

Original Article

교육 현장을 위한 시험 타당도 및 학업 역량 분석 연구

채한¹, 이해윤², 한상윤³

¹ 부산대학교 한의학과, ² 부산대학교 한의학전문대학원 인문사회의학교실, ³ 대전대학교 한의과대학

IRT-based analysis of test validity and academic competency for educational practice

Han Chae¹, Hye-Yoon Lee², Sang Yun Han³

¹School of Korean Medicine, Pusan National University, ²Division of Humanities and Social Medicine, School of Korean Medicine, Pusan National University, ³College of Korean Medicine, Daejeon University,

Received: July 23, 2023
Revised: August 17, 2023
Accepted: August 17, 2023

Background: Specific methodologies and application examples are required to make Item Response Theory (IRT) analysis more accessible for competency-based learning and personalized education in the field of traditional medicine education practice. The purpose of current study was to analyze the test and item validity along with the competency characteristics of four subjects in a mock National License Examination (NLE) as a representative showcase.

Methods: The 3-parameter IRT was used to analyze the pediatrics, preventive medicine, physiology and herbology subjects of mock NLE. The Test Character Curve (TCC), Test Information Function (TIF), and Test Standard Error Curve (TSE) were used to assess test validity, and the Wright map was used to attest item and test validity. Kernel density map was used to analyze the characteristics of competencies by gender and academic administration groups.

Results: TCC, TIF, and TSE, along with Wright map, provided intuitive understanding of test and item validity, and showed distinctive differences between subjects. The Kernel density map revealed group-specific competency characteristics for evidence-based personalized student management.

Discussion and conclusion: A practical learning analytics methods using IRT were presented, which can be easily utilized in the field of medical education. Systematized regular academic analysis and individualized student management with basic comprehensive assessment and NLE could be useful for improving the quality of evidence-based competency-based education in traditional Korean medicine.

Keywords: Item Response Theory, Learning Analytics, National License Exam, Evidence-Based Teaching, Competency-Focused Education

Corresponding Author:
Sang Yun Han
College of Korean Medicine,
Daejeon University. 62 Daehak-
ro, Dong-gu, Daejeon, 34520,
Korea
E-mail: drhan@dju.kr

서론

한의학 교육에 있어서의 역량중심 접근이란, 미래 한의사로서의 직무 수행에 필요한 실질적인 윤리적 태도(attitude), 임상적 기술(behavior), 한의학 지식(knowledge)을 중심으로 교육 평가 체계를 재구조화 해야 한다는 것으로서¹, 기존의

교과서와 원전(原典) 및 교수자 중심 교육이나, 경쟁과 수동성을 강조하는 획일적 교육에서 벗어나 학습자를 중심으로 실질적인 의료 현장 역량(力量, competency)을 기르는 것을 의미한다².

이러한 역량중심 한의학 교육을 위해서는 객관적으로 평가된 역량의 분석을 근거로 한의학 교육이 기획되고, 시행되고,



평가된 다음 수정이 새롭게 반영되어 개선된 교육으로의 환류가 이루어지는 근거에 기반한(evidence-based) 체계적 접근이 필수적이다. 그러나, 역량의 사전적 정의가 ‘잠재되어 드러나지 않는 개인 고유의 특성’이라는 점을 고려한다면³, 잠재하고 있는 역량을 정확하고 신뢰롭게 측정하고 분석하는 것이 역량중심 그리고 근거기반 교육의 핵심 요소라 할 수 있을 것이다.

문항반응이론(Item Response Theory, IRT)은 학습자의 잠재 역량을 분석하는 학업분석 방법의 하나로¹, 교육 평가를 비롯하여 심리적 특성의 측정과 임상 진단 등에 있어서 타당도 분석에 폭넓게 활용되고 있다. 기존의 고전검사이론(Classical Test Theory)이 학생들의 역량 분석에 개별 평가문항의 난이도 분포 또는 동시에 평가받는 동료의 수준 등에 크게 영향을 받는 상대평가 점수를 제시하는 것과는 다르게, IRT는 잠재되어 있는 학생들의 절대적 역량을 문항 및 시험 자체에 대한 피험자의 반응을 분석하여 보다 정확하게 그리고 연속형 절대평가 점수로 제시한다^{1,4}.

이와 함께, IRT를 사용한 학업역량의 평가는 교육의 두 주

체인 학습자와 교수자 모두에게 핵심적인 가치를 제공한다^{1,4}. 교수자(평가자)에게는 신뢰로우면서도 효율적인 학업성취도 평가와 시험 타당도 평가를 가능하게 하며, 학습자(피평가자)에게는 과목 또는 단위별로 정확한 역량 수준을 제시하여 학습자의 수준을 고려한 학생 맞춤형 교육을 가능하게 한다.

그러나, IRT의 활용이 해외 의료인 면허시험과 국내 기초 학력 평가 등에 확대⁵⁻⁷되어가는 과정에 있으며, 이십여 년 전부터 국내 의학교육⁸⁻¹⁵에 꾸준히 도입되어 왔음에도 불구하고, 한의학 교육에 있어서는 기본 개념과 분석 방법의 어려움으로 인하여 최근 연구^{1,4,16}를 통해서야 처음으로 소개되었다. 이러한 한의학 연구들은 본초학 중간/기말고사와 같은 일상적인 교육^{1,4}과 함께 면허 국가시험과 같은 종합평가¹⁶에 적용될 수 있음을 실증적으로 제시하였다.

