응급구조학과 교수

A study of the adequate number of questions in a mock test for the paramedic national examination using item response theory

Jung Eun Lee¹ · Jundong Moon^{2*} · Ajung Kim³

¹Professor, Department of Paramedicin, Dongnam Health University

²Professor, Department of emergency medical service, Kongju National University

³Professor, Depatment of Paramedicine, Kyungil University

= Abstract =

Purpose: To adjust item numbers in a national test, this study used item response theory to examine changes in average scores, reliability, difficulty, and discrimination according to the adjustment of item numbers.

Methods: We analyzed the dichotomous coding of correct and incorrect answers of 473 examinees in a mock test conducted in 2023. Additionally, as an explanatory pilot study, we used an online questionnaire to survey experts on their perceptions of the appropriate item numbers for each test subject from January 18, 2024, to February 15, 2024.

Results: Regarding the item numbers on the national exam, experts preferred to reduce the number of management of emergency patients (33.14 ± 6.09 , $p < .05$) and advanced emergency medical care:

Received June 15, 2024 Revised July 16, 2024 Accepted August 13, 2024

*Correspondence to Jundong Moon

Department of emergency medical service, Kongju National University, Gongjudaehak-ro 56
 (Singwan-dong 182), Gongju-si, Chungcheongnam-do, Republic of Korea

Tel: ***-****-**** E-mail: jdm02@kongju.ac.kr

†본 연구는 동남보건대학교 연구비 지원에 의해 수행된 것이다.

subtopics (104.49 ± 11.55 , $p < .05$), and the total number of questions (217.82 ± 20.95 , $p < .05$). In a simulation set in which items with low item fit were removed after fitting a two-parameter item response theory model, reliability was maintained at .910 until the 5th test consisting of 185 questions with little loss of difficulty, discrimination, and average score, and there was no correlation between item numbers and average score.

Conclusion: Experts responded that reducing the number of items on the national exam was appropriate. As a result of the item reduction simulation, there was no significant loss in the average score, difficulty, discrimination, or reliability. More reliable results could be obtained if the results were based on a validity analysis and analyzed using actual national exams.

Keywords: Item response theory, High-stake examination, Item number, Average score

I. 서 론

1. 연구의 필요성

1) 연구의 필요성

1급 응급구조사 국가시험은 전문적인 활동을 보장하기 위한 국가자격시험이다. 1급 응급구조사 국가자격시험은 실기시험과 필기시험으로 구분되어 있으며, 실기시험과 필기시험을 합격하여야 최종적으로 자격증을 취득하게 된다[1]. 응급구조사 국가시험은 2023년 현재 29회에 이르고 있으며 그동안 문항 기준과 문항수가 조정되어 왔으며 최종적으로 2016년에 조정되었다. 2023년 현재 응급구조사 1급 필기시험의 경우 기초의학 30문제, 응급환자관리학 40문제, 전문응급처치학총론 30문제, 응급의료관련법령 20문제, 전문응급처치학각론 110문제의 5개 과목 총 230문제로 이루어져 있다. 현행 230문제에 대한 문항수는 2016년 제22회 국가시험부터 적용되고 있으며 2016년 당시 기초의학 40문제에서 30문제로, 전문응급처치학총론이 40문제에서 30문제로, 전문응급처치학각론이 90문제에서 110문제로, 응급의료관련법령이 30문제에서 20문제로 변경되었다[2].

한편 한국보건 의료인 국가시험원에서 주관하는 보건 의료직종 중 대학교 3, 4년제로 운영

하여 응급구조사 1급 교육과정과 유사한 보건직종(임상병리사, 방사선사, 물리치료사, 작업치료사, 치과기공사, 치과위생사, 안경사)와 비교 결과, 보건직종 시험 과목 수는 필기 3과목과 실기 1과목으로 실시되고 있으나 응급구조사 1급은 5과목으로 운영되고 있는 것으로 나타나 응급구조사 직종의 과목 수가 많은 것으로 나타났다. 보건 의료 직종의 필기시험의 총문제수는 190 ~ 215개로 응급구조사 1급 국가시험 필기시험인 230문제보다 적게 나타나 문항수 역시 응급구조사 1급 직종이 많은 것으로 나타났다. 또한 과목당 문제 배점은 법률 관련 시험 과목의 문제수가 20문제였으며, 다른 시험 과목은 45 ~ 115점으로 나타났다. 응급구조사 1급 직종은 30 ~ 110점으로 나타나 응급구조사 1급 직종이 시험 과목 당 문제수가 적은 것으로 나타났다[3].

응급구조사 국가시험은 고부담 준거지향평가로 분류할 수 있다. 준거지향평가는 교수 목표에 기초해서 검사 문항을 제작하고, 다른 동료 학습자가 얼마나 성취했는지와는 상관없이 교수 목표에 비추어서 학습자의 성취 수준을 확인하고 판정[4]하므로 타당도가 중요하다고 할 수 있다. 따라서 응급구조사의 교육 목표는 사회가 응급구조사에게 요구하는 수행 능력과 대학의 교과 구조에 따른 요구 역량을 고르게 적

절히 반영하여야 하며, 응급구조사 국가시험은 이러한 교육목표에 기반한 타당도 높은 문항으로 구성되어야 한다.

