



계명대학교학술지

Keimyung Medical Journal

Original Article

pISSN 2092-8335 • eISSN 2733-5380
Keimyung Med J 2023;42(1):7-18
<https://doi.org/10.46308/kmj.2023.00094>

Received: April 10, 2023
Revised: April 28, 2023
Accepted: May 8, 2023

Corresponding Author:

Han Chae, PhD
School of Korean Medicine, Pusan National University, 49, Busandaehak-ro, Mulgeum-eup, Yongsan-si, Gyeongsangnam-do, 50612, Korea
Tel: +82-51-510-8470
Fax: +82-51-510-8470
Email: han@chaelab.org

Seul Lee, PhD
Hamsa Korean Medicine Clinic, Eastern 5th floor, 299 Omok-ro, Yangcheon-gu, Seoul 08001, Korea
Tel: +82-10-3077-0406
Fax: +82-10-3077-0406
Email: dewclear46@gmail.com

© 2023 Keimyung University School of Medicine
© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

<http://www.e-kmj.org>

문항반응이론을 사용한 한의사 국가시험 모의고사의 타당도와 학업 역량 분석

채한¹, 조은별², 김선경³, 최다혜⁴, 이슬⁵

¹부산대학교 한의학과
²한국한의학연구원 한의과학연구부
³동의대학교 한의학과 의학교육학교실
⁴부산대학교 한의학교육실
⁵함소아한의원

Analysis on Validity and Academic Competency of Mock Test for Korean Medicine National Licensing Examination Using Item Response Theory

Han Chae¹, Eunbyul Cho², SeonKyoung Kim³, DaHye Choi⁴, Seul Lee⁵

¹School of Korean Medicine, Pusan National University, Busan, Korea
²KM Science Research Division, Korea Institute of Oriental Medicine, Daejeon, Korea
³Department of Medical Education, College of Korean Medicine, Dong-Eui University, Busan, Korea
⁴Office of Korean Medicine Education, Pusan National University, Busan, Korea
⁵Hamsa Korean Medicine Clinic, Seoul, Korea

The national licensing examination is used to evaluate the medical competency at the time of graduation, however no study has been performed on the validity of traditional Korean medicine license examination yet. The purpose of this study was to develop learning analytics using item response theory (IRT) to examine the validity and academic competency of the mock test of the national licensing exam. Classical test theory and IRT were used to evaluate the validity of test items, and IRT was used for test validity and competency analysis. The correlation between competency score of 12 subjects was analyzed using Pearson's correlation. The distribution of students' latent competencies was examined by gender and administrative group using a Kernel density map, Latent Profile Analysis, and χ^2 . The guessing parameter of 340 items was relatively high, and the information level of 12 subjects were relatively low. Significant correlations ($r = 0.49-0.83, p < 0.05$) were observed between the competency scores of total and 12 subjects. Two (high and low) latent academic competency groups were identified based on the competency score of 12 subjects. The low academic competency group requiring intensive management has a significantly higher frequency of male students with the experience of academic fail in the seven-year course. This study presented the quantitative learning analytics for the national licensing exam of traditional Korean medicine. The multifaceted item and test validities of the mock license test were provided, and an evidence-based approach to competency-based student management and national licensing exam of traditional Korean medicine was suggested.

Keywords: Competency-focused education, Evidence-based teaching, Item response theory, Learning analytics, National license exam

Introduction

보건의료인 국가시험은 지난 4-7년간의 대학 학업을 통해 준비한 의료 전문가로서의 역량을 최종적으로 확인하는 고부담 시험(high-stakes testing)으로, 교수-학습 측면에서 볼 때 수십여명의 교수자들이 운영한 교육과정의 효율성에 대한 점검인 동시에 교육과정을 통해 성취하려는 의료역량의 구체화된 최종 목표라 할 수 있다[1].

이러한 국가시험의 결과는 대한민국 정부가 특정 의료 직종에서의 독점적인 자격을 허가한다는 면허(免許 또는 license)로 연결된다[2]. 면허의 기준이 되는 의료인의 역량이란 '잠재되어 밖으로 잘 드러나지 않는 개인 고유의 특성'에 해당하기에, 이를 객관적으로 측정한다는 것은 매우 어려우며 높은 수준의 측정학적 전문성이 요구된다[3]. 이에, 의료인에게 요구되는 윤리적인 태도, 임상적 기술과 지적 역량의 공정하고 안정적인 측정을 위하여, 대한민국에서는 한국보건의료인국가시험원(국시원)이라는 전문기관이 30여 직종에서의 국가시험을 주관하고 있다[4].

한의사 국가시험은 1952년 95명의 합격자(한의사) 배출에서 시작되었으며, 주관기관 또한 보건복지부로부터 시작하여 국립보건원을 거쳐 1998년부터 국시원으로 바뀌어 왔다. 2022년 시행되어 731명(97.1%)가 합격한 제77회 한의사국가시험은 기존의 지필형 시험과는 다르게 컴퓨터시험(computer-based Test, CBT)을 처음으로 활용하였는데, 한의학 지식에 대한 11과목 340문항(객관식 5지선다형)을 대상으로 전국 7개 시험센터에서 점심식사 1시간을 포함하여 7시간45분동안 진행되었다[4].

그러나, 한의사 국가시험의 오랜 역사에도 불구하고 이에 대한 연구는 만족스럽지 못하다. 의사 국가시험과 관련하여 필기시험의 시행 방법[5,6], 문항과 시험의 타당도 및 신뢰도 개선[7,8], 국가시험 분석 방법론[9], 대학에서의 학업성취도[10]나 임상술기시험, 학업성적 및 졸업시험 역량과의 상관성[11-13], 심리적 특성과의 상관성[14], 합격선의 설정[1,2], 새로운 시험 방법의 도입[15-17] 등 다방면의 연구가 지속되어 온 것과는 대조적으로, 한의사 국가시험의 타당도나 역량분석 등에 대한 분석은 찾아보기 어렵다[3,18]. 이와 함께, 임상적 기술에 대한 실기시험은 수년내 도입을 목표로 한 논의를 시작하는 단계이며, 직업적인 윤리, 인성, 환자에 대한 태도와 프로페셔널리즘 같은 윤리적 태도는 중요성이 제시되는 단계에 있다[19].

이는 한의학 교육 및 교육 공학이 다른 현안에 비하여 우선 순위가 낮았던 것도 원인이기는 하지만, 교육 측정학적 분석방법에 대한 교수자들의 낮은 관심과 성취역량을 분석하는 교육지원 환경을 구축하기 위한 준비의 부족이 새로운 학업분석(learning analytics)의 도입에 장애가 되어왔다. 이와 함께, 한의학 교육의 역량중심 재편과정에서도 측정과 분석을 전제로 하는 총체적 역량(competency)보다는 분절된 단위 능력(ability) 단위로서의 역량(competence)

을 나열하는 것에만 초점이 놓여 있었다[20]. 이에 따라, 한의학 교수자들이 국가시험의 미래지향적 변화나 한의과대학 교육과정 개선에 만족스러운 근거와 논리를 제시하지 못하였으며, 학업역량 증진에 대한 충분한 검토도 부족하였다.

이에, 본 연구에서는 본초학 시험 및 문항, 역량의 분석에 문항반응이론(item response theory, IRT)을 적용하였던 선행연구[3,18]를 토대로 한의사 국가시험에 대한 구체적인 연구방법론을 제시하고, 이를 한의학과에서 시행한 모의고사에 적용하여 교육 현장에서의 효용성을 살펴보고자 하였다.

