JDCS
디지털콘텐츠학회논문지
Journal of Digital Contents Society
Vol. 27, No. 10, pp. 2949-2962, Oct. 2026

Check for updates

Rasch 모형을 활용한 KEEPII AI 준비도 문항의 타당도 검토

조 윤 성*
현대해상화재보험주식회사 인재개발파트 책임연구원

An Investigation of the Validity of the AI Readiness Items in KEEPII Using the Rasch Measurement Model

Yunseong Jo*
Senior Researcher, Human Resource Development Team, Hyundai Marine & Fire Insurance, Seoul 03183, Korea

[요 약]

본 연구는 한국교육고용패널조사Ⅱ 8차년도에 포함된 AI 준비도 8개 문항에서 산출되는 점수의 측정 특성과 타당도 근거를 검토하였다. 이를 위해 AI 앱 사용 경험이 있는 청년 2,946명의 자료에 Rasch 평정척도모형을 적용하고, 내용·실질·구조·일반화 측면의 근거를 평가하였다. 분석 결과, 5번 문항은 부적합하였으나 해당 문항 제외 전후 응답자 측정치의 순위 변화는 미미하였다. 6개 응답범주는 대체로 적절히 기능했지만, 문항이 전반적으로 반응하기 쉬운 것으로 나타나 AI 준비도 상위 응답자를 변별하는 데 한계가 있었다. 다음으로 잔차 분석과 범주형 확인적 요인분석에서는 AI 활용과 AI 윤리·규범 이해 간 두 관련 차원이 확인되었다. 그러나 이중요인 지수는 총점의 조건부 사용을 지지한 반면, 특수요인의 고유 신뢰도는 낮아 하위점수의 독립적 사용은 지지되지 않았다. 일반화 측면을 확인하기 위한 성별 차별기능 문항은 3개 문항에서 통계적으로 유의했으나 효과크기는 모두 무시할 수준이었다. 이상을 바탕으로 총점 활용 조건과 중복 문항 수정 및 고난도 문항 추가를 제안하였다.

[Abstract]

This study examined validity evidence for the scores produced by the eight AI readiness items included in the eighth wave of KEEPII. The Rasch rating scale model was applied to data from 2,946 young adults with experience using AI applications, and evidence was examined across content, substantive, structural, and generalizability aspects. Item 5 showed misfit, although its removal produced little change in person-measure ordering. The six response categories generally functioned as intended, but the items were relatively easy and provided limited discrimination among highly prepared respondents. Residual analyses and categorical confirmatory factor analysis indicated two related dimensions: AI use and understanding of AI ethics and norms. Nevertheless, bifactor indices supported the conditional use of a total score, whereas the low unique reliability of the specific factors did not support separate subscale scores. Three items showed statistically significant gender-related differential item functioning, but all effects were negligible. These findings clarify the conditions for total-score use, and suggest revising overlapping items and adding more difficult items.

색인어 : AI 준비도, Rasch 평정척도모형, 타당도 근거, 이중요인모형, 한국교육고용패널조사II
Keyword : AI Readiness, Rasch Rating Scale Model, Validity Evidence, Bifactor Model, KEEPII

http://dx.doi.org/10.9728/dcs.2026.27.10.2949

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-CommercialLicense(http://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 03 August 2026; **Revised** 04 September 2026
Accepted 02 October 2026

***Corresponding Author; Yunseong Jo**
Tel: ※ 개인정보 표시제한
E-mail: jopry@hanmail.net

Ⅰ. 서 론

생성형 AI를 비롯한 인공지능 기술이 일의 방식과 노동시장을 근본적으로 재편할 것으로 전망되는 가운데[1], AI 활용 격차는 새로운 사회적 불평등의 축으로 부상하고 있다[2]. 특히 AI가 학습과 일의 방식을 전반적으로 바꾸어 놓는 과도기적 시기를 경험하고 있는 청년 세대에게 이 격차의 영향은 장기간 누적될 수 있다. 따라서 청년의 AI 활용 수준을 정확하게 파악하는 것은 디지털 격차를 넘어 사회경제적 격차 완화를 위한 교육적·정책적 개입의 출발점이 된다.

한편, AI 활용의 격차가 실제 AI 활용의 결과로서 관찰되는 것이라면, 격차를 완화하기 위한 전략적 개입은 결과의 확인보다는 개인이 AI를 활용하는 데 필요한 요건을 얼마나 갖추고 준비하고 있는가를 진단하는 데서 출발할 필요가 있다. AI 준비 수준에 대한 심층적 진단이 없는 경우 개입이 필요한 지점과 대상을 특정하기 어렵고 개입의 효과를 점검할 근거 또한 확보하기 어렵기 때문이다[3]. 이와 관련하여 AI 준비도(AI readiness)는 개인의 AI 활용 준비 수준을 포착하는 대표적인 개념이다. AI 준비도는 AI 활용에 요구되는 인지, 능력, 비전, 윤리를 갖춘 준비 상태로 개념화된다[4]. 특히 노동시장 이행과 경력 형성의 초기 단계에 있는 청년의 AI 준비도를 명확히 파악하는 것은 AI 활용 격차가 향후 사회경제적 격차로 확대되기 전에 개입할 수 있는 진단 근거를 제공한다는 점에서 중요하다.