그러나, 국내 교육계의 발전 동향을 고려할 때 한의학 교육에도 IRT가 적극적으로 활용되어야 하지만, 교육 현장에서의 현실적인 한계는 이를 지연시키고 있다^{1,4,16}. 현재의 한의학 교육실은 학업분석과 IRT 활용을 위한 하드웨어적 토대로서의 컴퓨터기반시험(Computer-based Test, CBT) 시스템

Table 1. Explanation and Example of Item Response Theory Terms in Current Study (modified from Chae et al.¹)

Term	Explanation	Example in current study
Competency or Theta (θ)	Unique and latent ability (θ) rated as a continuous variable ranging from -4 and +4. The competency is a major interest of Item Response Theory (IRT) and can be used for standardized and validated evaluation of academic achievement.	X axis of TCC/TIC and Kernel density map. Y axis of Wright map.
Test	Assessment tool consisted of test items for analyzing specific academic competency	Subjects of National License Examination
Item	Basic unit of question and answer for competency measure.	Figure 1 (Wright map, X axis of right panel)
True score	The score showing the latent academic competency of specific subject estimated using logistic probability function $P(\theta)$.	Figure 1 (TCC, Y axis)
Difficulty	The difficulty of item. Ability level which can achieve 0.5 probability of correct answer.	Figure 1 (Wright map, Y axis of right panel)
Test Characteristic Curve (TCC)	A curve providing estimated true score corresponding to the competency of specific subject.	Figure 1 (TCC, solid black line)
Test Information Function Curve (TIF)	A curve showing the amount of information yielded at specific competency level and representing reliability of test.	Figure 1 (TIF, dashed red line)
Test Standard Error Curve (TSE)	A curve displaying the amount of errors for estimating competency at specific competency level as reliability of test	Figure 1 (TSE, dashed blue line)
Wright Map or item-person Map	A figure showing the item-person map with two parts: left box is for the distribution of subjects and right box for difficulty parameter of items in relation to competency level.	Figure 1 (Wright map)
Kernel Density Map	A figure showing the relative density and distribution of subject population at specific ability levels.	Figure 2

과 분석을 시행하고 이해하여 현장에 적용하기 위한 소프트웨어적 토대로서의 측정분석학 전문가와 갖추지 못하고 있다. 이와 함께, 직접적으로 교육을 수행하는 내용 전문가들은 학업분석 결과를 적절히 활용할 만큼 IRT를 이해하지는 못하고 있다. 선행 연구 또한 IRT의 활용 가능성만을 제시하였을 뿐, 교육 현장에서 즉각적으로 활용하기 위한 손쉬운 예시와 핵심 내용을 제시하지는 못하였다.

이에, 본 연구에서는, 한의학 교육 현장에서 실무자와 내용 전문가가 IRT를 교육 평가에 일상적으로 활용할 수 있도록 보다 간략하게 정리된 그리고 현장에서의 활용을 중심으로 하는 구체적인 쇼케이스를 제시하고자 하였다. 가장 잘 기획된 5 지선다형 객관식 시험이라 할 수 있는 모의 한의사 국가시험을 대상으로¹⁶, 연구자의 주요 관심 과목이 포함된 4 교시(소아, 예방의학, 생리, 본초학)에 사용된 과목별 문항과 시험을 분석하였다. 이 과정에서 분석의 시행과 이해에 필수적인 IRT의 기본 개념들(Table 1)과 이를 직관적으로 표현하는 시험특성곡선과 시험정보곡선, 라이트 지도 및 커널 밀도지도 등을 사용하였는데¹, 이러한 시험 타당도 분석과 학생 역량 분석¹⁶은 교육 현장에 있어서 그대로 활용이 가능하다.

본 연구는 한의학교육 현장에 바로 활용될 수 있는 IRT 활용 학업분석의 방법론적인 토대를 제공할 수 있을 것이며, IRT를 사용한 학업역량 평가를 근거로 교수자-학습자 중심의 한의학 교육(Evidence-based competency-focused medical education)에 기여할 수 있을 것이다^{4,17,18}

연구 도구 및 방법

1. 연구 대상 및 절차

부산대학교에서 2019년에 실시된 첫번째 한의사 국가시험 모의고사 중 소아과학, 예방의학, 생리학 및 본초학의 4 과목에 대한 40명의 답안을 본 연구의 대상으로 사용하였다. 본 연구는 부산대학교 생명윤리위원회의 승인(PNU IRB/2022_75_HR)을 받은 이후 분석을 진행하였다.

2. 한의사 국가시험 모의고사

한국보건의료원 국가시험원(국시원)이 주관하는 한의사 국가시험은 내과학 1, 내과학 2, 침구학, 보건의약관계법규, 외과학, 신경정신과학, 안이비인후과학, 부인과학, 소아과학, 예방의학, 한방생리학, 본초학의 12 과목으로 이루어져 있으며, 5 지선다(긍정형, A형) 형태의 340 문항(문항당 1 점)으로 구성된다¹⁶.

각 대학에서 실시해온 모의시험은 국시원의 한의사 국가시험을 준비하기 위하여 형식이나 절차를 동일하게 사용하며, 부산대학교에서는 한의학과 졸업대체 기준에 따라 교육실의

관리하에 교수가 출제한 2 번의 모의고사를 4 학년 2 학기에 시행한다¹⁶.