그러나 응급구조사에게 요구되는 역량은 응급의료체계의 발전과 같은 사회의 변화에 따라 정의가 바뀌게 되기 때문에, 여러 의학 지침들의 개정과 응급구조사 업무범위 변화 등에 따라 2016년에 개정된 응급구조사 국가시험 출제 문항 기준은 다시 검토되어야 할 시점이라고 사료된다. 또한 응급구조사 국가시험에서는 문제해결형 문항의 비중이 높아지고 있는데, 이는 비구조화된 문제들로서 단편적인 암기나 공식의 적용을 통해 손쉽게 구조화되지 못하는 특성을 가지고 있으며[5], 문제해결형 문항에서 요구하는 분석, 종합, 평가 등의 고차원적 인지 능력은 단순 지식의 암기 및 이해를 전제로 하기 때문에, 측정하고자 하는 역량의 범위가 넓은 편이다. 또한 최근 보건의료직군을 대상으로 국가고시에서 통합형 문항의 도입에 대한 논의가 진행되고 있고[6], 이러한 추세들은 모두 국가시험의 출제기준의 변화와 함께 문항수의 조절을 동반하게 된다. 이러한 타당도 기준에 의해 문항수를 조절할 때, 검사의 양호도를 판단하는 또 다른 기준인 신뢰도를 고려해야 한다. 또한 문항 분석 지표인 난이도, 변별도, 문항반응분포 등도 변화할 수 있고, 이에 따라 평균 점수도 변할 수 있다는 점을 고려해야 한다. 국가시험 같은 고부담 시험에서는 더욱 그러하다. 선행연구에 의하면 다른 조건이 모두 동일할 때 신뢰도 계수는 문항수가 증가할 때 점수 분산이 커짐으로 해서 비례하는 것으로 알려져 있고[4], 문항반응이론에서 검사의 변별도가 증가할 때 신뢰도 계수는 계속 증가하며, 신뢰도 계수는 평균 문항난이도가 0일 때 대체로 큰 값을 갖는 걸로 보고하고 있다[7].

문항반응이론은 문항 특성의 불변성, 피험자 능력의 불변성이라는 특징 때문에 고전검사이

론을 보완 또는 대신하여 문항 분석에 널리 사용되고 있다. 문항반응이론을 의사국가시험, 간호사국가시험에 적용하여 분석한 연구들이 있으며[8, 9], 의사국가시험의 타당도와 신뢰도를 유지하는 적정의 범위를 산정한 연구[10], 컴퓨터 적응시험 도입에 대한 연구[11]들이 있다. 반면 응급구조사 국가고시에서 이러한 분석 시도는 아직 활발하지 못하며, 응급구조사 직종의 국가시험에 대한 연구는 과목개선에 대한 연구[12], 실기시험 개선에 대한 연구[13], 국가시험 응시자격기준에 대한 연구[14], 최근 9년간 1급 응급구조사 국가자격시험의 결과를 분석한 연구[15] 등으로 응급구조사의 국가 자격시험에 대한 연구, 특히 문항 분석에 따른 국가시험의 타당성에 대한 연구는 없는 실정이다. 시험의 타당도와 신뢰도를 검토하는 것은 검사 도구의 질을 측정학적으로 검토할 수 있는 의미있는 작업이며, 더불어 사회 환경의 변화에 맞추어 국가시험의 발전을 위해 필수적으로 수행해야 할 선결 과제라고 생각된다. 이에 본 연구는 향후 국가시험 문항수 조절을 대비한 예비연구로서 문항반응이론을 적용하여 1급 응급구조사 국가시험 모의고사 결과를 분석하고, 문항수 조절에 따른 평균점수와 신뢰도, 난이도, 변별도의 변화를 살펴보고자 한다.

2) 연구 목적

본 연구에서는 문항반응이론을 사용하여 응급구조사 국가시험과 같은 과목수와 문항수로 실시한 모의고사를 분석하여 국가시험 모의시험의 신뢰도, 난이도, 변별도를 검증하고, 산출된 결과를 바탕으로 타당도와 신뢰도를 높게 유지하는 적정 문항수의 범위를 산정하는 것을 목적으로 한다.

구체적인 연구 목적은 다음과 같다.

- (1) 1급 응급구조사 국가시험 문항수에 대한 전문가의 인식을 분석한다.

- (2) 1급 응급구조사 국가시험의 년도별 합격률과 난이도, 변별도, 신뢰도를 정리·분석한다.
- (3) 문항반응이론을 이용하기 위한 모형 적합도 검정과 일차원성 검정을 실시한다.
- (4) 문항수 제거에 따른 문항 분석을 실시하여 적정 문항수에 따른 신뢰도, 난이도, 변별도를 비교한다.

II. 연구방법

1. 연구설계

본 연구는 국가시험 모의고사 시험 결과를 이용하여 문항반응이론 2모수 모형을 적용하여 분석한 후향적 관찰연구이다.

2. 연구대상 및 자료수집 방법

2023년에 시행된 1급 응급구조사 국가시험 대비 모의고사에 대한 473명의 답안을 연구 대상으로 하였고, 대상 자료는 모의고사를 주관한 A출판사에 자료를 요청하였으며, 과목별 문항은 암호화되고, 각 문항의 답안이 정답과 오답, 이분형으로 코딩된 자료로 개인정보에 대한 사항, 문항에 대한 정보는 일체 알 수가 없다.

또한 사전 연구로서 2024년 1월 18일부터 2월 15일까지 온라인 설문지를 이용하여 전국의 현직 응급구조학과 교수를 대상으로 각 교과목의 적정 문항수에 대한 인식을 조사하였다. 설문 항목은 각 과목별 적정하다고 판단되는 문항수 및 전체 문항수이다. 동시에 한국보건의료인 국가시험원의 홈페이지에 공개된 2012년부터 2022년까지 시행된 1급 응급구조사 국가시험 문항 분석결과를 년도별로 문항수, 합격률, 난이도, 변별도, 신뢰도를 정리하였다.