이러한 AI 준비도는 유관 개념들과 구별되는 특성을 지닌다. 예컨대 AI 리터러시가 AI를 이해하고 비판적으로 평가하며 윤리적으로 사용하는 데 필요한 앎에 초점을 둔다면[5],[6], AI 활용 역량은 이러한 지식과 기술을 실제 문제해결과 과업 수행 등 다양한 상황에 사용하고 적용하는 능력을 의미한다[6],[7]. 이에 비해 AI 준비도는 특정 능력의 보유 여부를 넘어 활용에 요구되는 인지적 이해와 수행 능력, 그리고 규범적 인식까지 갖추어진 상태를 포괄한다. 이를 반영하듯 AI 준비도는 능력 요소와 인지·태도 요소를 함께 포함하는 다차원 구인으로, 응답자가 지각한 준비 수준에 대한 자기보고 방식으로 측정되어 왔다[8]. 요컨대 AI 준비도는 AI 활용 능력을 포함하되 이를 넘어서는 준비 요소까지 아우른다는 점에서[4],[8] 능력 자체에 초점을 두는 AI 활용 역량과 구별되며, AI 활용과 성과에 선행하는 조건의 성격을 갖는다. 실제로 선행연구에 따르면, 준비도에 대한 정확한 평가가 AI 도입의 성공 가능성을 높이는 조건임이 규명되었으며[9], AI 준비도가 청년 취업자의 디지털 기술 활용 수준과 직무만족도의 관계를 매개하는 요인으로도 실증된 바 있다[10].

AI 준비도 측정 및 진단의 중요성이 강조됨에 따라 실제 AI 준비도를 측정하기 위한 척도의 개발과 타당화 연구가 지속적으로 진행되어 왔다. 예컨대 대학생을 대상으로 인지, 능력, 비전, 윤리 등 4개 하위요인, 22개 문항으로 구성된 AI 준비도 척도 개발 연구가 수행되었다[8]. 이후 해당 척도는 다양한 나라의 언어로 번안·타당화를 거쳐 전 세계적으로 활용되고 있다[3],[11]. 국내의 경우 AI에 관한 인식, 태도, 이용의도를 종합하여 AI 준비도 수준을 측정하거나[12], AI 인식, 능력, 태도, 전망으로 구성된 별도의 AI 준비도 척도를 개발한 연구[13]가 진행된 바 있다. 이 밖에도 대학생[14], 교사[15], 직장인[16] 등 개인 수준뿐만 아니라 기업[17] 그리고 국가 수준[18]에서의 AI 준비도 측정 및 효과에 관한 연구가 활발하게 이루어지고 있다.

그러나 지금까지의 AI 준비도 척도 검증은 방법론적으로 특정 접근에 편중되어 있다. 원 척도의 개발[8]과 이후의 번안·타당화 연구[3],[11]는 모두 탐색적·확인적 요인분석과 내적일치도 계수를 중심으로 하는 고전검사이론에 기반하고 있으며, 문항 수준의 기능과 응답범주의 작동 방식을 검토한 문항반응이론 기반 검증은 확인되지 않는다. 이러한 편중은 실제적 문제로 이어질 수 있다. 예컨대 중국어판 타당화 연구에서는 원 척도의 인지 요인에 속하던 두 문항이 능력 요인으로 재배치되었는데[3], 이는 동일한 문항이라도 대상과 맥락에 따라 차원 배정이 달라질 수 있음을 보여준다. 따라서 원 척도에서 파생되거나 별도로 구성된 단축형 문항을 사용할 경우, 개발 단계에서 상정한 구조가 그대로 유지된다고 전제하기보다 해당 자료에서 구조를 직접 확인하는 추가 절차가 요구된다. 특히 문항 수가 적은 척도에서는 높은 내적일치도가 단일차원성을 보장하지 않으며 문항 간 국소의존이 신뢰도와 요인구조를 과대평가하게 만들 수 있다[19],[20]. 이에 본 연구는 문항반응이론에 기반한 문항 적합도 검토에 더하여 잔차 기반 차원성, 국소독립성, 범주형 확인적 요인분석 및 이중요인 지수를 종합적으로 검토한다는 점에서 고전검사이론 중심의 기존 타당화 연구와 구별된다.

한편, 한국교육고용패널조사Ⅱ(KEEPⅡ)는 총 8개 문항으로 구성된 AI 준비도 문항을 도입하여 대규모 표본을 바탕으로 청년의 AI 준비도를 측정하고 있으며, 이를 활용한 연구가 활발히 진행되고 있다[10],[21]. 다만 해당 8개 문항은 인지, 능력, 비전, 윤리의 네 영역을 포괄하는 원 척도[8]와 달리 AI 활용 수행과 AI 윤리·규범 이해에 해당하는 내용으로 구성되어 있다. 즉 원 구인 가운데 비전에 해당하는 전망적 인식은 포함되지 않으며 인지와 윤리 영역도 부분적으로만 반영된다. 따라서 본 척도로 산출되는 점수는 AI 준비도 구인 전체가 아니라 그중 활용과 규범 이해에 한정된 범위를 포착하는 것으로 이해할 필요가 있다. 그럼에도 해당 척도는 국가 수준에서 산출되는 측정치라는 점에서 향후 청년의 AI 격차 진단과 교육적·정책적 전략 수립을 위한 실증 연구에 지속적으로 활용될 것으로 예상된다.