본 연구에서는 지면의 제약을 고려하여 연구자의 주요 전공 및 관심 영역이 포함되어 있는 4 교시의 소아과학(24 문항), 예방의학(24 문항), 한방생리학(16 문항) 및 본초학(16 문항) 과목만을 대상으로 분석을 진행하였다.

3. 시험 및 문항의 타당도 분석

모의고사 각 과목의 시험 타당도로서, 역량 수준별 진점수를 제시하는 시험특성곡선(Test Character Curve, TCC)과 함께, 역량수준별 문항 분산 값의 역수를 모두 합한 값으로 역량 수준별 정보의 양을 표현하는 시험정보곡선(Test Information Function, TIF)과 역량 수준별로 제공되는 정보의 정확도를 표현하는 시험표준오차곡선(Test, Standard Error Curve, TSE)을 제시하였다(Table 1)^{14,16}.

TCC는 시험의 민감도와 특이도를 반영하는데, 좋은 시험은 낮은 역량수준에서는 낮은 진점수를 그리고 높은 역량수준에서 높은 진점수를 보이게 된다. 교육 현장에 있어서, 역량수준 최저(-5)에서 X 축과 만나는 점수인 최저점은 0 점에 가까울수록, 역량수준 최고(+5)에서의 점수인 최대점은 만점에 가까울수록 변별력이 좋은 시험이라고 해석할 수 있다.

TIF와 TSE는 시험의 신뢰성과 정확성을 반영하는데, 좋은 시험이라면 목표하는 역량수준에서의 TIF가 높고 TSE가 낮아야 한다. 교육 현장에 있어서, 통상적인 학생들의 역량수준(-2~+2)에 있어서 TIF가 높고 TSE가 낮다면 시험이 학생들의 역량에 대한 정보를 신뢰롭고 정확하게 측정하고 있다고 해석할 수 있다.

모의고사 각 과목의 문항 타당도(Table 1 and Figure 1)는 라이트 지도(Wright map)를 사용하여 제시하였는데, 이는 피험자의 역량을 측정하기 위한 문항 난이도의 적절성을 직관적으로 표현할 수 있다. 역량 수준을 나타내는 하나의 동일한 로그 지표를 기준으로 사용하여, 왼쪽에는 역량에 따른 학생의 분포를, 오른쪽에는 문항별 난이도를 동시에 제시하기에 문항-피험자 지도(item-person map)라고 불리우기도 한다^{1,19}.

교육 현장에서 사용할 때에는, 먼저 학생들의 역량수준의 범주를 왼쪽 그림에서 확인한다음, 이 난이도 범주에 해당되는 문항들을 오른쪽 그림에서 확인한다. 통상적인 학생들의 역량수준(-2~+2)에 해당하는 문항의 개수가 많다면 좋은 문항과 시험이라고 할 수 있으며, 그 외의 역량수준(-2보다 작거나 +2보다 높은)에 해당하는 문항들은 검토를 통해 적절한 난이도를 갖도록 수정되어야 한다.

4. 성별 및 학사행정 그룹에 따른 학생의 역량특성 분석

그룹 구별에 따른 역량수준별 학생 분포는 커널 밀도지도(Kernel Density Map)로 제시하여, 각 그룹의 역량수준 특성

을 직관적으로 표현하였다 (Table 1)¹. 커널 밀도지도를 사용하여 각 그룹의 역량수준별 분포를 손쉽게 비교할 수 있는데, 이 과정에서 낮은 또는 높은 역량 수준을 지닌 그룹을 특정할 수 있다. 교육 현장에서는 이러한 역량수준별 분포를 사용하여, 성취도가 낮은 그룹에서는 저성취도의 원인을 찾아 개선하기 위한 맞춤형 지도 방안을 찾도록 하며, 성취도가 높은 그룹에서는 교육과정에서의 장점을 극대화할 수 있도록 할 수 있다.

본 연구에 있어서, 성별 그룹으로는 남학생과 여학생의 2개 그룹을 사용하였으며, 학사행정 그룹으로는 유급과 휴학을 경험한 학생, 학석사 과정(7년, 학사 3년과 석사 4년)과 석사 과정(석사 4년)의 3개 그룹으로 구분하였다¹⁶.

5. 통계 분석

모의고사 4 과목의 문항별 응답치를 활용한 역량의 추정에는 3-모수 모형(3-parameter logistic model)과 사후기대법 (expected a posteriori, EAP)을 사용하였으며, 이를 통해 모의고사 각 과목에 대한 학생 개개인의 학업 역량과 함께 문항별 모수를 추정하였다¹⁶. 커널 밀도지도의 밀도 함수를 매끄럽게 함에 있어서는 가우시안 평활화(gaussian smoothing)를 사용하였다.

문항 및 시험 분석과 역량특성의 분석에는 jMetrik 4.1.1(J. Patrick Meyer, Charlottesville, VA)을 사용하였으며²⁰, 그 외의 분석에는 jamovi 2.3.28(The jamovi project, <https://www.jamovi.org>)이 사용되었다. 통계 분석의 결과는 평균±표준오차 또는 빈도(%)로 표기하였으며, 통계적 유의수준으로는 $p<0.05$, $p<0.01$ 및 $p<0.001$ 이 사용되었다.

결과

1. 시험 및 문항의 타당도

모의고사의 소아과학, 예방의학, 생리학, 본초학 과목의 역량수준별 진점수(TCC)와 제공되는 정보의 양(TIF) 그리고 표준 오차(TSE)는 Figure 1 과 같다. 네 과목 모두에 있어서 가장 낮은 역량수준(-5)도 배정 점수의 절반을 득점하고 있으므로 성취도 평가 시험으로서의 변별도는 비교적 낮은 것으로 확인되었다.