IRT는 잠재되어 있는 개인 고유의 역량을 분석할 수 있는 문항 및 시험 타당도 분석방법으로[3,18], 의사 국가시험을 비롯한 국내 보건의료계가 준비하는 컴퓨터 적응시험(computer adaptive test, CAT)의 측정분석학적 기반 이론이다[21]. IRT는 이론과 계산 방법이 어려워 교육현장에서의 사용이 최근까지 매우 제한적이었으나, 정보통신기술의 급격한 발전에 따라 기존의 고전검사이론(classical test theory, CTT)를 대신하여 다양한 학업분석에 적극적으로 도입되고 있다[21-23]. 아울러, 본 연구의 대상으로 활용된 국가시험 모의고사는 한의대를 비롯 많은 의대에서 4-7년간의 학업을 매듭짓는 졸업자격 시험 또는 기말고사 대체로 시행되고 있으므로[12], 국가시험에 적용할 학업분석 방법론의 활용성을 검토하기에 가장 적절하다고 여겨진다.

본 연구에서 제시하는 한의사 국가시험의 타당도 및 역량 분석법이 학업분석의 성취수준 확인을 위한 기본적인 방법으로 활용된다면, 향후 축적될 시계열 데이터는 한의사 국가시험 관련 제도 및 정책의 개선에 보다 능동적인 개선을 이끌어낼 수 있을 것이다. 이와 함께, 재학생들의 학업역량에 대한 다면적인 분석은 학생들의 수준과 진도에 따른 맞춤 학업지도 및 관리를 위한 구체적인 방안을 제시하여, 학습 효율성과 학생 만족도가 높은 한의학 교육을 현실화할 수 있을 것이다. 이는 한의학 교육을 근거기반 역량중심 의학교육(evidence-based competency focused medical education) 발전의 초석이 될 것이다[18,24,25].

Materials and methods

1. 연구 대상 및 절차

2019년에 시행된 1차 모의고사(12과목 340문항)에 대한 40명의 답안을 연구 대상으로 활용하였다. 본 연구는 부산대학교 생명윤리 위원회의 심의(PNU IRB/2022_75_HR) 이후 한의학교육실에서 데이터를 받아 분석을 진행하였다.

2. 한의사 국가시험 모의고사

한의사 국가시험은 내과학1(80문항) 내과학2(32문항), 침구학(48문항), 보건의약관계법규(20문항), 외과학(16문항), 신경정신과학(16문항), 안이비인후과학(16문항), 부인과학(32문항), 소아과학

(24문항), 예방의학(24문항), 한방생리학(16문항), 본초학(16문항)의 12과목으로 5지선다형(A형, 긍정형) 340문항(문항당 1점)으로 구성되어 있다. 국가시험의 최종 합격을 위해서는 340문항 전체에서 60%(204점)이상 득점과 함께 6개 묶음에 대하여 40% 이상 득점(면과락)하여야 한다. 6개의 묶음(면과락 점수/문항 개수)은, 내과학1과 내과학2(45점/112문항), 침구학(20점/48문항), 보건의약관계법규(8점/20문항), 외과학, 신경정신과학과 안이비인후과학(23점/48문항), 부인과학과 소아과학의 합계(23점/56문항), 예방의학, 한방생리학과 본초학(23점/56문항)이다.

한의사 국가시험 모의고사는 국시원에서 시행하는 한의사 국가시험을 대비하기 위한 것으로, 한의학과 졸업대체 기준에 따라 교수가 출제하고 교육실이 관리하여 매년 4학년 2학기에 2번 시행하며 졸업 사정에 그 결과를 반영한다.

3. 통계 분석

본 연구에 활용된 모의고사 340문항의 타당도는 CTT와 IRT를 사용하여 계산하였다. CTT에서의 문항난이도 또는 정답율은 문제를 맞출 확률(0-1.0)이며, 문항변별도는 문항점수와 과목 총점간의 상관계수(-1.0-1.0)를 말한다. IRT에서는 문항특성곡선(item character curve, ICC)을 사용하여 타당도를 제시하는데, 문항난이도(b)는 답을 맞출 확률로서 ICC 확률축 중간에 해당되는 ICC 능력축의 값(-4.0-4.0)을 의미하며, 문항변별도(a)는 문항이 역량수준을 변별할 수 있는 능력으로서 ICC 확률축 중간에 해당되는 지점에서의 기울기(0-2.0)이고, 문항추측도(c)는 제일 낮은 역량에서 답을 맞출 확률로서 ICC 확률축의 절편값(0-1.0)이다[3,18]. 12 과목에 포함된 문항들의 타당도는 과목별 평균과 95% 신뢰구간으로 제시되었다.

12개 과목에서의 타당도 분포는 IRT 타당도 지표와 국시원의 분류 기준을 사용하여 빈도로 제시하였다. 문항 난이도(b)가 -2.0 이하이면 '매우 쉽다', -2.0-0.5는 '쉽다', -0.5-0.5는 '중간이다' 또는 '보통이다', 0.5-2.0는 '어렵다', 2.0 이상이면 '매우 어렵다'고 해석하였으며, 문항 변별도(a)가 0.34 이하이면 변별력이 '거의 없다', 0.35-0.64는 '낮다', 0.65-1.34는 '적절하다', 1.35-1.69는 '높다', 1.70 이상이면 '매우 높다'로 해석하였으며, 문항 추측도(c)가 0-0.2 사이를 적절한 영역, 0.2-0.3 사이를 경계 영역으로, 0.3이상을 과도하게 높은 영역으로 구분하였다[3,18,26,27].

모의고사(340문항)와 12과목에 있어서 시험의 타당도는 역량 수준(-4-4)별 진점수(true score)를 의미하는 시험특성곡선(test character curve, TCC)과 시험이 역량 수준(-4-4)별로 제공하는 정보의 양을 의미하는 시험정보곡선(test information curve, TIC)으로 제시하였다. TIC는 시험 또는 과목에 포함된 문항들의 문항정보곡선(item information curve, IIC)을 모두 합한 값으로, IIC는 문항의 능력추정 과정에서 확인되는 역량수준(-4-4)별 분산 값의 역수이다[3,18].

학생 개개인의 학업 역량 추정에는 3-모수 모형(3-parameter logistic model)과 사후기대법(expected a posteriori, EAP)을 사용하였으며, 모의고사와 12과목에 있어서의 학생 개개인의 학업 역량도 함께 추정되었다. 모의고사와 과목의 학업 역량간 상관성을 분석함에 있어서는 Pearson's correlation을 사용하였다.

역량수준별 학생의 분포는 커널 밀도지도(kernel density map)를 사용하였는데, 이는 역량수준별 빈도를 직관적으로 표시하는 비모수 밀도 함수로서, 직관적인 표현을 위해 평활화(gaussian smoothing)를 사용하여 곡면을 매끄럽게 하였다[3,18]. 이와 함께, 역량수준별 학생 분포를 성별, 학사행정 그룹으로 나누어 제시하였는데, 학사행정을 기준으로 한 집단으로는 유급과 휴학을 경험한 학생들, 학석사 과정(7년, 학사 3년과 석사 4년) 재학생, 석사 과정(석사 4년) 재학생의 3개 그룹으로 나누어 제시하였다.

12개 과목별 역량을 기준으로 잠재되어 있는 학생 그룹을 확인하기 위하여 잠재프로파일분석(latent profile analysis, LPA)을 시행하였다. 잠재되어 있는 프로파일 그룹의 개수는 모델 적합도 통계량의 분석을 통해 결정되는데, 개수가 증가함에 따라 변화하는 프로파일 모델의 적합도 지수가 더 이상 유의미하게 개선되지 않을 때를 가장 적절한 모델이 확인된 것으로 하였다. 본 연구에서는 프로파일 모델의 적합도 지수의 해석에 있어서, 1) Akaike information criterion(AIC), consistent Akaike information criterion(CAIC), Bayesian information criterion(BIC), sample size adjusted BIC(SABIC)가 작을수록 유의하며, 2) p-value for bootstrapped likelihood test(BLRT-p)가 유의($p < 0.05$)하며, 3) entropy 지수가 0.8보다 큰 것을 유의한 것으로 보았다. 역량 프로파일의 성별(남-여), 학사행정(유급-학석사[7년]-석사[4년])별 분포의 차이는 χ^2 를 사용하여 분석하였다.