다만 제기될 수 있는 문제는 해당 8개 문항에서 산출되는 점수의 해석과 활용을 뒷받침할 측정학적 근거가 충분히 제시되지 않은 상태에서 이미 실증 연구에 투입되고 있다는 점이다. 선행연구[10],[21]는 8개 문항을 하나의 총점 또는 평균으로 합산하여 사용하고 있으나 이러한 합산이 정당화되는지에 대한 측정 수준의 근거는 제시된 바 없다. 타당성이 확

인되지 않은 척도로 산출된 점수는 응답자의 실제 AI 준비도 수준을 과대 또는 과소 추정할 수 있으며, 문항군이 서로 다른 차원을 형성하는 경우 단일 총점은 이질적인 정보를 하나의 값으로 단순 처리하게 된다. 또한 특정 집단에 유리하게 기능하는 문항이 존재한다면 집단 간 점수 차이가 실제 준비도 차이인지 문항 편향의 결과인지 구분할 수 없다. 이러한 심층적인 타당성을 검증하기 위한 대표적인 방안으로 Rasch 측정모형 적용을 고려할 수 있다. Rasch 측정모형은 문항특성의 불변성과 피험자 능력의 불변성을 유지하여 개별 문항의 기능과 응답자의 능력을 명확히 추정할 수 있다는 장점이 있다[22]. 실제로 다양한 심리·행동 척도[23],[24]뿐만 아니라 국가 패널조사에 수록된 척도의 타당성을 검증하기 위하여 Rasch 측정모형을 활용한 연구가 활발하게 진행되고 있다[25]-[27].

이상의 맥락을 바탕으로 본 연구는 KEEPⅡ 8차년도 조사에 수록된 AI 준비도 8개 문항을 대상으로 Rasch 측정모형을 활용하여 해당 문항에서 산출되는 점수의 측정특성과 타당도 근거를 검토하는 것을 목적으로 한다. 구체적으로는 Wolfe와 Smith가 제안한[28] 준거에 따라 내용 측면에서 개별 문항의 적합도와 문항–측정치 상관, 실질 측면에서 응답범주의 기능성과 응답자 능력 분포와의 정합성, 구조 측면에서 척도의 차원성과 문항 간 국소독립성, 일반화 측면에서 성별에 따른 측정의 불변성을 검토한다. 특히 구조 측면에서는 차원성 확인에 그치지 않고, 확인된 구조가 8개 문항의 총점 사용을 지지하는지를 함께 검토함으로써 후속 연구에서의 활용 조건을 제시하고자 한다. 이를 위한 연구 문제는 다음과 같다. KEEPⅡ AI 준비도 8개 문항에서 산출되는 점수는

첫째, 문항 기능 및 내용 측면에서 어떠한 타당도 근거를 보이는가?
둘째, 실질 측면에서 어떠한 타당도 근거를 보이는가?
셋째, 구조 측면에서 어떠한 타당도 근거를 보이는가?
넷째, 일반화 측면에서 어떠한 타당도 근거를 보이는가?

Ⅱ. 연구방법

2-1 자료수집 및 연구대상

본 연구는 한국직업능력연구원에서 실시한 한국교육고용패널조사Ⅱ 8차년도 원시자료를 활용하였다. KEEPⅡ는 청년층의 학교교육 및 일 경험, 진로·직업에 대한 인식과 가치관 변화, 학교에서 노동시장으로의 이행 과정을 종단적으로 파악하기 위해 구축된 패널조사이다. 2016년 고등학교 2학년 학생 10,558명을 원패널로 구축한 이후 동일한 패널을 지속적으로 추적하고 있다. 8차년도 조사는 유효표본 8,772명 중 7,521명이 조사에 참여하여 85.7%의 조사 성공률을 보였다[29]. 분석 대상의 선정은 두 단계로 이루어졌다. 먼저 8차년도 조사에 참여한 7,521명 중 “AI 앱 또는 프로그램을 사용해 본 경험이 있다”고 응답한 2,951명(39.2%)을 1차 선정하였으며, 이 가운데 AI 준비도 8개 문항에 모두 응답한 2,946명을 최종 분석 대상으로 확정하였다. 한편 AI 사용 경험이 없다고 응답한 4,570명(60.8%)은 조사 설계상 AI 활용에 해당하는 1번부터 5번 문항에 응답하지 않았으므로 8개 문항 전체를 대상으로 하는 본 연구의 분석에서 제외되었다. 다만 AI 윤리·규범 이해에 해당하는 6번부터 8번 문항은 AI 사용 경험 여부와 무관하게 전체 응답자에게 조사되었다. 1차 선정 이후 문항 무응답으로 제외된 사례는 5명이었다.