3. 분석 방법

문항 및 시험 분석에는 jMetrik 4.1.1(Psychomeasurement Systems, Charlottesville, VA, USA)을 사용하였으며 일차원성 가정 분석과 피어슨 상관계수 분석에는 Jamovi 2.3.28(The jamovi project, <https://www.jamovi.org>) 을 사용하였다. 빈도, 평균과 표준편차와 같은 기술통계치는 모두 SPSS 29.0 (IBM copr., Chicago, USA)로 시행하였다. 구체적인 분석 방법은 다음과 같다.

1) 1급 응급구조사 국가시험 문항수에 대한 전문가 인식 분석 방법

1급 응급구조사 국가시험 문항수 적정성에 대한 온라인 설문 응답은 총 78명 이였고, 국가고시 과목별 적정 문항수 설문 결과와 현행 국가고시 문항수 차이를 모평균과 비교하였다.

2) 년도별 1급 응급구조사 국가시험 분석

한국보건의료인 국가시험원 홈페이지에서 제공하는 각 직종별 국가시험 결과 분석 자료 중 응급구조사 1급의 2012년 ~ 2022년까지 시험 분석 결과를 분석하였다. 고전검사이론에 근거하여 국가시험 종료 후 매년 발간하는 한국보건의료인 국가시험원의 문항은 문항수 변경이 있던 2016년을 전후로 하여 합격률과 문항 분석 지표의 변화를 정리하였다. 지표의 표준편차가 제시가 되지 않은 년도가 있어, 표준편차는 제시하지 못했다. 문항수가 240문항에서 230문항으로 바뀐 2016년 이전과 이후 자료를 분석하여 합격률, 난이도, 변별도, 신뢰도를 조사하였다.

3) 문항반응이론을 이용한 모형의 적합도 검정

문항반응이론 1모수 모형과 2모수 모형 중 어떤 모형이 표본 데이터에 적합한지 상대적인

비교를 하기 위해 모형선택기준인 AIC, BIC 통계량을 사용하였다. AIC, BIC 값이 더 적은 값이 나온 모형이 더 나은 적합도를 의미한다. 문항반응이론 2모수 모형의 AIC와 BIC가 문항반응이론 1모수 모형에 비해 값이 더 적게 나왔으며, 카이제곱 검정에서 유의확률이 $p < .000$ 으로 유의하여 문항반응이론의 1모수 모형보다는 2모수 모형이 더 적합한 것으로 나타나 이후 문항 분석은 문항반응이론 2모수 모형에 의거하여 진행하였다.

4) 문항반응이론을 이용한 2모수 모형의 일차원성 검정

문항반응이론의 2모수 모형을 적용하였으며, 전체 230문항 중 모두 정답으로 나타난 기초의학에서의 1문항은 문항반응이론의 분석에 부적절하여 제외하였다. 또한 응급구조사 1급 국가시험의 과목인 기초의학, 응급환자관리학, 전문응급처치학총론, 응급의료관련법령, 전문응급처치학각론의 5과목 중 다른 과목과 비교하여 측정하고자 하는 특성이 다르기 때문에 응급의료관련법령 20문항을 제외한 209문항을 최종 분석 대상으로 하였다.

문항반응이론의 전제조건인 일차원성 가정을 검정하기 위하여 주성분 요인분석을 실시하였으며 회전방법은 Verimax 방법으로 하였으며 Bartlett의 구형성 검증에서는 $\chi^2 = 30982$, 자유도 21736, $p < .001$ 로 나타났으며, KMO 구형성 검증은 0.625로 조금 낮게 나오기는 하였으나 요인분석에 적절한 척도임을 확인하였다.

5) 문항수 제거에 따른 문항 분석

문항반응이론 2모수 모형을 적용하여 응급의료관련법령 20문항과 모두 정답으로 나온 기초의학 1문항을 제외한 총 209문항이 최종 분석 대상이 되었다. 209문항에 대해 문항반응이론 2모수 모형으로 분석 결과를 토대로 시뮬레이션 세트를 구성하였다. 시뮬레이션 세트 구성

시 문항제거 우선순위는 모형적합도가 낮은 문항, 난이도와 변별도가 낮은 순서로 1차 세트(Test 1)는 4문항을, 나머지는 5문항씩 제거하여 총 7세트(Test 2 ~ Test 7)를 구성하였다. 전체 209문항과 7개의 테스트 세트의 문항수와 100점 만점으로 환산한 점수의 상관관계를 분석하였으며, 각 테스트는 문항반응이론 2모수 모형으로 분석하여 평균점수와 변별도, 난이도, 신뢰도를 비교분석하였다.

III. 연구결과

1. 1급 응급구조사 국가시험에 대한 전 문항 인식분석

응급구조학과 교수를 대상으로 온라인 설문을 통해 수집한 1급 응급구조사 국가시험 적정 문항수에 대한 응답은 (Fig. 1)과 같다. 응급환자관리학(33.14 ± 6.09)과 전문응급처치학각론(104.49 ± 11.55)에서 축소의 응답이 많았고, 전체 문항수도 현행 230문항에서 감소한 217.82 ± 20.95 의 응답을 보였다.