문항 및 시험 분석에는 jMetrik 4.1.1(J. Patrick Meyer, Charlottesville, VA, USA)을 사용하였으며[28], 그 외의 분석에는 jamovi 2.3.24(The jamovi project, <https://www.jamovi.org>)를 사용하였다. 통계 분석의 결과는 평균±표준오차, 95%신뢰구간(상한값, 하한값) 또는 빈도(%)로 표기하였으며, 통계적 유의수준으로는 $p < 0.05$, $p < 0.01$ 및 $p < 0.001$ 을 사용하였다.

Results

1. 문항 분석

모의고사 340문항을 대상으로 CTT와 IRT 타당도 분석을 시행한 결과는 Table 1과 같다. 통상적으로 제시되는 국시원의 시험결과 분석에는 문항별 분석을 상세하게 제시하나, 본 연구에서는 지면상의 제한으로 12개 과목별 평균과 95% 신뢰구간(하한값, 상한값)을 제시하였으며, 국시원의 타당도 분류 기준을 사용한 빈도분석을 진행하였다.

분석 결과를 살펴보면, 가장 어려운 과목은 내과학2였으며, 가장

Table 1. Calculated item parameters of each subjects using CTT and IRT

Subject (items)	CTT		IRT																	
	Disc	Diff	Diff	Disc	Diff	Diff	Guss	Disc	VL	Lo	Ad	Hi	VH	VE	Ea	Mi	Di	VD	Ad	Mg
IM1 (80)	0.19, [0.16, 0.23]	0.67, [0.61, 0.73]	0.89, [0.80, 0.99]	-1.65, [-2.37, -0.93]	0.44, [0.41, 0.48]	9	13	48	8	2	36	19	8	7	10	7	16	57		
IM2 (32)	0.26, [0.18, 0.33]	0.50, [0.41, 0.59]	0.97, [0.86, 1.09]	0.94, [-0.79, 2.66]	0.41, [0.35, 0.46]	2	2	24	4	0	9	8	4	1	10	6	2	24		
AM (48)	0.14, [0.10, 0.18]	0.76, [0.68, 0.84]	1.01, [0.92, 1.11]	-1.94, [-2.96, -0.92]	0.48, [0.41, 0.54]	1	5	33	8	1	28	7	2	4	7	3	11	34		
PL (20)	0.23, [0.15, 0.30]	0.67, [0.57, 0.76]	0.85, [0.66, 1.05]	-1.05, [-2.00, -0.10]	0.48, [0.41, 0.56]	1	8	9	2	0	7	9	2	0	2	2	1	17		
DS (16)	0.32, [0.25, 0.39]	0.73, [0.63, 0.82]	0.93, [0.70, 1.17]	-2.15, [-3.05, -1.24]	0.52, [0.49, 0.54]	2	3	8	3	0	8	6	0	2	0	0	0	16		
NP (16)	0.13, [0.04, 0.21]	0.60, [0.47, 0.72]	0.87, [0.69, 1.06]	-0.39, [-1.87, 1.09]	0.46, [0.37, 0.56]	1	4	9	2	0	4	5	2	2	3	1	4	11		
OO (16)	0.19, [0.11, 0.26]	0.67, [0.56, 0.77]	0.77, [0.61, 0.93]	-0.81, [-2.29, 0.66]	0.51, [0.43, 0.58]	3	0	12	1	0	6	4	1	1	4	0	2	14		
GY (32)	0.20, [0.15, 0.25]	0.72, [0.65, 0.79]	0.86, [0.71, 1.00]	-1.64, [-2.42, -0.85]	0.52, [0.46, 0.57]	4	3	20	5	0	12	11	4	4	1	1	3	28		
Pd (24)	0.15, [0.08, 0.23]	0.69, [0.58, 0.80]	0.86, [0.73, 0.99]	-1.22, [-2.58, 0.15]	0.48, [0.41, 0.56]	2	4	16	2	0	11	6	2	0	5	2	3	19		
Pv (24)	0.17, [0.10, 0.24]	0.66, [0.54, 0.77]	0.85, [0.66, 1.04]	-1.09, [-2.44, 0.25]	0.46, [0.41, 0.51]	3	6	10	3	2	11	4	2	2	5	2	1	21		
Ph (16)	0.20, [0.10, 0.29]	0.61, [0.49, 0.72]	0.77, [0.59, 0.95]	-0.94, [-2.34, 0.47]	0.42, [0.34, 0.49]	1	6	7	2	0	4	4	5	1	2	2	3	11		
Hb (16)	0.25, [0.19, 0.31]	0.65, [0.56, 0.74]	0.95, [0.84, 1.07]	-1.49, [-2.13, -0.85]	0.52, [0.46, 0.58]	0	2	13	1	0	3	11	2	0	0	1	0	15		

Data shown as mean, [95% CI]. CTT, classical test theory; IRT, item response theory; Disc, discrimination; Diff, difficulty; Guss, guessing; IM1, internal medicine 1; IM2, internal medicine 2; AM, acupuncture medicine; PL, public health & medicine related laws; DS, dermatology & surgery; NP, neuropsychiatry; OO, ophthalmology & otorhinolaryngology; GY, gynecology, Pd, pediatrics; Pv, preventive medicine; Ph, physiology; Hb, herbiology. VL, very low discrimination (<0.35); Lo, low discrimination (0.35-0.64); Ad, adequate discrimination (0.65-1.34); Hi, high discrimination (1.35-1.690); VH, very high discrimination (≥1.70); VE, very easy difficulty (<-2.0); Ea, easy difficulty (-2.0--0.5); Mi, middle difficulty (-0.5-2.0); VD, very difficult (>2.0); Ad, adequate guessing (0-0.2); Mg, marginal guessing (0.2-0.3); VH, very high guessing (≥0.3).

쉬운 과목은 외과학으로 확인되었다. 적절한 IRT 타당도를 보인 과목도 확인할 수 있었는데(Fig. 1), 적절한 난이도(-0.5-0.5)의 비율은 예방의학에서 가장 낮았으며 본초학에서 가장 높았고, 적절한 변별도(0.65-1.34)의 비율은 외과학에서 가장 낮았고 한방생리학에서 가장 높았으며, 적절한 추측도(0-0.2)의 비율은 외과학과 부인과학에서 가장 낮았으며 내과학2에서 가장 높았다. 12과목이 전반적으로 적절한 변별도를 지닌 것을 확인할 수 있었으며, 높은 추측도로 인한 매우 낮은 난이도가 특징적이었다.

2. 시험 분석

12개 과목에서의 역량 수준별 진점수와 정보의 양을 IRT를 사용하여 확인하였으며, 역량수준별 진점수를 표현하는 TCC와 역량수준별 정보의 양을 제시하는 TIC을 제시하였다(Fig. 2). 모의고사(12과목 전체)를 볼 때 평탄하게 낮은 TIC 곡선과 직선에 가까운 TCC 곡선을 확인할 수 있었다. 내과학, 예방의학, 생리학, 본초학 등에서는 역량수준 0--3사이에서 높아지는 시험정보의 양을 확인할 수 있었으나, 내과학1이나 부인과 등에서는 평탄하게 낮은 시험정보의 양을 확인할 수 있었다.

3. 12개 과목 역량간 상관성

12개 과목간 역량의 상관성을 피어슨 상관계수를 제시하였으며(Table 2), 총점과 12개 과목간에 유의한 상관성($r = 0.49-0.83, p < 0.01$)을 확인할 수 있었다. 이는 총점으로 대표되는 한의사로의 전반적인 역량과 하위 12개의 과목이 모두 유의미한 관련성을 지니고 있음을 의미한다.