이상과 같은 선정 절차로 인해 분석 표본이 AI 사용 경험자에 한정되므로 표본 구성의 차이를 확인하기 위해 전체 조사응답자 7,521명을 AI 사용 경험 집단과 미경험 집단으로 구분하여 성별과 최종 학력 분포를 비교하였다. 먼저 성별 분포는 두 집단 간에 유의한 차이가 없었다($\chi^2(1)=0.583$, $p=.445$, Cramer's V=.009). 최종 학력은 중학교 졸업으로 응답한 1명을 고등학교 졸업 범주에 포함하여 고등학교 졸업, 전문학사, 학사 및 석사의 네 범주로 재구성하였다. 분석 결과, 최종 학력 분포는 두 집단 간에 유의한 차이가 있었으나 효과크기는 작았다($\chi^2(3)=85.223$, $p<.001$, Cramer’s V=.106). 구체적으로 AI 사용 경험 집단은 미경험 집단보다 학사 비율이 높고 전문학사 비율이 낮았다. 따라서 본 연구의 분석 표본은 AI 미사용자와 성별 구성에서는 차이가 없었으나 학사 비율이 상대적으로 높은 특성이 있으므로 연구 결과를 전체 청년층이나 AI 미사용자에게 일반화할 때 이러한 표본 특성을 고려할 필요가 있다.

최종 분석 대상의 인구통계학적 변인에 따른 분포를 살펴보면, 먼저 성별은 남성 1,591명(54.0%), 여성 1,355명(46.0%)이었다. 최종 학력은 고등학교 졸업 1,295명(44.0%), 전문학사 390명(13.2%), 학사 1,238명(42.0%), 석사 23명(0.8%)이었다. 이때 전문대학 학사학위 전공심화과정 졸업자 37명은 학사학위 취득자에 해당하므로 학사 범주에 포함하였으며, 앞서 학력 분포 비교에도 동일한 기준을 적용하였다. 대학교 전공계열의 경우 공학계열이 821명(33.2%)으로 가장 많았으며, 사회계열 543명(22.0%), 예체능계열 284명(11.5%), 자연계열 255명(10.3%), 인문계열 222명(9.0%), 의학계열 216명(8.7%), 교육계열 129명(5.2%)의 순으로 나타났다. 나머지 476명(16.2%)은 전공계열 정보가 없거나 해당되지 않았다.

2-2 연구도구

본 연구에서는 KEEPⅡ 8차년도에 포함된 AI 준비도 8문항을 분석하였다. 해당 문항 중 1번부터 5번 문항은 문제해결에 적합한 AI 앱 선택, AI 정보와 기존 지식의 결합, AI의 문제해결 절차 설명 능력, AI를 활용한 성과 향상 및 여가활동에서의 AI 활용 등 AI 활용 수행 수준을 측정한다. 6번부터 8

번 문항은 AI 기술의 강점과 한계, 윤리적 의무 및 법적 책임에 대한 이해 등 AI 윤리·규범 이해 수준을 측정한다. 각 문항의 구체적인 내용은 표 1과 같으며, 문항의 국문 원문은 한국직업능력연구원이 공개한 KEEPⅡ 8차년도 조사 원자료와 설문지에서 확인할 수 있다.

표 1. 문항 정보
Table 1. Item information

Item Label	Contents
AIR1_SEL	I can select an AI application or program that can solve a problem assigned to me.
AIR2_INT	I can combine and use information provided by AI with my existing knowledge.
AIR3_PRC	I can explain the process by which AI provides an answer to a problem.
AIR4_PRF	I can improve my performance when carrying out work or assignments with the help of AI applications or programs.
AIR5_CRT	I can use AI applications or programs for entertainment or hobbies, such as image creation, video editing, or music composition.
AIR6_LIM	I know the strengths and limitations of AI technology.
AIR7_ETH	I know the ethical responsibilities required in the AI era.
AIR8_LEG	I know the legal responsibilities that should be observed when using AI technology.

해당 8개 문항은 인지, 능력, 비전, 윤리의 네 영역으로 구성된 원 척도[8]와 달리 활용 수행과 윤리·규범 이해에 해당하는 내용으로 구성되어 있으며 전망적 인식에 해당하는 비전 영역은 포함하지 않는다. 아울러 각 문항은 응답자의 객관적 수행이나 지식을 측정하는 것이 아니라 스스로 지각한 수준을 보고하도록 구성되어 있다. 이에 본 연구에서 산출되는 값은 실제 AI 준비도 능력이 아니라 자기보고에 기초한 AI 준비도 수준의 추정치로 해석하였다.

실제로 본 연구에서는 조사표의 구성에 따라 1번부터 5번까지를 AI 활용, 6번부터 8번까지를 AI 윤리·규범 이해로 구분하였다. 구체적으로 두 문항군은 조사표에서 별도로 배치되어 있으며, 문두도 각각 “귀하는 다음을 수행할 수 있습니까?”와 “귀하는 다음에 대해 알고 있습니까?”로 구분되어 있다. 아울러 두 문항군은 문항의 제시 대상도 다르다. 1번부터 5번 문항은 AI 사용 경험자에게만 제시된 반면, 6번부터 8번 문항은 AI 사용 경험 여부와 무관하게 전체 응답자에게 제시되었다. 이는 조사 설계 단계에서 두 문항군이 서로 다른 성격의 문항으로 취급되었음을 시사한다. 다만 조사 설계 단계에서 두 문항군을 이론적 하위요인으로 명명하거나 별도의 하위척도로 규정한 근거는 확인되지 않았다. 따라서 본 연구에서는 이 구분을 확정된 요인구조가 아니라 검증의 대상이 되는 잠정적 구분으로 취급하였으며, 실제 자료에서 두 문항군이 구별되는지 여부는 본문 3-4절의 구조 측면 검토를 통해 확인하였다. 각 문항은 ‘전혀 그렇지 않다(1점)’에서 ‘매우 그렇다(6점)’ 까지 6점 척도로 측정되었다. 해당 척도는 ‘보통이다’에 해당하는 중립 범주를 포함하지 않으며 점수가 높을수록 응답자가 지각하는 AI 준비도 수준이 높음을 의미한다.