2. 년도별 1급 응급구조사 국가시험 분석

고전검사이론에 근거한 한국보건의료인국가시험원의 문항 분석결과 보고서를 이용하여 년도별 합격률과 난이도, 변별도, 신뢰도를 정리한 결과는 (Table 1)과 같다. 변경된 출제문항 기준이 적용된 2016년 이전과 이후를 비교하면 난이도는 쉬워졌으며, 변별도는 감소하고 합격률은 조금 증가하였으며, 신뢰도는 조금 감소하였다. 그러나 2021년 국가시험의 특이점 효과를 제외하면, 소폭의 난이도 하락과 이에 따른 합격률 향상을 보였고, 변별도와 신뢰도는 큰 차이를 보이지 않았다.

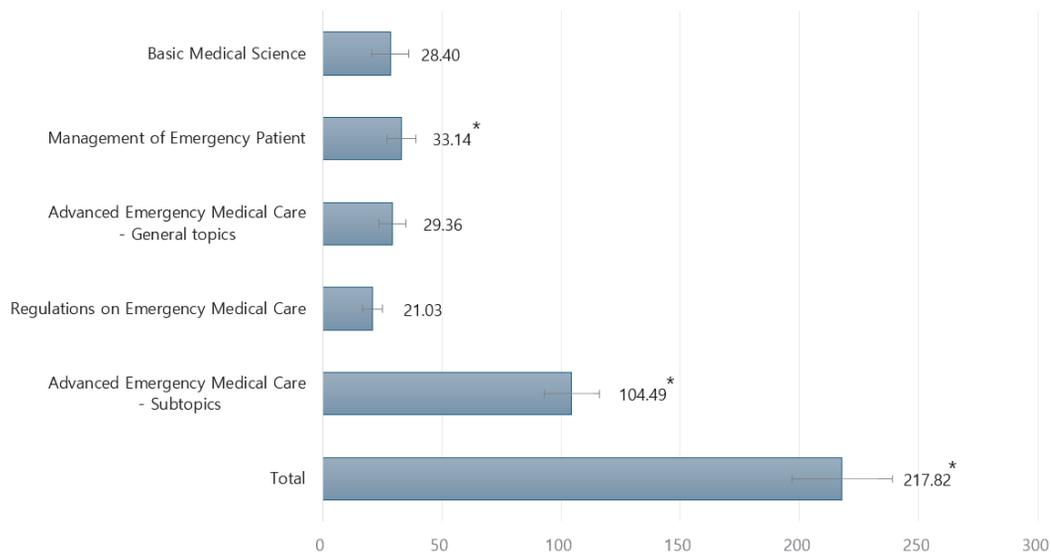


Fig. 1. Online survey summary of the appropriate item numbers (marked with numbers) for the level 1 emergency medical technician national exam. The asterisk(*) indicates that it is statistically significant in a two-tailed test at a significance level of 0.05 when compared to the item numbers in the current national examination.

Table 1. Item numbers, passing rate, difficulty, discrimination, and reliability of the national examination for level 1 emergency medical technicians by year

Year	Item numbers	Passing rate(%)	Difficulty	Discrimination*	Reliability
2012	240	86.3	68.8	.27	.943
2013	240	87.6	69.4	.25	.938
2014	240	81.5	68.2	.27	.945
2015	240	84.5	67.7	.25	.934
Mean		85.0	68.5	.26	.940
2016	230	84.5	69.3	.25	.937
2017	230	88.1	70.5	.24	.935
2018	230	88.3	71.8	.23	.932
2019	230	88.1	70.5	.24	.934
2020	230	85.9	68.9	.25	.938
2021	230	75.3	67.1	.23	.926
2022	230	88.1	74.0	.26	.945
Mean		85.5	70.3	.24	.935
Mean[†]		87.2	70.8	.25	.937

the difference in difficulty between the top 27% and bottom 27% groups

[†]Average of indicators from 2016 to 2022, excluding 2021

3. 문항반응이론을 이용한 모형의 적합도 검증

1급 응급구조사 국가시험 모의고사 결과를 바탕으로 한 모형의 선택 결과는 문항반응이론 1모수 모형인 Rasch 모형보다는 문항반응이론 2모수 모형이 적합하였다(Table 2). 모의고사 과목 중 응급의료관련법령은 문항들이 모두 동일한 특성을 측정하여야 한다는 문항반응이론의 일차원성 가정에 개념적으로 위배된다고 판단하였고, 실제 문항특성곡선에서도 다른 과목과 비교하여 능력치가 낮은 수험성과 높은 수험생에서 다른 과목과 비교하여 표준오차가 매우 커 제외하고 검증하였다.

4. 문항반응이론을 이용한 2모수 모형의 일차원성 검증

문항반응이론의 전제 조건인 일차원성 가정을 검증하기 위해 요인분석을 실시하여 고유값의 차이를 분석하였다. 주성분 분석으로

varimax방법으로 하였을 때 요인분석의 모형 적합지수는 $\chi^2=30982$, $df=21736$, $p<.001$ 로 적합한 것으로 나타났다. 요인분석을 하면 고유값은 문항수 만큼 계산된다. 1요인과 2요인의 차이는 10.9475로 나타나 2요인과 3요인의 차이인 0.3637과 3요인과 4요인의 차이인 0.2519보다 큰 것으로 나타났다(Table 3). 따라서 1요인과 2요인의 차이가 다른 요인의 차이보다 크게 나타나 일차원성 가정을 충족하였다고 볼 수 있다. 또한 일차원성 가정이 만족되었기 때문에 지역독립성 가정도 충족한다고 판단할 수 있다.