12개 과목별로 타 과목과의 상관성이 상이한 것을 확인할 수 있었는데, 내과1(9개), 내과2(8개), 부인과학(7개)은 다른 과목과의 유의한 상관성을 많이 보였으나, 보건의법(2개)을 비롯해 침구의학(3개), 안이비인후과학(3개), 소아과학(3개), 예방의학(3개)은 유의한 상관성을 적게 보였다.

과락 과목들 사이의 상관성을 분석한 결과, 내과학1과 내과학2($r = 0.55, p < 0.001$), 외과학, 신경정신과학 및 안이비인후과학($r = 0.41 [p < 0.01], 0.27$ 및 0.29), 부인과학과 소아과학($r = 0.44, p < 0.01$), 예방의학, 한방생리학 및 본초학($r = 0.35 [p < 0.05], 0.29$ 및 0.27) 등 과목 과락 안에서 단위 과목간 상관성은 내과학을 제외하고는 높지 않은 것을 확인할 수 있었다.

4. 성별 및 학사행정에 따른 학생들의 역량 수준

성별과 학사행정 그룹에 따른 학생들의 역량수준을 분석한 결과(Fig. 3), 남녀의 분포에 확연한 차이를 확인할 수 있었다. 전체 학생들의 역량수준에 있어서 두 개의 봉우리가 확인되었는데, 각각 남자와 여자 또는 유급/휴학과 정상적으로 진급한 그룹을 반영하는 것으로 보인다. 분석 결과 두 가지 흥미로운 결과를 확인할 수 있었는데, 첫째, 높은 역량수준을 가진 학생들이 학-석사 과정에

비하여 석사 과정에 더 많이 존재하며, 둘째, 유급/휴학군에서도 학-석사 과정과 유사한 역량수준을 보이는 학생들이 상당수 존재하고 있었다.

5. 잠재 학업역량 프로파일의 추출과 분석

잠재되어 있는 학업역량 그룹을 추출하기 위하여 12개 과목의 역량점수를 활용한 잠재 프로파일을 적합도를 확인한 결과(Table 3), 최종 2개(높은 성취도와 낮은 성취도 그룹)의 잠재 그룹(Fig. 4)을 제시한 두 번째 프로파일 모델이 가장 만족스러운 타당도를 지니고 있었다. 두 번째 모델은 AIC, BIC가 낮았으며, entropy가 0.9 이상을 보였으며, BLRT-p가 더 유의하였다(Table 3). 제시된 두 잠재 그룹에 있어서, 높은 성취도 그룹은 12개 과목 모두에서 높은 역량수준을 지니고 있었으며, 낮은 성취도 그룹은 12개 과목 모두에서 낮은 역량수준을 지니고 있었다.

성별, 학사행정 그룹별로 성취도 그룹의 분포를 분석한 결과(Table 4), 학-석사(7년) 과정 및 유급/휴학 경험자에 있어서 남녀 간의 성취도 그룹 분포에 큰 차이가 나타났으나, 흥미롭게도 석사(4년) 과정에 있어서는 유의한 차이가 없었다. 높은 성취도 그룹에서는 유의한($\chi^2 = 8.97, p = 0.011$) 성별 차이가 확인되었으며, 낮은 성취도 그룹에서는 여학생의 빈도가 0이어서 통계 자체를 시행할 수 없었다.

Discussion

본 연구에서는 한의사 국가시험 모의고사를 대상으로 문항 및 시험의 타당도와 학생들의 역량에 대한 체계적 분석을 시행하였다. 문항 및 시험의 타당도에 있어서 문항 변별도는 적절하였으나 높은 추측도와 낮은 난이도를 확인할 수 있었으며(Fig. 1), 모의고사의 역량수준별 정보량은 높지 못하였다(Fig. 2). 모의고사에서 확인된 학생들의 학업역량을 분석한 결과, 남학생 특히 학-석사(7년) 과정 및 휴학-유급 경험자의 낮은 학업성취도를 확인할 수 있었다(Fig. 3, Table 4).

본 연구에서는 한의사 국가시험의 타당도를 분석하기 위한 IRT의 체계적 활용 방법이 처음으로 제시되었는데, 기존에 보고되지 않았던 다음과 같은 새로운 지견이 확인되었다. 첫째, 본 연구를 통해 두가지 직관적인 문항 타당도 분석법, 즉 평균과 95%CI를 사용한 평가(Table 1)와 12개 과목별 타당도 수준의 빈도라는 평가(Fig. 1)가 제시되었으며, 모의고사에 있어서 추측도의 개선이 시급함이 확인되었다[9]. 본 연구에서의 모의고사는 재학생들의 국가시험 준비 정도를 점검, 평가하는 것 보다는 공부를 촉진한다는 목적에 더 적합했던 것으로 보이는데, 이는 한의대에서 중간 시험[3,18]이나 의과대학에서의 종합시험[21]보다 낮은 난이도와 낮은 정보수준을 보였기 때문이다.

기존의 IRT를 사용한 문항 타당도 분석에는 통상적으로 ICC가

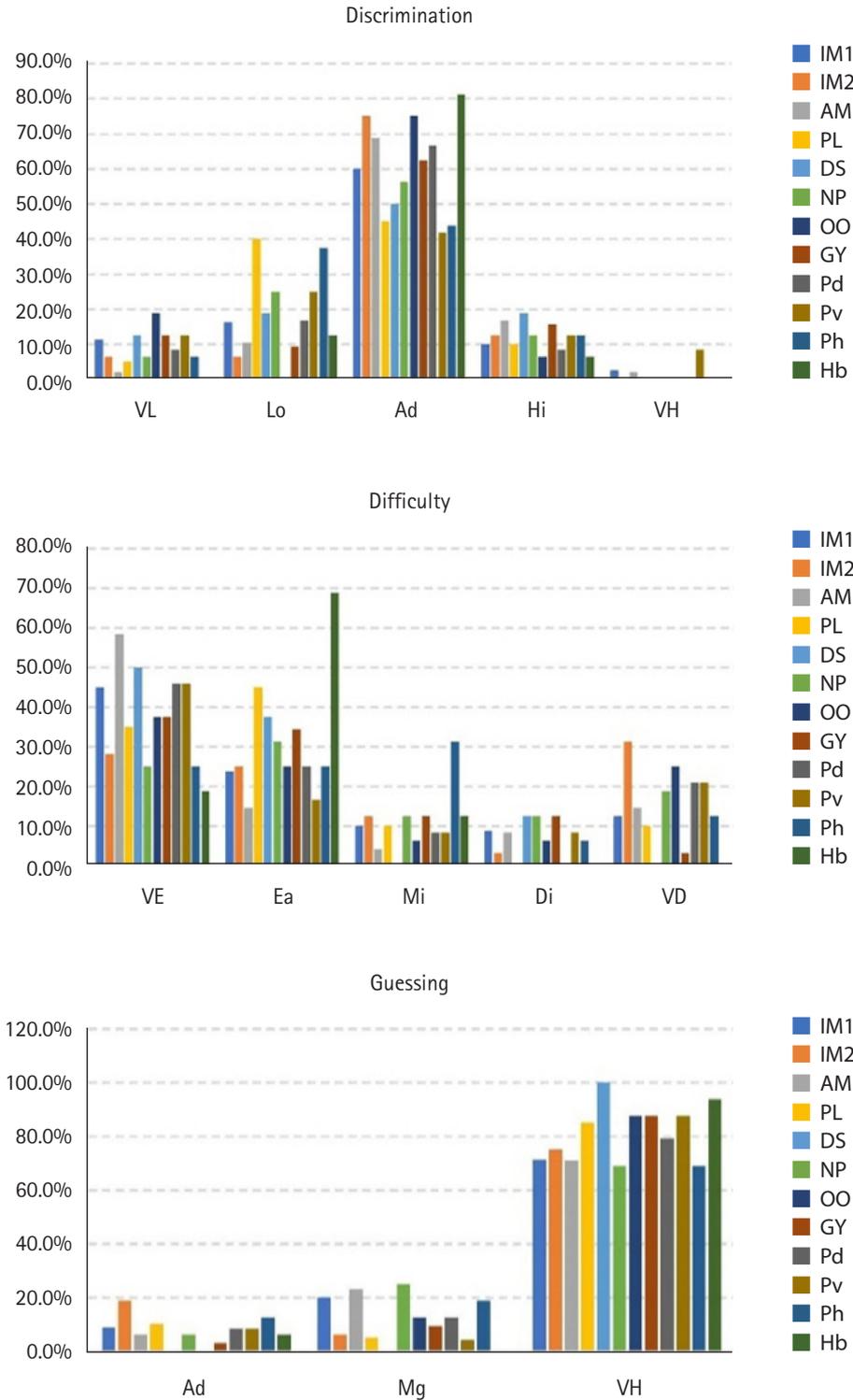


Fig. 1. Item validity of 12 subjects in national license mock exam.