2-3 분석방법 및 연구절차

Rasch 측정모형은 문항반응이론 유형 중 하나로 사회적, 심리적 변인의 근본적인 측정방법에 대한 이론적·수학적 설명이며, 따라서 과학적이고 엄격한 측정에 관한 처방적인 안내를 제공한다[19]. 실제로 해당 모형은 문항 난이도, 문항의 질, 응답자의 능력 분포 등에 대하여 검사 원점수와 관련하여 수학적으로 혼동되는 고전검사이론의 단점을 극복하고 응답자 능력 및 문항 난이도 등에 대한 불변적인 측정치를 제공함으로써 검사 도구에 대한 실제적인 심리측정학적 특성을 파악할 수 있는 장점이 있다[30]. 본 연구에서는 AI 준비도 척도에 대한 Rasch 측정모형을 적용하기 위해 Rasch 평정척도 모형을 적용하였다[31]. 해당 모형식은 아래와 같다.

$$\ln\left[\frac{P(X_{ni}=x)}{P(X_{ni}=x-1)}\right]=\beta_n-\delta_i-\tau_x \quad (1)$$

해당 수식에서 $P(X_{ni}=x)$는 응답자 n이 특정 문항 i의 응답범주 x를 선택할 확률을 의미하며, β_n은 응답자 n의 측정치(또는 특정 평정에 대한 응답자의 인식), δ_i는 문항 i의 위치(endorsability), τ_x는 $x-1$ 범주와 x 범주 간 경계모수를 의미한다.

한편, Messick은[32] 체계적인 타당성 평가 차원으로서 여섯 가지 영역을 제안하였으며, 이후 해당 기준은 미국교육연구협회(AERA), 미국심리학회(APA), 그리고 교육측정학회(NCME)에서 권고하는 타당화 연구의 표준 기준으로 자리매김하였다[33]. Wolfe와 Smith는 해당 타당성 평가 차원을 바탕으로 Rasch 측정모형을 적용한 타당화 검증 방안을 제안하였으며[28], 본 연구에서도 해당 타당도 근거 검토 절차에 따라 분석을 진행하였다. 해당 논의에서 제시한 타당성 유형별 Rasch 측정모형 적용 및 검토 내용은 다음과 같다.

첫째, 내용 측면의 타당도(The content aspect of validity)는 척도를 구성하는 개별 문항 기능 및 내용의 질과 관련된다. 이를 평가하기 위해 각 문항별 적합도와 문항–측정치 상관(point-measure correlation)을 분석하였다. 문항 적합도 지수는 문항에 대한 응답자의 실제 반응이 Rasch 모형에서 도출된 기댓값에 부합하는 정도를 나타내는 지표이다. 그 중 내적합도(Infit MNSQ)는 정보가중 적합도 지수로서 응답자의 측정치와 문항 곤란도가 가까운 구간에서 나타나는 비예측적 반응 패턴에 민감하며, 외적합도(Outfit MNSQ)는 응답자 수준과 문항 곤란도가 멀리 떨어진 구간에서 나타나는 극단적이고 예상 밖의 반응에 민감하다[34]. 리커트 척도의 경우 내적합도와 외적합도 값이 0.6~1.4 범위를 벗어나면 부적절한 것으로 판단한다[35]. 다만 본 연구의 표본은

2,946명으로 규모가 크기 때문에 표준화 적합도 지수(ZSTD)는 사소한 편차에도 과민하게 반응한다. 이에 MNSQ를 1차 판정 기준으로 삼고 ZSTD를 참고 지표로 병기하되, 기준을 상회하는 부적합은 측정의 왜곡으로, 기준을 하회하는 과적합은 문항 간 중복이나 예측 가능성의 문제로 구분하여 해석하였다. 문항-측정치 상관은 각 문항의 관찰 응답과 Rasch 모형에서 추정된 응답자 측정치 간의 Pearson 상관으로, 문항의 점수 방향이 전체 잠재특성의 방향과 일치하는지를 확인하는 지표이다. 값이 낮거나 음의 값을 보이는 경우 응답 방향의 오류, 해당 문항이 다른 구인을 측정할 가능성, 채점 및 코딩 오류, 응답자의 비일관적 반응 등 복수의 원인이 존재할 수 있다[20]. 아울러 문항 적합도 해석을 보완하기 위하여 각 문항의 관찰 응답 일치율과 Rasch 모형에서 산출된 기대 일치율을 비교하였다.