5. 문항수 제거에 따른 문항 분석

총 209문항 중 문항반응이론 2모수 모형에 문항적합도가 낮은 순서대로 5문항씩 문항을 제거(1차 시험은 4문제 제거)한 과목별 제외 문항 결과는 <Table 4>와 같다. 8개의 시험에서와 100점 만점으로 환산한 평균 점수는 상관관계가 없었다(Table 5).

각 시험의 평균점수와 변별도, 난이도, 신뢰

Table 2. Identification of item response theory model

Model	AIC*	BIC†	LogLik‡	χ^2	df	p
Rasch	96215.53	97088.94	-47897.77	-	-	-
2PL§	95035.14	96773.65	-47099.57	1596.388	208	0

*Akaike's Information Criterion

†Bayesian Information Criterion

‡log likelihood ratio

§2-parameter logistic model

Table 3. Difference in Eigenvalues for testing the unidimensionality assumption

Items	Factor	1	2	3
Total(209)	Eigenvalues	14.9570	4.0095	3.6458
	Difference of Eigenvalues	10.9475*	0.3637†	0.2519‡

*Difference between 1 and 2 factor

†Difference between 2 and 3 factor

‡Difference between 3 and 4 factor

Table 4. Results of excluded questions by subject

Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Total
Basic Medical Science							
1	1	1					3
Management of Emergency Patient							
1	1	3	1	2	2	1	11
Advanced Emergency Medical Care - General topics							
			1			1	2
Advanced Emergency Medical Care - Subtopics							
2	3	1	3	3	3	3	18
Number of excluded items							
4	5	5	5	5	5	5	34
Total Item numbers							
205	200	195	190	185	180	175	

Table 5. Correlation between number of questions and average % score

Pearson's r	95% CI (Upper)	95% CI (Lower)	df	p
-0.124	.636	-.762	6	.769

도는 <Table 6>과 같다. 변별도는 '적절'의 범위인 0.65에서 1.34사이를 유지하고 있었으며, 난이도는 '쉽다'의 범위인 -2.0에서 0.5 사이를 보였다. 총 185문항으로 구성된 5차 시험까지 신뢰도 .910을 유지하면서 변별도의 변동이 이전 시험과 비교하여 적은 것으로 나타났다. 난이도는 3차 시험부터 급격히 낮아지는데, 변별도와 신뢰도의 변동은 크지 않았기 때문에 난이도가 높지만 질이 높지 않은 문항이 제거된 것이라 사료된다. 원 시험과 5차 시험은 모두 피험자의 능력이 -1.667에서 가장 높은 설명력을 보여주었다. 합격 하한선에 해당하는 진점수 60점에 해당하는 피험자 능력은 -0.465에서 -0.483으로 다소 감소하였다(Fig. 2). 또한 원 시험과 5차

시험은 모두 합격선인 진점수 60점보다 낮은 수준의 수험생을 더 잘 설명하는 검사 특성을 가지고 있었다.

IV. 고 찰

본 연구는 1급 응급구조사 국가시험대비 모의시험에 문항반응이론을 적용하여 문항 분석을 시행하고, 그 결과를 이용하여 적절한 국가시험을 문항수를 산출해 보고자 하였다.

문항 분석의 여러 지표를 충분히 고려하지 않고, 타당도만을 고려하여 문항수를 조절하는 것은 평가도구의 양호도 측면에서 적절하지 않

Table 6. Number of questions, percent score, discrimination, difficulty and reliability per test

Test	Mean±SD	% Score	Discrimination	Difficulty	Cronbach α	KR21
0	133.8±20.2	64.0	0.702±0.39	-0.300±5.84	0.918	0.886
1	130.3±19.8	63.56	0.701±0.39	-0.246±5.85	0.916	0.883
2	127.3±19.6	63.67	0.701±0.39	-0.447±5.40	0.915	0.884
3	124.0±19.2	63.59	0.709±0.40	-0.757±4.25	0.914	0.883
4	120.8±18.7	63.58	0.705±0.40	-0.631±4.62	0.911	0.879
5	119.0±18.3	64.33	0.701±0.40	-0.661±4.98	0.910	0.878
6	117.0±17.7	64.99	0.698±0.40	-0.725±5.14	0.906	0.874
7	113.8±17.2	65.06	0.698±0.40	-0.710±5.02	0.903	0.871

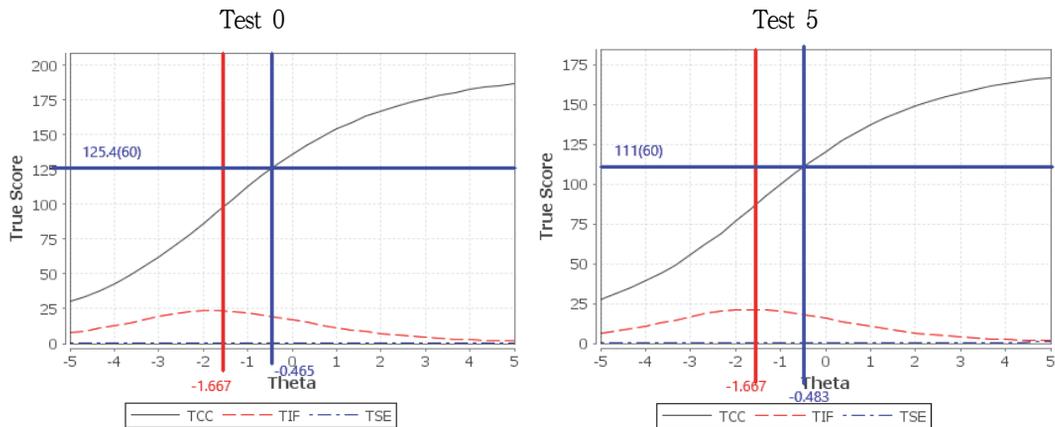


Fig. 2. Test characteristic curve (TCC), test information function (TIF) and test standard error curve (TSE) of the original test (209 questions) and the 5th test (185 questions).