IM1, internal medicine 1; IM2, internal medicine 2; AM, acupuncture medicine; PL, public health & medicine related laws; DS, dermatology & surgery; NP, neuropsychiatry; OO, ophthalmology & otorhinolaryngology; GY, gynecology; Pd, pediatrics; Pv, preventive medicine; Ph, physiology; Hb, herbology; VL, very low discrimination (<0.35); Lo, low discrimination (0.35-0.64); Ad, adequate discrimination (0.65-1.34); Hi, high discrimination (1.35-1.690); VH, very high discrimination (≥ 1.70); VE, very easy difficulty (<-2.0); Ea, easy difficulty(-2.0--0.5); Mi, middle difficulty (-0.5-0.5); Di, difficult (0.5-2.0); VD, very difficult (>2.0); Ad, adequate guessing (0-0.2); Mg, Marginal guessing (0.2-0.3); VH, Very High guessing (≥ 0.3)

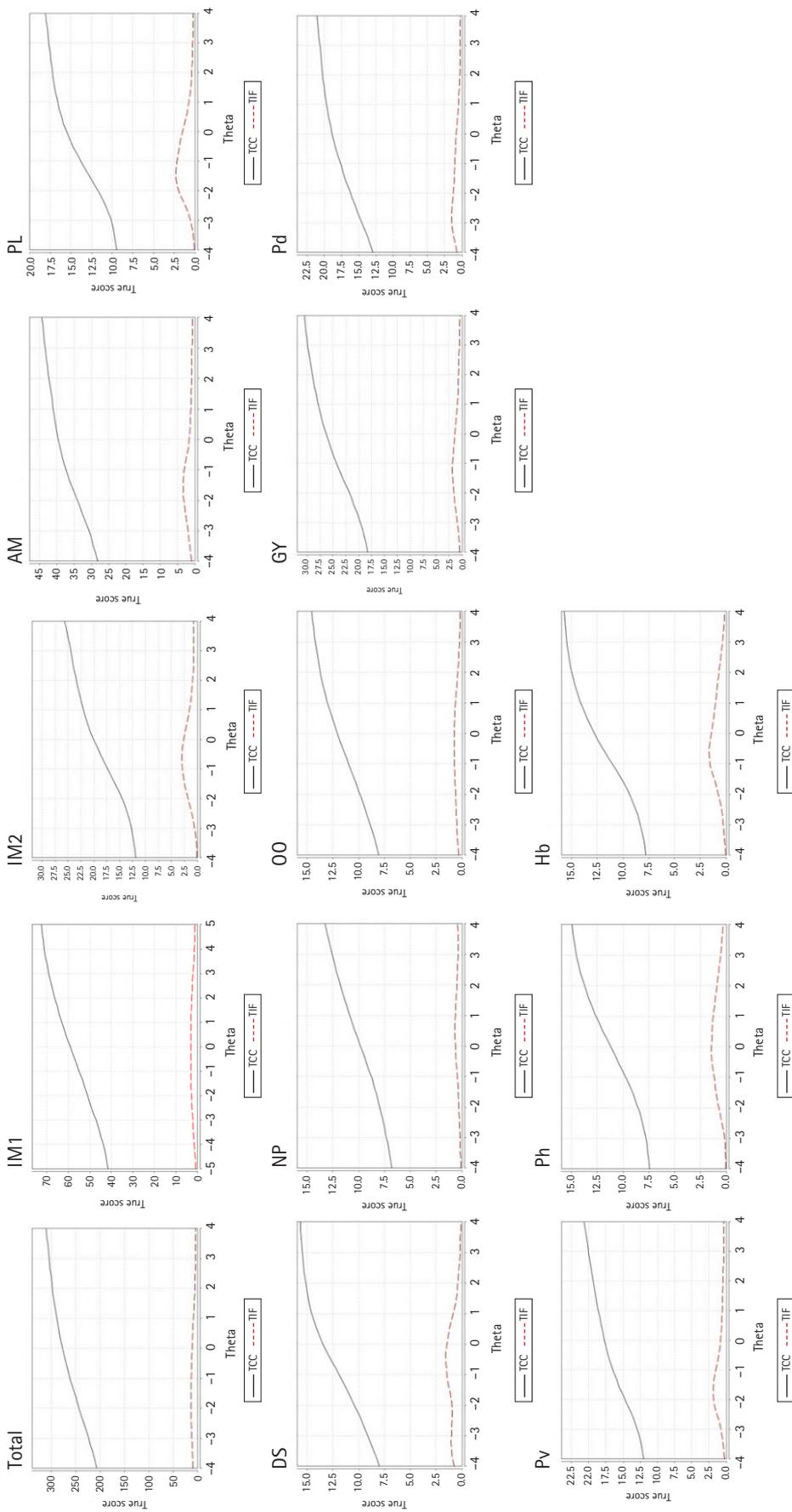


Fig. 2. Test character curve (true score) and test information curve of 12 subjects in national license mock exam. IM1, internal medicine 1; IM2, internal medicine 2; AM, acupuncture medicine; PL, public health & medicine related laws; DS, dermatology & surgery; NP, neuropsychiatry; OO, ophthalmology & otorhinolaryngology; GY, gynecology; Pd, pediatrics; Pv, preventive medicine; Ph, physiology; Hb, herbology

Table 2. Correlation coefficient among total and subject competencies

	Total	IM1	IM2	AM	PL	DS	NP	OO	GY	Pd	Pv	Ph
IM1	0.83***											
IM2	0.79***	0.55***										
AM	0.59***	0.45**	0.35*									
PL	0.52***	0.33*	0.33*	0.18								
DS	0.62***	0.67***	0.46**	0.15	0.32*							
NP	0.65***	0.56***	0.53***	0.29	0.27	0.41**						
OO	0.49**	0.37*	0.43**	0.04	0.15	0.29	0.27					
GY	0.74***	0.66***	0.52***	0.45**	0.15	0.46**	0.53***	0.45**				
Pd	0.50**	0.59***	0.45**	0.34*	0.01	0.35*	0.23	0.33*	0.44**			
Pv	0.64***	0.45**	0.39*	0.54***	0.54***	0.33*	0.28	0.27	0.33*	0.28		
Ph	0.55***	0.45**	0.5**	0.16	0.43**	0.29	0.11	0.42**	0.35*	0.32*	0.35*	
Hb	0.60***	0.54***	0.54***	0.11	0.31	0.31*	0.54***	0.3	0.28	0.27	0.27	0.29

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Bold represents coefficient bigger than 0.4. IM1, internal medicine 1; IM2, internal medicine 2; AM, acupuncture medicine; PL, public health & medicine related laws; DS, dermatology & surgery; NP, neuropsychiatry; OO, ophthalmology & otorhinolaryngology, GY, gynecology, Pd, pediatrics; Pv, preventive medicine; Ph, physiology; Hb, herbology.