둘째, 실질 측면의 타당도(The substantive aspect of validity)는 문항 내용 및 응답자 반응의 관련성 정도를 의미한다. 이는 평정척도기능 분석과 응답자–문항 측정치 분포 비교를 통해 확인할 수 있다. 평정척도기능 분석은 척도에 응답한 응답값이 애초에 설계한 대로 잘 기능하는가와 관련 있다. 본 연구에서는 Linacre가 제시한[36] 여섯 가지 조건, 즉 범주별 관찰빈도가 10개 이상일 것, 범주별 평균 측정치가 단조 증가할 것, 외적합도 지수가 2.0 미만일 것, 경계모수가 단조 증가할 것, 인접 경계모수 간 간격이 1.4에서 5.0 logit 사이일 것, 각 범주의 확률곡선이 최고점을 형성할 것을 판단 기준으로 적용하였다. 다음으로 문항 난이도가 응답자의 자기보고 AI 준비도 수준에 적절히 대응하는지 확인하기 위해 응답자와 문항의 측정치 분포를 비교하였다. 구체적으로 응답자 평균 측정치와 문항 평균 측정치의 차이, 최고난도 문항보다 높은 측정치를 보인 응답자의 비율, 극단응답자의 수와 비율을 산출하였다.

셋째, 구조 측면의 타당도(The structural aspect of validity)는 실제 측정결과를 통해 나타난 척도의 내적 구조와 애초에 의도한 구인의 구조와의 일치 정도를 의미한다. 본 연구에서는 다음의 순서로 검토를 진행하였다. 먼저 표준화 잔차의 주성분분석을 통해 Rasch 측정치가 설명한 분산의 비율과 잔차 제1대조의 고유값을 확인하였으며, 고유값이 2.0 미만일 때 단일차원성을 지지하는 것으로 판단하였다[19]. 다음으로 제1대조에 대한 문항별 적재량과 문항 군집을 확인하고, 분리된 군집으로 산출된 응답자 측정치 간 상관 및 비감쇠 상관을 검토하였다. 이어서 문항 간 표준화 잔차상관을 산출하여 국소독립성을 확인하였으며, 절댓값 0.30 이상인 문항쌍은 국소의존의 가능성이 있는 것으로 판단하였다[20]. 단일차원성 또는 국소독립성이 충족되지 않는 경우의 후속 분석으로는 하위 문항군별 Rasch 분석과 범주형 확인적 요인분석을 계획하였다.

넷째, 일반화 측면의 타당도(The generalizability aspect of validity)는 측정 결과가 다른 맥락에서도 동일하게 유지되는가와 관련되어 있다. Rasch 측정모형에서는 차별기능문항 분석을 통해 일반화 타당성에 대한 근거를 확인할 수 있다. 차별기능문항이란 특정 문항이 둘 이상의 집단에서 서로 다른 방식으로 측정되는 것을 의미한다. 측정 결과의 일반화 가능성이 지지되려면 서로 다른 집단에서 문항 측정값의 차이가 실질적으로 의미 있는 수준이어서는 안 된다[20]. 본 연구에서는 이를 위해 성별에 따른 차별기능문항 분석을 수행하였다. 실제로 AI 준비도의 성별 차이는 남성의 점수가 높게 보고된 연구[37]와 유의한 차이가 확인되지 않은 연구[38]가 혼재되어 있어, 관찰된 차이가 실제 준비도의 차이인지 문항이 성별에 따라 다르게 기능한 결과인지는 문항 수준의 검증 없이 판별하기 어렵다. 이와 같은 이유에서 기존 AI 준비도 척도 개발 연구에서도 성별을 측정 불변성 검증을 위한 대상 변인으로 설정하였으며[8], 따라서 본 연구에서도 성별을 차별기능문항 분석을 위한 변인으로 설정하였다. 판정은 통계적 유의성과 실질적 크기를 구분하여 실시하였다. 통계적 유의성은 Rasch-Welch t 검정으로 확인하되 8개 문항에 대한 다중비교를 고려하여 Bonferroni 보정을 적용하였으며, 실질적 크기는 차별기능문항 대비값의 절댓값을 기준으로 0.43 미만은 무시할 수준, 0.43 이상 0.64 미만은 중간 수준, 0.64 이상은 큰 수준으로 분류하였다[39].

이상의 네 측면에 대한 검토에 더하여 세 가지 보조분석을 실시하였다. 첫째, 내용 측면에서 부적합이 확인된 문항이 척도 전체의 측정 속성에 미치는 영향을 확인하기 위하여 해당 문항을 포함한 모형과 제외한 모형의 신뢰도, 분리도, 설명분산, 잔차 제1대조 고유값 및 응답자 측정치를 비교하는 민감도 분석을 실시하였다. 둘째, 구조 측면에서 확인된 문항 군집의 구조를 검토하기 위하여 단일차원 모형과 상관 2요인 모형에 대한 범주형 확인적 요인분석을 실시하였다. 문항이 6점 서열척도로 측정되었으므로 WLSMV 추정법을 적용하였으며 모형 간 차이는 Muthén과 Muthén이 제안한 DIFFTEST 절차를 통해 검정하였다[40]. 또한 국소의존이 확인된 문항쌍의 잔차공분산을 자유추정한 모형을 추가로 검토하였다. 셋째, 확인된 구조가 8개 문항 총점의 사용을 지지하는지 판단하기 위하여 일반요인과 두 특수요인을 설정한 이중요인 모형을 추정하고, Rodriguez 등의 산출 절차에 따라[41], 설명공통분산(ECV; explained common variance), 오염되지 않은 상관의 비율(PUC; percentage of uncontaminated correlations), 총점의 omega hierarchical(ωH) 및 하위척도의 omega hierarchical subscale(ωHS)을 산출하였다. 판정에는 Reise 등이 제안한 기준, 즉 PUC가 .80 미만인 경우 ECV가 .60을 초과하고 ωH가 .70을 초과하면 다차원성이 존재하더라도 본질적 단일차원으로 해석할 수 있다는 기준을 적용하였다[42].