생리학과 본초학은 각각 역량수준 0 과 -1 에서 비교적 많은 정보를 제공하고 있었으나, 이와 달리 소아과학과 예방의학은 해당 역량수준에서 제공하는 정보의 양이 많지는 않았다. 소아과학과 예방의학은 역량수준 -5~+5 에서 고르게 낮은 표준오차를 보였으나, 이와 달리 생리학과 본초학은 매우 낮은(<-3) 또는 높은(<3) 역량수준에서 높은 표준오차를 보였다. 이러한 결과는 생리학과 본초학이 중간 역량수준(-

3~+3)에 최적화되어 있으며, 소아과학과 예방의학은 모든 역량수준에 사용하는 평가에 최적화되어 있음을 확인할 수 있었다.

모의고사 네 과목에서의 문항 분석 결과를 제시하는 라이트 지도(Wright Map)는 Figure 1 과 같다. 매우 높거나(+3<) 또는 낮은 (<-3) 역량수준에 해당하는 문항이 소아과학(8 개)과 예방의학(6 개)에서 비교적 많았던 것과 달리, 생리학(1 개)과 본초학(1 개)에서는 비교적 적은 것으로 확인되었다. 이러한 문항 난이도의 분포가 위에서 살펴본 역량수준별 정보의 양 (TIF)과 표준오차(TSE)의 크기를 결정한 주요 원인이었던 것으로 사료되는데, 합격 불합격의 구분을 위해서는 소아과학과 예방의학의 시험 문항들이 생리학과 본초학처럼 문항 난이도 중간(-2~+2) 수준에 밀집되도록 수정되어야 할 것이다.

문항 분석의 결과를 사용하여 생리학 시험 문제를 분석한 결과는 다음과 같았다. 생리학에 있어서 통상적인 역량 수준보다 높은 난이도 또는 어려운 문항은 2 번과 5 번 문항으로, 각각 ‘五行相乘에 해당되는 내경 원문’과 ‘營氣 不足의 임상 증상’을 제시된 답가지에서 찾아내기를 요구하였다. 2 번 문항은 답가지를 꼼꼼히 읽고 분석해야 한다는 점에서, 5 번은 임상 증상들의 기전을 하나씩 분석해야 한다는 점에서 난이도가 출제 의도보다 높았던 것으로 사료되며, 보다 낮은 난이도를 지니도록 수정되어야 할 것이다.

이와 함께, 생리학에 있어서 통상적인 역량수준보다 낮은 난이도 또는 쉬운 문항은 3 번과 9 번으로, 각각 ‘三陰三陽의 標本 從化規律의 옳은 예시’와 ‘升降의 外輸에 해당하는 장부’를 제시된 답가지에서 찾아내기를 요구하였다. 3 번은 개념만 옳게 이해하고 있으면 답가지를 쉽게 분석할 수 있었기 때문에, 그리고 9 번은 비교적 단순한 개념이 매우 짧은 단어로 제시되었기에 출제 의도보다 쉽게 받아들여졌던 것으로 사료되며, 보다 높은 난이도를 지니도록 수정되어야 할 것이다.

2. 성별 및 학사행정 그룹에 따른 학업 역량 특성

성별(남학생과 여학생)과 학사행정(유급 및 휴학, 학석사(7년) 과정 및 석사(4년) 과정) 그룹에 따른 과목별 역량분포 특성을 분석한 커널 밀도함수는 Figure 2 와 같다.

전체 학생들에 있어서, 소아과학과 예방의학에서는 중간 수준의 역량에 몰려 있는 경향성이 생리학과 본초학보다 비교적 높았는데, 이는 생리학과 본초학에서의 학업량이나 학업 성취도가 비교적 다양함을 의미한다.

성별에 있어서는, 모든 과목에서 여학생의 역량수준이 남학생보다 비교적 높은 것을 확인할 수 있었으며, 생리학에 있어서는 남학생 그룹에서 두개의 역량그룹이 존재함을 추정할 수 있었다. 이러한 결과는, 성별 차이를 고려할 때에는 남학생들의 전반적인 학업 성취도를 향상시킬 수 있도록 지도해야