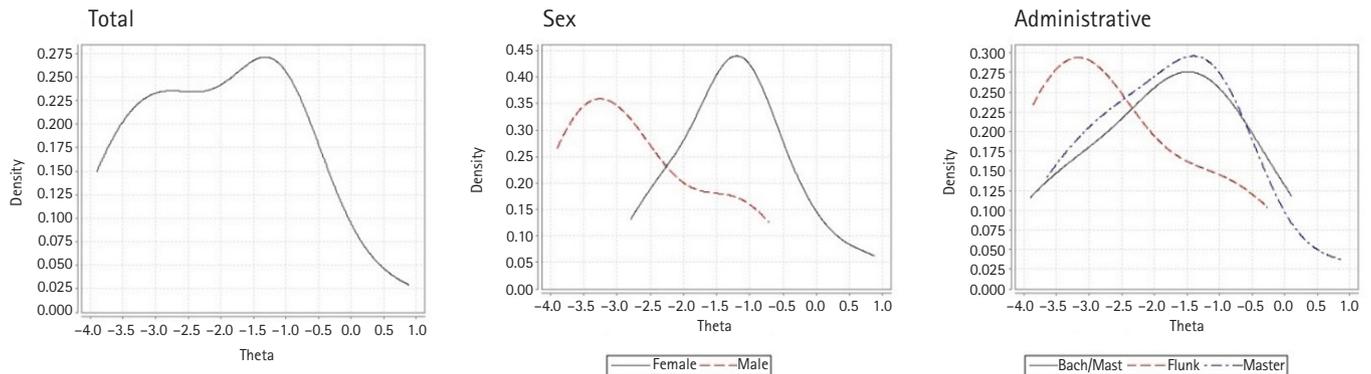


Fig. 3. Kernel density map according to the sex and administrative groups.

활용되는데, 이러한 방법은 국가시험의 많은 문항(12개 과목 340 문항)들을 일목요연하게 비교하기에는 한계를 지닌다. 본 연구에서 제시된 타당도 제시 방법은 모의고사 또는 국가시험의 차수 별 비교나 시행 연도별 차이의 분석에 매우 유용하게 활용될 수 있을 것이다.

둘째, 12개 과목별 시험 타당도 특성은 TCC와 TIC로 제시되었으며(Fig. 2), 12개 과목 사이의 관련성은 상관 계수로 제시되었다(Table 2).

의료인의 면허취득을 결정짓는 국가시험에서는 합격과 불합격을 결정짓는 합격선 점수(cut-off score) 또는 역량 수준(competency level)에서의 변별력(discrimination)과 정보(information)의 양이 매우 중요하다[1,2]. 본 연구에서의 TCC는 합격-불합격을 결정짓는 합격선 시험점수 또는 역량수준을 분석하는데 유용하며, TIC는 합격-불합격을 결정짓는 역량수준에서 시험의 변별력 또는 정보의 양을 직관적으로 제시하고 있다(Fig. 2). 국가시험이 공정성과 타당

성, 신뢰성을 충분히 담보하기 위해서는 합격선 점수(역량수준)에서 제공되는 정보의 양이 충분해야 하는데, 매년 95%를 상회하는 합격률이나 의사 국가시험(모의고사)에 대한 선행연구를 고려할 때 역량수준 -2~-1에서의 정보량이 많아야 한다[21].

이와 함께, 과목별 역량간 상관성(Table 2)은 각 과목이 측정하는 역량 요인 사이의 관련성이나 시험을 준비하는 시기(예를 들어, 시험 준비 처음에 또는 시험 직전)의 차이 등을 설명할 수 있을 것으로 보인다. 본 연구 결과를 기준으로, 내과학과 부인과학이 타 과목과의 유의한 상관성을 많이(7-9개) 보이고 있는 것은 이들이 타 과목 역량과 공통점을 많이 지니고 있다고 해석할 수 있을 것이다. 아울러, 다른 과목과 상관성이 적은(2개) 보건법규는 타 과목과 독립적인 성격을 지니고 있기 때문이거나 시험 직전에 공부하는 것으로 해석할 수 있을 것이며, 침구의학(3개)이나 예방의학(3개) 등은 시험공부를 시작하는 시기가 학생들 사이에 서로 상이할 수 있음을 의미할 것으로 사료된다.

Table 3. Model fit scores of Latent Profile Analysis using academic competency in 12 subjects

Model n.	# of profiles	AIC	BIC	CAIC	SABIC	Entropy	BLRT	BLRT-p
1	1	1320.91	1361.44	1385.44	1286.34	1		
2	2	1204.43	1266.92	1303.92	1151.14	0.95	142.47	0.01
3	3	1190.17	1274.61	1324.61	1118.15	0.94	40.27	0.03

AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; CAIC, Consistent Akaike Information Criterion; SABIC, sample size adjusted BIC; BLRT, bootstrapped likelihood ratio test; BLRT-p; p-value for BLRT.

Table 4. The distribution of high and low competency groups according to the administrative group and sex

Competency group	Administrative group	Sex		Total	Statistics
		Female	Male		
High	Master	6	7	13	$\chi^2 = 8.97, p = 0.011$
	Bachelor-Master	8	0	8	
	Flunk	4	0	4	
	Total	18	7	25	
Low	Master	0	7	7	
	Bachelor-Master	0	2	2	
	Flunk	0	6	6	
	Total	0	15	15	

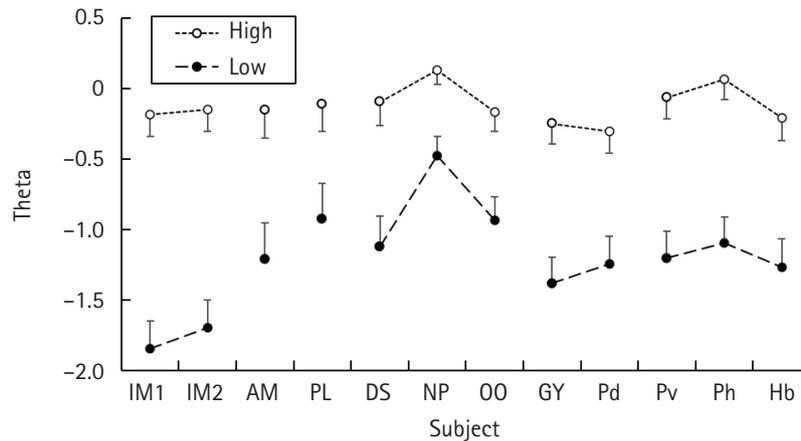


Fig. 4. High and low academic competency groups in 12 subjects of national license mock exam.

Six groups for attesting failing were connected with dashed lines. White circle is for high academic competency group and back circle for low academic competency group. Data shown as Mean and Standard Error. IM1, internal medicine 1; IM2, internal medicine 2; AM, acupuncture medicine; PL, public health & medicine related laws; DS, dermatology & surgery; NP, neuropsychiatry; OO, ophthalmology & otorhinolaryngology; GY, gynecology; Pd, pediatrics; Pv, preventive medicine; Ph, physiology; Hb, herbology

셋째, 학생들의 역량분석에 커널 밀도함수(Fig. 3)와 잠재 프로파일 분석(Table 3, Fig. 4)을 사용함으로써 학생 그룹간 차이를 다면적으로 확인할 수 있었는데(Table 4), 이는 평균과 표준편차를 사용하는 기존의 기술통계로는 확인하기 어렵다.

밀도함수를 사용한 역량 분포(Fig. 3)에서는, 평균 수준히 공부하는 학업 습관과 자기관리를 특징으로 하는 여학생이 남학생보다 비교적 높은 역량 수준을 지니고 있음을 확인할 수 있었다[10,11,14].

이와 함께, 역량수준 0 이상(최우수 수준)에서 학-석사과정(7년)과 석사과정(4년) 학생들 사이의 의미 있는 빈도 차이를 확인하였는데(Fig. 3), 이는 전문대학원 과정을 통해 입학한 성인학습자인 석사과정생이 평소의 꾸준한 자기관리를 통해 높은 역량수준을 지니고 있음을 의미하는 것으로 사료된다[10,11,14].