한편 KEEP Ⅱ는 층화표집에 기초한 종단 패널조사로서 조사 가중치를 제공한다. Rasch 측정모형은 문항모수의 표본불변성을 전제로 하며[19], 본 연구의 목적은 모집단 수준의 평균 추정보다 문항 수준의 측정 속성을 검토하는 데 있다. 이에 본 연구에서는 조사 가중치를 적용하지 않았다. 이상의

분석을 진행하기 위해 Rasch 분석에는 Winsteps 3.91.0을, 범주형 확인적 요인분석 및 이중요인 분석에는 Mplus 8.5를 활용하였다. 아울러 AI 사용 경험 여부에 따른 표본의 선택편향 검토에는 교차분석을 적용하였다. 이상의 분석 절차와 타당도 근거의 종합적 판단 과정을 아래 그림 1에 제시하였다.

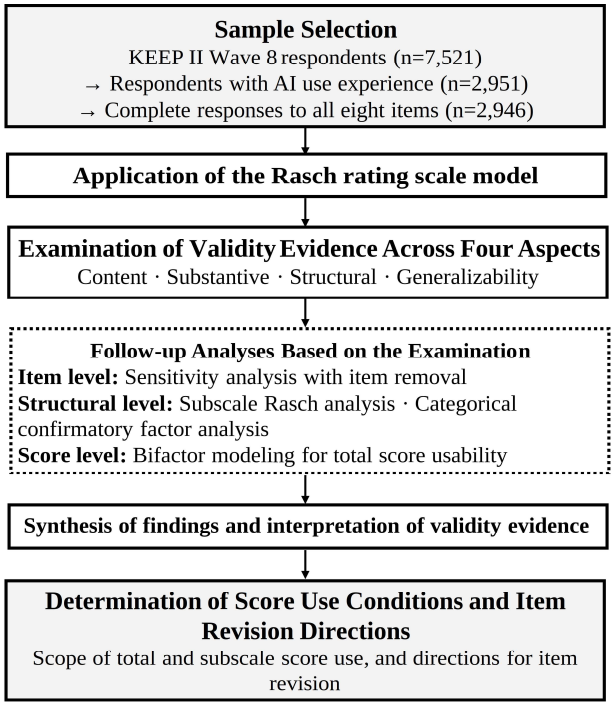


그림 1. KEEP Ⅱ AI 준비도 8문항의 타당도 근거 검토 절차
Fig. 1. Procedure for examining validity evidence for the eight AI readiness items in KEEPII

Ⅲ. 연구결과

3-1 측정의 요약 통계

Wolfe와 Smith의 준거에 따른 타당도 검토에 앞서 척도 전반의 측정 속성을 확인하였으며 그 결과는 표 2와 같다. 응답자 신뢰도는 실제 추정치 기준 .87, 모형 추정치 기준 .90이었으며, 응답자 분리도는 각각 2.59와 2.97로 나타났다. 이는 본 척도가 응답자를 약 네 개 수준으로 변별할 수 있음을 의미한다. 문항 신뢰도는 1.00, 문항 분리도는 21.15로 나타났는데, 이는 충분한 표본 규모(n=2,946)를 바탕으로 문항 난이도가 높은 정밀도로 추정되었음을 의미한다. 측정의 표준오차는 응답자의 경우 실제 추정치 기준 0.64 logit, 문항의 경우 0.03 logit이었다. 한편 8개 문항 모두에서 최고 응답범주를 선택한 극단 응답자는 55명(1.9%)이었으며, 최저 응답범주만을 선택한 극단 응답자는 없었다. 이들을 제외한 비극단 응답자 2,891명을 기준으로 한 응답자 신뢰도는 .86, 분리도는 2.46이었다. 원점수를 기준으로 산출한 Cronbach's α는 .89였으며, 원점수와 측정치 간 상관은 .96으로 나타났다.

표 2. 측정의 요약 통계
Table 2. Summary statistics of person and item measures

Statistic	Person (all)	Person (non-extreme)	Item
n	2,946	2,891	8
Mean measure	1.09	.97	.00
P.SD	1.77	1.57	.55
Real RMSE	.64	.59	.03
Real separation	2.59	2.46	21.15
Real reliability	.87	.86	1.00
Model separation	2.97	2.91	21.83
Model reliability	.90	.89	1.00
Maximum extreme scores	55(1.9%)	–	–
Minimum extreme scores	0(0.0%)	–	–

*n=2,946 persons and 8 items. Non-extreme statistics are based on 2,891 persons. RMSE = root mean square standard error. Cronbach's α = .89; correlation between raw scores and measures = .96.