다. 문항수가 많지 않더라도 평균점수가 평가도구의 목적에 부합하고, 최적에 가까운 문항 분석지표 범위를 유지하면서 평가도구를 개발하게 되면 문항수의 제한에서 벗어나 그 유효성을 유지할 수 있게 된다[16]. 본 연구에서 전문가들의 인식은 본 연구의 배경인 문제해결문항 비중의 증가, 통합형 문항의 필요성 대두 방향과 일치하게 국가시험 문항수를 줄이는 것이 적절하다고 응답했다. 특히 응급환자관리학과 전문응급처치학각론 과목에서 문항 축소의 필요성이 제기되었으며 전체 문항수 역시 감소요

구를 확인할 수 있었다. 문항수가 불필요하게 많다면 학생들에게 높은 수준의 에너지를 필요로 하며 압박감으로 이어질 수 있다[17]. 연구 [18]에 의하면 문항수가 적을수록 집중력이 향상되고 성적에 긍정적 영향을 미치는 것을 보고하고 있다.

1급 응급구조사 국가시험은 총 230문항으로 60점을 취득하려면 진점수는 138점이 되어야 한다. 최근 경향을 보았을 때 2021년도를 제외하면 난이도는 점차 쉬워지는 경향을 보였으며 합격률은 증가하였다. 고전검사이론의 난이도

는 0 ~ 100의 수치를 가지며 수치가 작을수록 난이도가 어려움을 의미한다. 문항지표들은 서로 상호작용을 하는데 극단적인 난이도 구간을 제외한 구간에서, 난이도가 높을수록 변별력이 증가하는 것으로 알려져 있다[19]. 또한 문항지표가 반영하지 못하는 요인으로, 난이도가 높을수록 불안과 스트레스가 증가하여 성적이 실제 능력보다 낮게 측정될 우려가 있다[19, 20]. 2016년 이후 국가시험의 평균 난이도는 70.3점으로 보통 수준의 난이도를 가진 것으로 나타났다[21]. 또한 신뢰도는 평균 .937를 보였으며, 변별도는 .25 였다. 이는 보건의료 26개 직종 국가시험의 난이도는 61.3 ~ 83.9 범위에 있으며, 변별도는 0.17 ~ 0.38 범위이며, 신뢰도는 0.886 ~ 0.978 사이에 있는 것[22]과 비교하여 1급 응급구조사 국가시험의 경우 비교적 높은 신뢰도와 보통의 난이도와 변별도를 나타내고 있다.

문항과 검사의 질을 분석하는 이론을 검사이론이라 하며, 검사이론에는 고전검사이론(Classical Test Theory, CTT)과 문항반응이론(Item Response Theory, IRT)이 있다. 고전검사이론에 따르면 관찰점수는 진점수와 오차점수에 의하여 합성되었음을 가정하고 총점에 의하여 문항을 분석하고 피험자 능력을 추정하는 검사이론으로 개별문항의 난이도 분포나 동료 피험자의 수준에 큰 영향을 받는다. 반면 문항반응이론은 문항이 불변하는 고유의 속성을 가지고 있다고 가정하고, 이 속성을 나타내는 고유한 문항특성곡선과 그들의 합인 검사특성곡선을 통해 문항과 평가도구를 분석하는 것이다[23]. 문항난이도(Item difficulty)는 고전검사이론에서는 총 피험자 중 답을 맞힌 피험자의 비율인 정답률이 되며 .20 미만이면 매우 어려운 문항에 속하며 .80 이상 ~ 1.00이면 매우 쉬운 문항이 된다. 문항반응이론에서는 50%의 정답률을 보이는 문항특성곡선상의 학습자 능

력이 난이도를 의미하며, 일반적으로 -2에서 +2사이의 값이 존재하며 값이 클수록 그 문항은 어렵다고 해석된다[23, 24]. 문항변별도(Item discrimination)는 고전검사이론에서는 .10미만은 변별력이 없는 문항으로 그리고 .40 이상은 변별력이 높은 문항으로 구분한다. 문항반응이론에서의 변별도는 50% 정답률에 해당하는 문항특성곡선상의 기울기로 0.34 이하이면 변별력이 거의 없고, 0.35 ~ 0.64는 낮음, 0.65 ~ 1.34는 적절, 1.35 이상이면 변별력이 높은 것으로 추정한다[23].

문항반응이론 1모수 모형(Rasch 모형)은 문항변별도와 문항추측도를 고려하지 않고 문항난이도만 고려하는 모형이다. 이 모형에서는 문항변별도는 모든 문항이 동일하고, 문항추측도는 0이라 간주한다. 만약 모든 문항의 문항변별도가 상수로서 동일하다면 1모수 문항반응모형을 선택하여 문항특성곡선을 그리지만 그렇지 않다면 2모수 문항반응모형이나 3모수 문항반응모형을 선택하여야 한다. 문항반응이론 2모수 모형은 문항의 추측도를 고려하지 않고 문항난이도와 문항변별도만 고려하는 문항이다. 즉 모든 문항의 문항추측도는 0이고 각 문항들의 난이도와 변별도가 다를 것이라고 가정하는 것이다. 따라서 모수가 증가할수록 모형이 제공하는 정보가 많아지지만, 대상자료의 수가 충분해야 하며, 모형 적합도 문제가 발생할 가능성이 크다[24, 25]. 본 연구에서 모의시험에 대한 문항반응이론의 모형은 1모수 모형보다 2모수모형이 적합한 것으로 나타났다.