유급/휴학을 경험한 학생들의 역량수준은 -2.0~-0.5로 상당히 폭넓은 것으로 확인되었는데, 이는 유급/휴학을 경험한 학생들의

학업역량 수준이 예상보다 더 다양하다는 것을 의미한다. 한의사 국가시험이 12개 과목에 있어서 7년(또는 4년)간의 학업 성취도를 종합적으로 측정하기에 평소의 자기관리와 학업습관을 반영한다는 것을 고려한다면, 기존의 획일적인 관리보다 개인의 상황과 특성을 고려한 맞춤형 학업관리가 필요하다는 것을 재확인하는 것으로 보인다.

잠재 역량프로파일 분석(Table 3, Fig. 4)에서는 두개의 학업성취도 학생 그룹을 추출할 수 있었는데, 추가적인 빈도분석(Table 4)에서는 성별, 학사행정별로 빈도가 다른 것이 명료하게 드러난다. 석사과정(4년)을 제외했을 때, 상위권 그룹은 여학생이 하위권 그룹은 남학생이라는 이분법적 해석이 너무도 명확하게 드러났는데, 이는 여학생들이 평소 자기관리와 학업 관리를 잘 한다라는 통념을 객관적으로 확인하는 결과로 사료된다[10,11,14].

의료인 국가시험은 해당 의료 지역의 전문가로 활동하기 위한 필수 역량들을 국가가 최종적으로 확인하여 면허로 인정한다는 권위와 상징성을 지니기에, 한의학 교육에 있어서도 다음 사항들을 고려한 포괄적인 접근이 필요하다.

첫째, 국가시험은 서로 얽혀 있는 태도, 술기, 지식 역량에 대한 최종적인 확인이므로, 국가시험을 통해 확인할 한의사로서의 인성, 의료 기술, 한의학 지식의 구체적인 조작적 정의와 이를 안정적으로 측정, 분석할 수 있는 방법론들이 체계적으로 검토되어야 한다[3].

국시원은 의료인의 세 가지 역량에 대한 분석법을 체계적으로 개발하여 왔으나[4], 한의학 교육계는 기존 정책들을 개선하기 위한 인적, 물적 자원이 부족한 현실이다. 보건의료계에서는 인지적 지식 역량을 측정하는 CAT의 구체적인 도입방안이 이미 논의되었음에도[21-23], 한의계는 토대가 되는 IRT도 도입하지 못하고 있다. 의료 술기에 있어서는 조작적 정의와 안정적 평가법, 효율적 교육 체계에 이어 실기시험의 개선이 논의되고 있으나, 한의계는 제도의 시급한 도입에 쫓겨 필수 술기역량의 조작적 정의나 술기-지식 간의 상호관련성은 고려하지 못하고 있다[12,13]. 술기와 지식 역량간 상관성이 너무 높으면 불필요한 절차가 늘어날 뿐이며, 너무 낮으면 한의사의 역량과 무관한 시험이 될 것이다.

의료인의 기본적 인성에 대한 논의 또한 구성개념을 넘어 구체적인 측정 방법이 논의되고 있음에도, 과거 한의사의 인성 또는 의도(醫道)를 상징하던 '의자의야(醫者意也)'의 '의사에게는 환자의 마음과 증상을 잘 살펴 헤아리고 이해하는 것이 가장 중요하다'라는 경구는 잊혀지고 이제는 도리어 역수입을 고민해야 할 상황이다. 인간과 생명에 대한 존중, 사회적 책무, 훌륭한 인성, 인격이나 품성, 의료인의 이해심 등에 대한 유효한 교육프로그램과 다면적 평가법에 대한 실증적 논의가 필요하다[19].

둘째, 교육기관의 핵심역량은 각종 평가에 대한 학습분석에 있으며, 대학에 설치되는 CBT 시스템에는 IRT 학습분석과 experience application programming interface(xAPI)와 같은 국제표준이

기본적으로 포함되어야 한다[3,18]. xAPI는 온라인과 오프라인을 포괄하는 다양한 환경에서의 학습 데이터를 저장 및 공유하는 표준으로, 즉각적인 학습분석을 위한 이터닝의 토대이다.

한의학 교육에 대한 기존의 불만들은 대부분 학업분석의 부족에 기인하는데, 본 연구에서의 IRT를 사용한 학습분석은 이미 중고등 학생 대상 국가수준 학업성취도평가에서부터 의료인의 국가시험까지 폭넓게 적용되고 있다[21-23]. 한의학교육평가원의 한의과대학 교육 평가인증에 교육실의 설치 여부 또는 CBT 시스템의 설치 및 활용 횟수와 같은 피상적인 항목은 지양하고, IRT를 사용한 학업분석 보고서의 양과 질, 그리고 학업 분석을 기획, 해석하고 현장에 적용하기위한 교육평가 전문가의 보유와 활동과 같은 실질적 내용이 포함되어야 한다[29].

셋째, 국가시험의 개선을 위한 장기적인 연구가 필요하며, 상황에 따라 급변하는 제도만을 따라가기에 급급해 미숙한 혼란을 경험하기 보다는 3-5년 앞을 바라보는 비전이 요구된다. 기초종합평가나 과목통합 국가시험의 논의에 있어서 특정과목의 포함여부만큼 중요한 것이 기존의 합격선(과락기준 40%와 총점기준 60%)의 개선 문제이다[1,2]. 과락이 유사한 역량을 지닌 과목들을 묶어 최저 기준을 제한한다는 취지를 갖고 있으나, 본 연구(Table 2)는 이를 지지하지 않을 뿐만 아니라 의사국시에서는 이미 폐지되었다. 총점 60% 득점도 합격-불합격의 신뢰성과 공정성, 배출 의료인 숫자의 안정성 등을 고려하여 modified-Angoff 또는 bookmark 방법으로 수정하는 것을 고려하여야 한다[30].

과목통합 국가시험으로 개선하는 과정에서도 전체 문항의 개수 또한 필수역량 측정의 타당성과 신뢰성을 기준으로 결정되어야 한다[3,18]. 의사 국가시험의 경우 550문항(2006년)에서 320문항(2022년)을 거쳐 300문항 이하로 축소할 것이 제시되었으나[15], 한의사 국가시험의 경우 지난 문항축소(420문항에서 340문항으로)에 대한 근거도 확인할 수 없다. 실기시험의 신설에 있어서도 과목의 명칭과 문항 개수를 성급하게 결정하기보다는, 실기 역량의 구성 개념, 실기 역량을 고려한 평가 목표 및 출제 계획표(blueprint), 실기 모듈의 설계, 모듈별 세부문항의 개수, 세부 문항의 채점 방식(정오 또는 부분점수 부여), 합격선 산정 방법(합격 모듈의 개수 또는 전체 세부문항 점수 등) 등 술기 역량과 측정의 타당성부터 신중하게 검토되어야 할 것이다.

본 연구에서는 IRT를 사용한 한의사 국가시험의 타당도 분석방법과 역량의 다면적 분석방법을 처음으로 제시하였으며, 국가시험을 통해 한의학 교육의 질을 높일 수 있는 방안에 대하여 고찰하였다. 본 연구에서의 분석결과를 일반화하여 한의학 교육현장에 적용하기 위해서는 다음과 같은 한계를 극복하기위한 후속연구가 진행되어야 한다.

첫째, 본 연구를 통해 제시된 성별, 학사행정 그룹간 역량수준의 차이는 일개 대학에서 40여명을 대상 진행된 모의고사를 통해 확인된 것이다. 비록 의료인 국가시험 및 모의고사에 대한 선행연구

[10,11,14]와 유사한 결과이지만, 다년간의 모의고사 또는 여러 대학에서의 시행결과, 실제 국가시험 결과 등을 대상으로 재확인되어야 할 것이다. 또한, 분석결과가 외부로 공개될 수는 없더라도, 성별, 학사행정 그룹, 학교, 학제간 학업역량의 비교를 통해 미래 한 의학의 가장 적절한 교육과정이 기획되어야 한다.