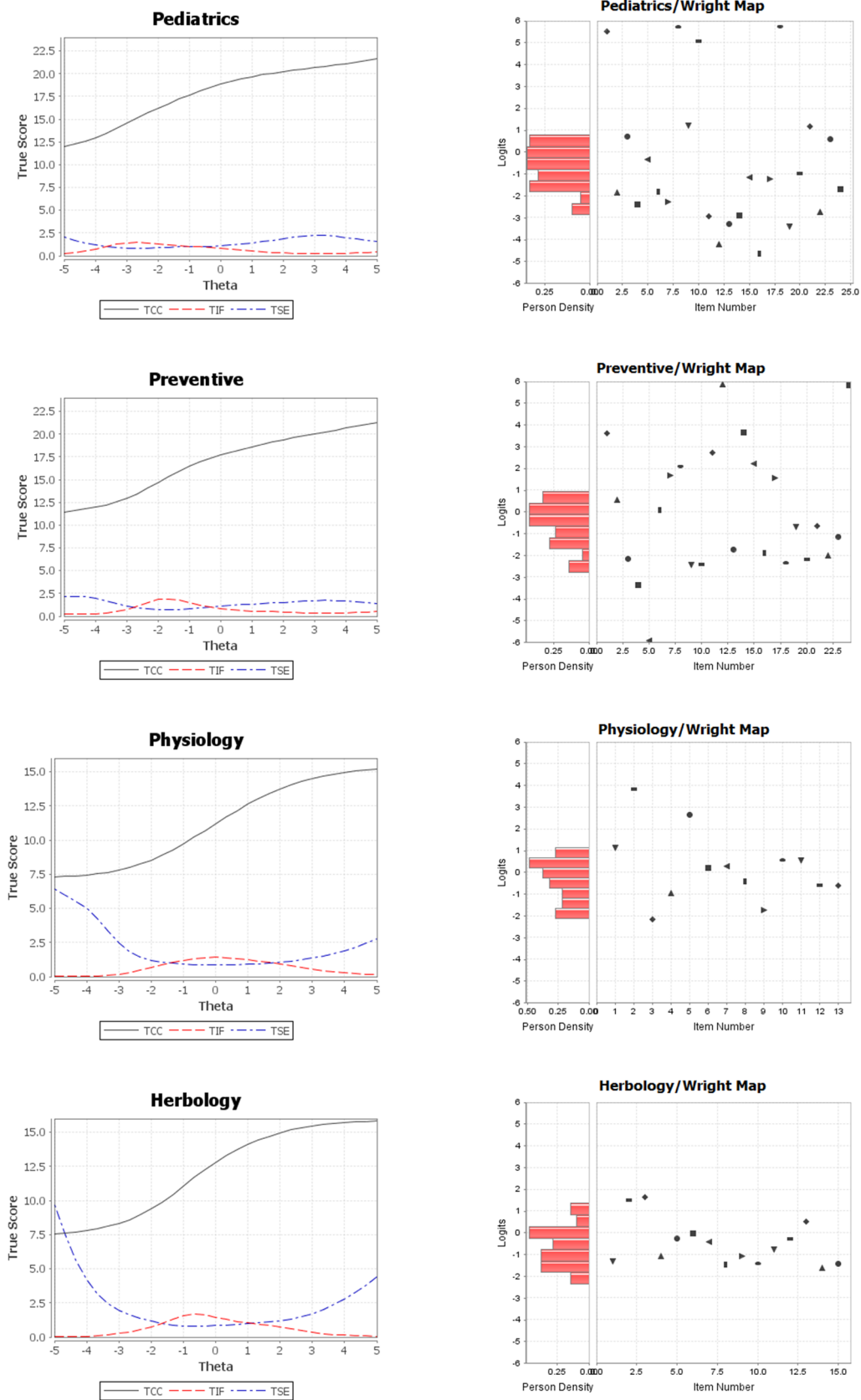


Figure 1. Test curve and Wright map for analyzing test and item validity in four subjects of mock national license exam