또한 문항반응이론을 전개하기 위해 일차원성 가정(unidimensionality assumption)이 충족되어야 한다. 일차원성 가정은 검사도구는 하나의 특성을 측정해야 함을 전제로 한다[24]. 즉 하나의 검사는 피험자가 지니고 있는 하나의 특성을 측정해야 문항반응이론을 적용하여 문항 분석뿐 아니라 피험자의 능력까지 정확하게

추정할 수 있다는 것이다. 이러한 가정에서 응급의료관련법령은 응급구조사의 역량을 측정하는 다른 과목과 비교하여 개념적으로 이 가정을 위배한다고 판단하였으며, 실제 예비 모형 분석에서도 응급의료관련법령을 포함한 모형은 유의하였으나, 적합도가 낮게 나왔다. 본 연구에서는 일차원성 가정을 검증하기 위해 요인분석을 사용하였다. 의사국가시험 자료를 분석한 연구에서 일차원성 가정을 검증하기 위한 요인 분석에서 고유값은 문항수만큼 계산된 것을 볼 수 있다. 요인별 고유값의 차이로 제1요인과 제2요인의 차이가 다른 요인간의 차이에 비해 큰 차이를 보여 일차원성 가정을 충족한 것으로 추정하고 있다[26].

문항을 순차적으로 제거한 결과 난이도는 3차 시험 이후 변화의 경향을 보이지 않았으나, 변별도와 신뢰도는 문항수 감소에 따라 감소하는 경향을 보였다. 7개 시험에서 모두 적절 변별도의 기준 0.65 ~ 1.34에 포함되어 있었고, 난이도는 쉬움에 속하는 -2.0 ~ -0.5에 포함되어 있었으며, 고부담 시험의 신뢰도 기준인 0.9 이상이었다. 그러나 다소 쉬운 난이도의 시험이기 때문에 -0.7 이하의 감소를 보이지 않고, 변별도는 0.7 이상 그리고 신뢰도는 0.91이상으로 높게 유지되는 5차 시험을 문항수 제거로 문항평가지표의 손실을 보지 않는 시뮬레이션 시험이라 판단하였다. 5차 시험까지 평균점수와 문항수의 상관관계는 없었다. 그러나 적합도가 낮은 문항 중 난이도가 높은 문항이 있었고, 제거 후 난이도가 3차 시험 이후 낮아지는 점, 175문항으로 구성된 7차 시험에서도 신뢰도는 0.9 이상으로 유지되는 점, 낮은 역량수준을 잘 반영한 시험이었다는 점, 국가시험은 이보다 많은 학생이 응시하고 분산이 커져 신뢰도가 높아질 수 있다는 점 등을 고려할 때 175문항 또는 그 이하의 문항에서도 문항평가지표의 손실을 가져오지 않고 문항을 구성할 수 있는 가

능성이 있다고 판단된다.

본 연구는 1급 응급구조사 국가시험의 문항수 조절과 시험의 양호도 판단을 위해 문항반응이론을 처음 적용한 시도라는데 의의가 있으며, 모의시험에서 185문항 수준으로 문항수를 줄이는 것이 시험의 난이도와 변별력, 신뢰도의 큰 손실없이 평균점수를 유지할 수 있다는 것을 확인하였다. 본 연구결과는 향후 응급구조사 1급 국가시험의 신뢰도와 타당도가 높은 적정 문항수를 산정하는데 도움이 되는 기초자료를 제공할 것이다.

연구의 제한점은 첫째, 응급구조학과 교수들의 의견을 온라인 설문으로 수집한 것이며, 개선의 의사가 강한 전문가들이 주로 응답을 한 선택 편견(selection bias)의 가능성이 있다. 실제 조절은 응급구조사 역량과 교과 구조의 요구에 따른 타당성 기준이 그 시작점이므로 보다 적절한 타당성 분석 연구가 필요하다. 둘째, 실제 국가시험 결과가 아니라 모의시험의 결과를 연구 대상으로 하였고, 역량 수준을 분석하는데 유용한 검사특성곡선은 합격선인 진점수 60점보다 낮은 역량의 수험생을 더 잘 반영하고 있었으며, 합격선 점수에서 제공되는 정보의 양이 부족한 편이었다. 합격선 점수에서 많은 정보를 제공하는 시험이 된다면 보다 합리적인 판단기준이 제공될 것으로 사료된다.

V. 결 론

국가시험 문항 축소 가능성을 알아보기 위한 예비조사로, 응급구조학과 교수들은 전체 문항수의 축소가 적절하다고 판단하였으며, 모의시험 결과를 기반으로 시행한 전체 문항의 축소 시뮬레이션 결과 평균점수, 난이도, 변별도, 신뢰도의 큰 손실은 없었다. 국가시험 출제영역에

대한 타당도 분석에 기반하고 실제 국가시험을 통해 분석한다면 더욱 신뢰할 수 있는 결과를 얻을 수 있다고 본다.