둘째, 본 연구를 통해 성별, 학사 그룹별 학업역량의 유의한 차이를 확인할 수 있었는데, 이는 학업역량 수준에 따른 교육 다양성이 한의학 교수자들에게 요구되고 있음을 시사한다. 높은 성취도 그룹은 전반적으로 높은 학업성취도와 안정된 자기 관리와 학업 습관을 지닌 것으로, 낮은 성취도 그룹은 이와 상반된 것으로 여겨지므로, 학생 고유의 특성을 고려한 개별화 교육(differentiated instruction) 또는 개별화 학습(individualized learning)이 필요하다. 향후 두 역량수준 그룹을 대상으로 자기 관리와 학업 습관에 대한 후속 질적 연구를 진행하여, 맞춤형 성취도 향상을 위한 구체적인 학습지도 방안이 도출되어야 한다. 이와 함께, 성별, 학사 그룹별 학업역량의 차이를 특정할 수 있는 시험 문항 또는 모의고사 과목에 대한 연구는 진행하지 못하였다. 이에 추가연구를 통해 그룹간 차별 문항 혹은 과목을 추출하고, 부족한 역량을 보완하기 위한 교육 커리큘럼, 맞춤 학습법과 학생 관리방법 또한 시급히 제안되어야 할 것이다.

셋째, 본 연구에서 확인된 성별, 학사 그룹별 차이가 높은 난이도를 지닌 모의고사 또는 국시원 주관 국가시험에서도 유사하게 나타날 것인지에 대한 추가연구가 필요하다. 한의사 국가시험이 12개 과목으로 이루어진 고부담시험이기에 시험 날짜에 다가갈수록 준비 정도가 급격히 높아질 수 있음을 고려한다면, 본 연구에서 사용된 잠재프로파일 분석이 지닌 데이터 의존성이 역량그룹의 추출이나 역량간 상관성에 영향을 줄 수 있기 때문이다.

본 연구에서는 문항반응이론을 사용하여 한의사 국가시험을 준비하기 위한 대학 자체 시행 모의고사를 분석하여 문항 및 시험의 타당도와 학생 그룹별 역량수준을 객관적으로 분석하였다. 이와 같은 다면적 학업분석은 한의학 교육 측정학의 토대에 기여할 수 있을 것이며, 이러한 과정에서 한의학 교육에 근거기반 역량중심 의학교육(evidence-based competency-focused medical education)을 적용할 수 있을 것이다[18,24,25].

Acknowledgements

본 연구는 부산대학교의 연구비지원을 받았음 (This work was supported by a 2-Year Research Grant of Pusan National University.)

Conflict of interest

The authors declare no conflicts-of-interest related to this article.

References

1. 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;1:67–76.
2. Lee G. A psychometric approach to setting a passing score on Korean national medical licensing examination. *J Educ Eval Health Prof.* 2004;1:5–14.
3. Chae H, Lee SJ, Han CH, Cho YI, Kim HW. Study on the academic competency assessment of herbology test using rasch model. *J Korean Med.* 2022;43:27–41.
4. Korea Health Personnel Licensing Examination Institute. [cited 2023 Mar 17]. Available from: <https://www.kuksiwon.or.kr/>.
5. Yang E. Perceptions on item disclosure for the Korean medical licensing examination. *Korean J Med Educ.* 2015;27:167–75.
6. Baik SH. Major reforms and issues of the medical licensing examination systems in Korea. *Korean Med Educ Rev.* 2013;15:125–35.
7. Yim MK, Huh S. Testing unidimensionality and goodness-of-fit-ness for the application of Item Response Theory to the Korean medical licensing examination. *Korean J Med Educ.* 2007;19:163–9.
8. Lee C, Lee S, Cho K, Lee EI, Park ES, Lee SJ, et al. Development of CAT for national licensure examinations (medical and nurse) using item response theory. *J Health Sci Med Technol.* 2005;31:55–73.
9. Lim EY. Comparison of item analysis according to item response theory and classical test theory. Seoul: Korea Health Personnel Licensing Examination Institute; 2001. Report No: RE1-0103-00.
10. Park J, Young H. Depending on admission types, the comparison of GPAs, dropout rates, and passing rate of medical license examination. *J Educ Eval.* 2017;30:317–40.
11. Ahn SS, Seo YK, Baek SE, Bae SY, Seol JH, Lee HY, et al. The correlation of grade point average of medical school and the score of Korean medical licensing examination. *Korean J Med Educ.* 2004;16:25–32.
12. Jung KH, Jung HK, Lee K. The relationship between senior year examinations at a medical school and the Korean medical licensing examination. *Korean J Med Educ.* 2009;21:17–22.
13. Kim MS, Uh Y, Lee JI, Chang SJ, Park KC, Park JY, et al. Correlation between the academic score, the trial examination score, and the Korean medical licensing examination score. *Korean J*

- Med Educ. 2005;17:73–82.
14. Han ER, Chung EK, Oh SA, Oh CK, Woo YJ. Medical students' failure experiences and their related factors. *Korean J Med Educ.* 2012;24:233–40.
 15. 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;2:113–8.
 16. Lee YH, Park JH, Park IY. Estimation of an examinee's ability in the web-based computerized adaptive testing program IRT-CAT. *J Educ Eval Health Prof.* 2006;3:4.
 17. Choi K, Cho JK. Statistical analysis of national examination for radiological technologists in convergence perspective. *J Korea Converg Soc.* 2017;8:93–9.
 18. Chae H, Han SY, Yang GY, Kim H. Study on the herbology test items in Korean medicine education using item response theory. *Korea J Herbol.* 2022;37:13–21.
 19. Kim BS. Concept of talent on the doctor of Korean medicine to pursue university education in vocational performed for Korean medical doctor. *J Physiol Pathol Korean Med.* 2015;29:256–66.
 20. Macmillan dictionary. London: Macmillan Education Limited; 2023. Competence.
 21. Heo S. Metric review, item bank construction, and administrative/financial support for the introduction of computerized adaptive testing in the national licensing examination for Korean healthcare professionals. Seoul: Korea Health Personnel Licensing Examination Institute; 2018. Report No: RE01-1816-00.
 22. Seo M. A study on designing for the computer-based NAEA. Jincheon-Gun: Korea Institute for Curriculum and Evaluation; 2020.
 23. Kim J. A study on online evaluation system based on individualized education. Daegu: Korea Education and Research Information Service; 2018. Report No: RR 2018-1.
 24. Park J, Yoon T, Ko J. Development and validation of a questionnaire to evaluate medical students' evidence-based medicine competencies. *Korean J Med Educ.* 2009;21:259–67.
 25. Schaubert SK, Hecht M, Nouns ZM. Why assessment in medical education needs a solid foundation in modern test theory. *Adv Health Sci Educ.* 2018;23:217–32.
 26. Korea Health Personnel Licensing Examination Institute. Development of multiple choice test item. Improvement of item development ability (Workshop); Yangsan campus, Pusan National University: Korea Health Personnel Licensing Examination Institute; 2019.
 27. Seong T. Understanding and application of item response theory. 2nd ed. Paju: Educational Science Publishing; 2016.
 28. Meyer JP. Applied measurement with jMetrik. New York: Routledge; 2014.
 29. Huh S. Can computerized tests be introduced to the Korean medical licensing examination? *J Korean Med Assoc.* 2012;55:124–30.
 30. Yim M. Comparison of results between modified-Angoff and bookmark methods for estimating cut score of the Korean medical licensing examination. *Korean J Med Educ.* 2018;30:347–57.