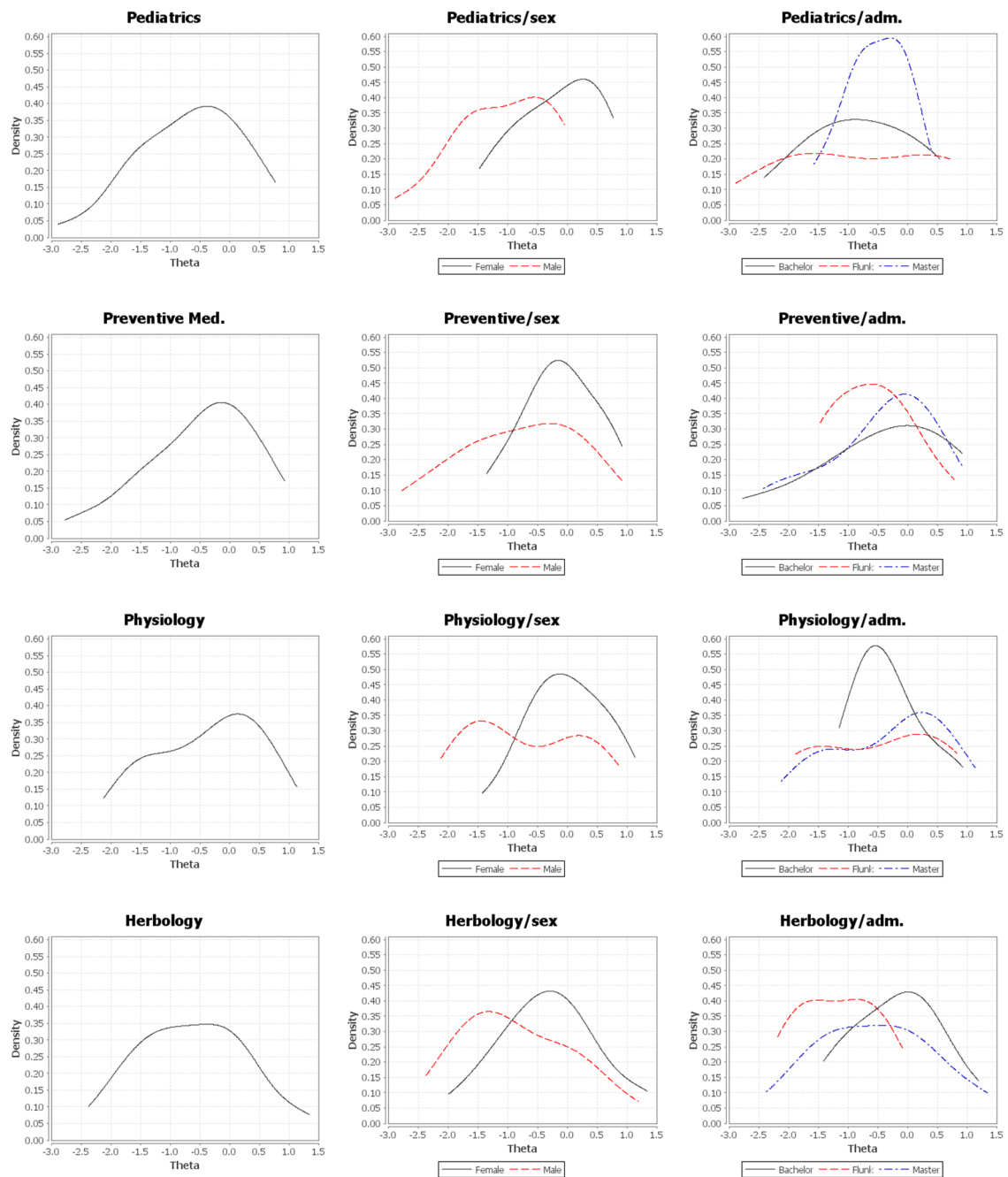


Figure 2. Kernel density map according to the sex and administrative groups for analyzing group-specific academic competency in four subjects of mock national license exam

하며, 특히 생리학에 있어서는 남학생 중에서 낮은 학업성취도 그룹에 더 많은 관심을 보여야 함을 의미하는 것이다.

학사행정 에 따른 역량수준별 분포는 과목마다 상이한 특성을 보였다. 소아과학에서는 석사(4년) 과정의 역량 수준이 더 높은 수준에 밀집되어 있었으나, 생리학에서는 학석사(7년) 과정의 역량수준이 중간에 밀집되어 있었고, 본초학에서는 유급/휴학과 학석사(7년) 과정 사이의 역량 수준의 차이가 보다 명료한 것을 확인할 수 있었다.

학사행정 그룹에 따른 맞춤 관리를 위하여 낮은(-1.5 이하) 역량수준에서의 학생 분포를 살펴보았는데, 이들은 교수자의 관심과 집중적인 학업관리가 필요한 학생이라고 할 수 있다. 분석 결과, 소아과학은 유급/휴학과 학석사(7년) 과정에, 예방의학은 학석사(7년) 및 석사(4년) 과정에, 생리학과 본초학은 유급/휴학과 석사(4년) 과정에 더 많은 학생들이 분포함을 확인할 수 있었다.

고찰 및 결론

본 연구에서는 IRT 학업 분석을 한의학 교육 현장의 행정 실무자와 내용 전문가가 일상적인 교육평가 현장에서 활용할 수 있도록, 실제 시행되었던 국가시험 모의고사를 활용하여 필수적인 IRT 분석 개념과 직관적인 분석 결과를 제시하였다. 이와 함께, 별도의 평가 측정분석 전문가의 도움이 부족하더라도 한의학 교수자가 직접 문항 및 시험의 타당도와 학습자의 그룹별 역량 특성을 직관적으로 이해하기 위한 현장 활용 가이드라인을 제시하고자 하였다.

본 연구에서 시험 및 문항의 타당도를 과목에 따른 역량수준별 진점수, 정보의 양, 표준오차, 그리고 라이트 지도(Figure 1)를 통해 입체적으로 분석하였으며, 평가의 목적에 따른 문항 개선 방향을 제시할 수 있었다. 이와 함께 커널 밀도함수(Figure 2)를 활용하여 학생 그룹에 따른 역량분포의 차이를 직관적으로 확인할 수 있었는데, 이러한 역량 특성은 맞춤 교육의 근거로 활용될 수 있다.

현재 한의학 교육을 개선하기 위한 노력이 다양한 측면에서 진행되고 있다. 그러나, 교육 이들 연구가 현장에서 실무자를 위한 가이드 라인이나 구체적인 쇼케이스의 제시보다는, 교육학적 담론, 학습자 만족도나 인식, 교육 과정이나 내용 제안 등 한의학 교육의 '지향점과 내용'에 편중되어 있다²¹. 학습자의 만족도를 높이기 위한 한의학 교육의 효율성, 효과성을 향상시키기 위해서는 학생 및 교육에 대한 타당도 높은 평가 및 환류 시스템의 도입이 필수적이며, 본 연구와 같은 근거중심 학습분석법(learning analytics)의 도입이 시급하다. 특히 한의학교육실의 역할에 대한 명료한 정립이 부족한 현실을 고려한다면²¹⁻²⁴, 기초학과 임상 및 한의사 국가시험에서의 다양한 학습분석 모델을 도입하는 것은 역량중심 한의학 교육을 위한 교육실의 주요 기능이라 할 것이다^{1,4,16,25}.

최근 한의계에서 추진되고 있는 컴퓨터기반시험(Computer-based Test, CBT)은 효율적이고 효과적인 교육을 위한 정보통신기술을 의미하는 에듀테크(EduTech)의 한 부분으로, 교육 현장에서 등록, 교육 및 평가와 같은 학습 전반을 효율적으로 관리하는 학습관리시스템(Learning Management System, LMS)의 일부로 활용되고 있다^{1,16}.