ORCID ID

Jung Eun Lee: 연구설계, 연구방법, 고찰

0000-0002-5805-9158

Jundong Moon: 연구설계, 서론, 연구결과, 교신저자

0000-0003-4742-8744

Ajung Kim: 연구설계, 연구결과, 고찰

0000-0002-5468-9777

References

- Emergency medical service act. Available at <https://www.law.go.kr/LSW/eng/engLsSc.do?menuId=2§ion=lawNm&query=emergency&x=47&y=26#liBgcolor4>
- Korea health personnel incensing examination institute. National exam analysis results. Available at https://www.kuksiwon.or.kr/analysis/brd/m_91/list.do
- Korea health personnel incensing examination institute. Health professional categories. Available at https://www.kuksiwon.or.kr/cnt/c_1/view.do?seq=7
- Lee SE, Park YJ, Lee DJ. Curriculum and educational evaluation. KNOU press, 2022. 290-341.
- Yim CI, Yim JH, Lee DJ. Educational technology. KNOU press, 2022. 249-50.
- Kim MY. Development of practice-based integrated question type for national nursing examination. Korea Health Personnel Licensing Examination institute, 2023.
- Sung TJ, Kim KH. Changes in reliability coefficient and test information function according to changes in number of items, item difficulty, and item discrimination. *J Curric Eval* 1993;6(2):123-54.
- Yim MK, Huh S. Testing unidimensionality and goodness-of-fitness for the application of item response theory to the Korean Medical Licensing Examination. *Korean J Med Educ* 2007;19(2):163-9. <http://doi.org/10.3946/kjme.2007.19.2.163>
- Jeong GH, Yim MK. Applicability of item response theory to the Korean Nurses' Licensing Examination. *J Educ Eval Health Prof* 2005;22(1):23-9. <http://doi.org/10.3352/jeehp.2005.2.1.23>
- Kim AJ, Park TJ, Bak YS, Kim JH, Kim YS, Son YM et al. Review of the qualification criteria for the national examinations for emergency medical technicians. *Korean J Emerg Med Ser* 2021;25(2):39-53. <http://doi.org/10.14408/KJEMS.2021.25.2.039>
- Kim MY, Lee YH, Huh S. Correlations between the scores of computerized adaptive testing, paper and pencil tests, and the Korean Medical Licensing Examination. *J Educ Eval Health Prof* 2005;2(1):113-8. <http://doi.org/10.3352/jeehp.2005.2.1.113>
- Yoou SK, Jo JM, Kim JH, Uhm TH. Proposal of a new EMT national registry exam using the Delphi method. *Korean J Emerg Med Ser* 2010;14(2):57-69.

13. Youu SK, Choi ES, Shin DM, Jo JM, Lee JE, Roh SG et al. Proposal of a new emergency medical technician national practical examination. *Korean J Emerg Med Ser* 2012;16(3):45-62 (Korean).
<http://doi.org/10.14408/KJEMS.2012.16.3.045>
14. Park HG, Kim DW, Kang YJ, Bae SY. Analysis of the appropriate number of questions that satisfy the validity and confidence level of the test. *Korea Health Personnel Licensing Examination Institute*, 2022.
15. Kim YH, Hang MJ. Trend analysis of the results of the national examination for EMT-paramedics for the Past 9 Years. *Journal of Health Informatics and Statistics* 2022;47(2):133-8.
<http://doi.org/10.21032/jhis.2022.47.2.133>
16. Brogden HE. Variation in test validity with variation in the distribution of item difficulties, number of items, and degree of their intercorrelation. *Psychometrika* 1946;11(4):197-214.
17. Baiju RM. Academic stress and psychological well-being among college students. Available at: <https://ijip.in/wp-content/uploads/2021/08/18.01.022.0210903.pdf>. Accessed May 18, 2024.
18. Suruchi S, Rana SS. Test item analysis and relationship between difficulty level and discrimination index of test items in an achievement test in biology. *PIJR*. 2014;3(6):56-8.
19. Smoleń T, Chuderski A. The quadratic relationship between difficulty of intelligence test items and their correlations with working memory. *Front Psychol*. 2015;6:156499.
20. Stankovska G, Dimitrovski D, Angelkoska S, Ibraimi Z, Uka V. Emotional intelligence, test anxiety and academic stress among university students. *Bulgarian Comparative Education Society*. 2018. Available at: <http://files.eric.ed.gov/fulltext/ED586176.pdf>. Accessed May 18, 2024.
21. Kaur M, Singla S, Mahajan R. Item analysis of in use multiple choice questions in pharmacology. *Int J Appl Basic Med Res*. 2016;6(3):170-3.
<https://doi.org/10.4103/2229-516x.186965>
22. Kim MY. Development of practice-based integrated question type for national nursing examination. *Korea Health Personnel Licensing Examination Institute*. 2023.
23. Seong TJ. Educational evaluation. Seoul: Haksisa, 2019. 259-64.
24. Seong TJ. Understanding and applying item response theory. Paju: Kyoyookbook 2019. 30-4, 56-7.
25. Columbia University Mailman School of Public Health. Item response theory. Available at: <https://www.publichealth.columbia.edu/research/population-health-methods/item-response-theory>. Accessed May 18, 2024.
26. Lim EY, Park JH, Kwon I, Song GL, Huh S. Comparison of item analysis results of Korean Medical Licensing Examination according to classical test theory and item response theory. *J Educ Eval Health Prof* 2004;1(1):67-76.
<http://doi.org/10.3352/jeehp.2004.1.1.67>