그러나, CBT 도입 본연의 목표인 학업성취도 증진보다, 기존의 종이기반 필기시험이 단순히 모니터로 옮겨진 외형적 특성만이 중시되고 있다. 학습자의 사용성과 편리성, 시험 환경 등에 대한 만족도를 주로 연구하거나, 출제, 인쇄, 보안, 시행, 채점 등에서의 비용과 시간을 중심으로 한 교수자의 행정적 편의 등이 CBT 도입의 가장 중요한 이유로 제시되어 왔다^{6,26,27}. 부산대학교 한의학전문대학원의 두 번에 걸친 CBT 시스템 도입과정에서도 학업분석을 위한 IRT 시스템은 탑재되

지 못하였는데, 이는 CBT 및 IRT에 대한 한의계의 전반적인 현실을 반영한다고 할 수 있을 것이다²⁸.

이에, 향후 CBT 시스템의 구매 및 업그레이드 과정에서 본 연구에서 제시된 최소한의 학습자 역량분석과 시험 타당도 분석이 포함되도록 요구하여야 할 것이며²⁸, 보유하고 있는 CBT 시스템에서 평가 결과를 외부에서 다운로드할 수 있다면 자체적인 학습분석 시스템을 새롭게 구축하여야 할 것이다.

이와 함께, 교육 현장에서 평가분석과 내용 교육을 담당하는 교수자들이 IRT를 직접 활용하면서 한의학 교육의 효율성과 효과성을 높이기 위한 현장 매뉴얼 작성과 보수교육도 함께 시행되어야 한다. 아무리 좋은 교육 제도 또는 시스템이라도 현장의 상황을 고려하여 손쉽게 시행될 수 없다면, 기대하는 성과를 얻을 수 없거나 또는 획일적인 분석 과정에서 역효과를 초래할 수 있기 때문이다.

예를 들어, 본 연구를 통해 제시된 과목별 문항분석 결과(Figure 1)에 있어서, 평가 목적이 특정 학업수준에서의 합/불합격의 결정과 같은 규준지향 평가 또는 의료인 면허를 위한 국가시험에 있다면 생리학과 본초학이 가장 적절하다 할 것이다. 그러나, 학습자의 현재 수준을 평가함으로써 완전학습과 자기개발을 목표로 하는 모의고사라는 측면을 더 강조한다면 역량수준 -5~+5를 고르게 포괄하는 소아과학과 예방의학이 더 적절하다 평가할 수도 있기 때문이다.

이와 함께, 학생 그룹별 역량분포의 분석(Figure 2)에 있어서도, 맥락과 상황을 고려한 저성취도 학습자의 스크리닝과 맞춤 관리를 중심으로 해석되어야 한다. 본 연구와 같은 단 한번의 시험 만을 근거로 '예방의학에서는 유급/휴학 그룹에 대한 관심이 불필요하다'라고 과도하게 일반화된 결론을 도출하거나, 생리학과 본초학을 근거로 '학석사(7년) 과정이 더 성취도가 높은 학제' 또는 소아과학을 근거로 '석사(4년) 과정이 더 성취도가 높은 학제'라는 획일적 주장을 제시한다면, 교육 현장에 대한 그릇된 인식이 고착화될 것이기 때문이다.

본 연구의 결과를 교육 현장에서 일반화함에 있어서는 다음과 같은 연구에서의 제한 사항이 고려되어야 할 것이다.

첫째, 본 연구는 한의사 국가시험 모의고사에 있어서 연구와 지면의 제약을 고려하여 4 과목에 대한 분석만을 제한적으로 진행하였다. 이에 12 과목 전부, 그리고 1 차와 2 차의 평가 결과를 동시에 고려하는 한의학교육실의 관점에 따른 추가 분석이 필요할 것이며, 이러한 과정에서 요구되는 추가적인 행정적 지원, 전산시스템, 소프트웨어 및 평가분석 전문가의 역할이 고려되어야 할 것이다.

둘째, 내용 전문가를 대상으로 하는 현장 수용성 검토 및 본 연구가 지향하는 교육학적 효과를 확인하기 위한 시범사업이 진행되어야 한다. 최신 학업분석 기법에 대한 교수자 연수의 효과성 분석, 그룹별 역량분석을 통한 교육 개선의 효과 분석,

교육실의 고유업무로 수행되기 위한 행정적 비용 분석, 그리고 시범 사업을 통한 교수자-학습자의 학교 만족도 개선효과 등의 분석 등을 토대로 한의학교육 현장에 적용할 수 있는 전문 학회의 권고안 마련 등이 진행되어야 할 것이다.

본 연구에서는 한의학교육 현장에 바로 활용하기위한 IRT 활용 학업역량 분석과 이를 근거로 한 한의학 교육의 개선에 대하여 고찰하였다. 이러한 현장중심 역량근거 한의학 교육은^{4,17,18}, 한의학 교육에 대한 학습자 만족도와 교수자 편의성 개선에 기여할 수 있을 것이다.

Acknowledgements

본 연구는 부산대학교의 연구비지원을 받았다

This work was supported by a 2-Year Research Grant of Pusan National University.

Conflict Of Interest

No potential conflict of interest relevant to this article was reported.

ORCID iD

채 한 <http://orcid.org/0000-0002-8698-8229>

이혜윤 <http://orcid.org/0000-0002-9486-1703>

한상윤 <http://orcid.org/0000-0002-6174-1078>

CRediT Statement

Conceptualization: HC and SYH; Methodology: HC; Software: HC; Validation: HC; Formal analysis: HC; Investigation: HC; Resources: HL; Data curation: HC, HL; Funding acquisition: HC; Project administration: HC and SYH; Writing-original draft: HC, HL and SYH; Writing-review & editing: HL and SYH; Visualization: HC; Supervision: SYH.

References

1. Chae H, Lee SJ, Han c-h, Cho YI, Kim HW. Study on the Academic Competency Assessment of Herbology Test using Rasch Model. *J Korean Med*. 2022;43:27-41. <https://doi.org/10.13048/jkm.22017>
2. Lee J, Han JH, Kim MS, Lee HS, Han SY, Lee SJ, et al. Teaching Yin-Yang biopsychology using the animation, "Pororo the Little Penguin". *European Journal of Integrative Medicine*. 2020;33:101037. <https://doi.org/10.1016/j.eujim.2019.101037>
3. Seong T. *Understanding and application of Item Response Theory*. 2nd ed. Paju: Educational Science Publishing; 2016.
4. Chae H, Han SY, Yang G, Kim H. Study on the herbology test items in Korean medicine education using Item Response Theory. *Kor J Herbology*. 2022;37:13-21. <https://doi.org/10.6116/kjh.2022.37.2.13>.
5. Seo M. *A Study on Designing for the Computer-based NAEA*. Jincheon-Gun: Korea Institute for Curriculum and Evaluation; 2020.12.31. 2020.
6. 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*: Korea Helath Personnel Licensing Examination Institute;2018.
7. Kim J. *A Study on Online Evaluation System Based on Individualized Education*. Daegu: Korea Education and Research Information Service 2018.
8. Lim EY. *Comparison of item analysis according to item response theory and classical test theory*. Korea Helath Personnel Licensing Examination Institute;2001. RE1-0103-00.
9. Yang E. Perceptions on item disclosure for the Korean medical licensing examination. *Korean J Med Educ*. 2015;27:167-175. <https://doi.org/10.3946/kjme.2015.27.3.167>
10. Baik S-H. Major Reforms and Issues of the Medical Licensing Examination Systems in Korea. *Korean Med Educ Rev*. 2013;15:125-135. <https://doi.org/10.17496/kmer.2013.15.3.125>
11. 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;19:163-169. <https://doi.org/10.3946/kjme.2007.19.2.163>
12. 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. *Journal of Health Science & Medical Technology*. 2005;31:55-73.
13. 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;2:113-118. <https://doi.org/10.3352/jeehp.2005.2.1.113>
14. 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. <https://doi.org/10.3352/jeehp.2006.3.4>
15. Choi K, Cho JK. Statistical analysis of national examination for radiological technologists in convergence perspective. *Journal of the Korea Convergence Society.* 2017;8:93-99. <https://doi.org/10.15207/JKCS.2017.8.7.093>
 16. Chae H, Cho E, Kim S, Choi D, Lee S. Analysis on validity and academic competency of mock test for Korean Medicine National Licensing Examination using Item Response Theory. *Keimyung Med J.* 2023;24:7-18. <https://doi.org/10.46308/kmj.2023.00094>
 17. 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;21:259-267. <https://doi.org/10.3946/kjme.2009.21.3.259>
 18. Schauber SK, Hecht M, Nouns ZM. Why assessment in medical education needs a solid foundation in modern test theory. *Adv in Health Sci Educ.* 2018;23:217-232. <https://doi.org/10.1007/s10459-017-9771-4>
 19. Park JC, Kim KS. A Comparison between Discrimination Indices and Item-Response Theory Using the Rasch Model in a Clinical Course Written Examination of a Medical School. *Korean J Med Educ.* 2012;24:15-21. <https://doi.org/10.3946/kjme.2012.24.1.15>
 20. Meyer JP. *Applied measurement with jMetrik.* New York: Routledge; 2014.
 21. Gwak D, Kim M, Kwon J, Yin C. Trends in Korean Medicine education research from 2003 to 2022. *The Journal of Korean Medical History.* 2023;36:89-98. <https://doi.org/10.15521/jkmh.2023.36.1.089>
 22. Mannadeurimnida D. From Teacher to Traditional Korean Medicine Education Center, Professor Han Sangyoon (1). *Brunch Story*2022.
 23. Han SY, Kim HY, Lim JH, Cheon J, Kwon YK, Kim H, et al. The past, present, and future of traditional medicine education in Korea. *Integr Med Res.* 2016;5:73-82. <https://doi.org/10.1016/j.imr.2016.03.003>
 24. Han SY, Lee S-H, Chae H. Developing a best practice framework for clinical competency education in the traditional East-Asian medicine curriculum. *BMC Medical Education.* 2022;22:352. <https://doi.org/10.1186/s12909-022-03398-4>
 25. Newspaper Ds. 'Future Engine' in sync with the Advancement of Medicine and Social Transformation <1>. *Doctor's Newspaper.* 2006.10.25, 2006.
 26. Kim SG, Lee A, Hwang I. Analysis of the Satisfaction with Computer Based Test Program and Test Environment in Medical School *Korean Med Educ Rev.* 2020;22:198-206. <https://doi.org/10.17496/kmer.2020.22.3.198>
 27. Im E-J, Lee W-K, Lee Y-C, Choe B-H, Chung S-K, Lee T-H, et al. Development of Computer-Based Test(CBT) and Student Recognition Survey on CBT. *Korean J Med Educ.* 2008;20:145-154.
 28. Chae H. Competency-Based Traditional Korean Medicine Education, Absolute Assessment, and Examination Analysis. *AKOM News.* 2022. 06. 02., 2022.