3-2 문항 기능 및 내용 관련 경험적 타당도 근거 검토 결과

내용 측면의 타당도 근거 검토 결과는 표 3과 같다. 먼저 문항별 측정치인 난이도를 살펴보면, 8번 문항("나는 AI 기술을 사용하는 과정에서 지켜야 할 법적 책임을 알고 있다.")이 0.84 logit으로 가장 높아 응답하기 가장 어려운 문항으로 나타났다. 반면 1번 문항("나는 내게 주어진 문제를 해결할 수 있는 AI 앱(프로그램)을 선택할 수 있다.")은 -0.75 logit으로 가장 낮아 응답하기 가장 쉬운 문항으로 나타났다. 8개 문항의 최대-최소 난이도 범위는 1.59 logit이었다.

다음으로 문항별 내적합도와 외적합도를 확인하였다. 그 결과 5번 문항("나는 오락이나 취미활동을 위해 AI 앱(프로그램)을 활용할 수 있다.")은 내적합도 1.41, 외적합도 1.53으로 판단 기준인 0.6~1.4를 벗어났다. 2번 문항("나는 AI가 제공하는 정보와 내가 가진 지식을 결합하여 활용할 수 있다.")은 내적합도 0.71, 외적합도 0.72로 기준 범위 이내였으나 8개 문항 중 가장 낮은 값을 보였다. 나머지 6개 문항의 적합도 지수는 0.79~1.17의 범위에 분포하였다.

한편 표준화 적합도 지수(ZSTD)는 5번 문항이 9.9, 2번 문항이 -9.9로 나타나는 등 8개 문항 중 7개 문항에서 절댓값 2를 초과하였다. 이는 표본 규모가 큰 조건에서 표준화 적합도 지수가 작은 편차에도 민감하게 반응한 결과로 볼 수 있다. 이에 본 연구에서는 평균제곱 적합도(MNSQ)를 기준으로 문항 적합도를 판정하고 표준화 적합도 지수는 참고 지표로 활용하였다. 또한 적합도 해석을 보완하기 위해 응답 일치율을 확인한 결과, 5번 문항의 관찰 일치율은 47.9%로 모형 기대 일치율 48.0%와 유사하였다. 반면 2번 문항의 관찰 일치율은 63.0%로 기대 일치율 55.5%를 상회하여, 응답이 모형

표 3. 문항 적합도 분석 결과
Table 3. Item fit statistics

Item	Measure	SE	Infit MNSQ (ZSTD)	Outfit MNSQ (ZSTD)	rpm
AIR1_SEL	−0.75	0.03	0.99 (−0.2)	0.99 (−0.2)	0.66
AIR2_INT	−0.71	0.03	0.71 (−9.9)	0.72 (−9.9)	0.75
AIR3_PRC	0.07	0.02	0.93 (−2.6)	0.95 (−1.7)	0.75
AIR4_PRF	−0.37	0.03	0.79 (−8.1)	0.79 (−8.4)	0.76
AIR5_CRT	0.67	0.02	1.41 (9.9)	1.53 (9.9)	0.67
AIR6_LIM	−0.05	0.03	0.81 (−7.2)	0.83 (−6.7)	0.73
AIR7_ETH	0.30	0.02	1.06 (2.4)	1.07 (2.7)	0.74
AIR8_LEG	0.84	0.02	1.12 (4.4)	1.17 (6.3)	0.74

*n=2,946. Values in parentheses are standardized fit statistics (ZSTD). r_{pm} = point-measure correlation.

의 기대보다 상대적으로 예측 가능한 양상을 보였다.

끝으로 문항-측정치(r_{pm}) 상관 분석 결과, 모든 문항에서 0.66에서 0.76의 정적 상관이 나타나 문항의 점수 방향이 전체 잠재특성의 방향과 일치하였다. 모형의 기댓값과 비교하면 5번 문항의 관찰값은 0.67로 기댓값 0.75보다 낮았으며, 나머지 문항의 관찰값은 기댓값과 같거나 이를 상회하였다. 이상의 결과를 종합하면, 5번 문항은 MNSQ 기준을 벗어나고 문항-측정치 상관도 기댓값보다 낮은 부적합 양상을 보였다. 반면 2번 문항은 MNSQ 기준을 충족하였으나 다른 문항보다 상대적으로 높은 응답 예측 가능성을 보였다.

앞선 분석 결과에서 적합도 기준을 벗어난 5번 문항이 척도 전체의 측정 속성에 미치는 영향을 확인하고자 민감도 분석을 실시하였으며 그 결과는 표 4와 같다. 응답자 신뢰도는 8문항 모형과 7문항 모형 모두 .87이었고, 응답자 분리도는 각각 2.59와 2.61이었다. 문항 신뢰도는 두 모형 모두 1.00, 문항 분리도는 각각 21.15와 21.10이었다. 두 모형 모두에서 유한한 측정치가 산출된 2,879명을 대상으로 산출한 응답자 측정치 간 상관은 .982였다. 한편 Rasch 측정치가 설명한 분산은 57.6%에서 60.6%로 증가하였으나, 잔차 제1대조의 고유값 역시 2.79에서 2.87로 증가하였다. 응답자 평균 측정치는 1.09 logit에서 1.30 logit으로 상승하였으나, 두 모형은 문항 구성이 다르고 각각 재중심화되었으므로 이 차이는 응답자 수준의 실제 변화로 해석하지 않았다.

즉, 5번 문항을 제외하더라도 신뢰도와 분리도, 응답자 측정치의 상대적 순위는 거의 변하지 않았으며, 잔차 제1대조의 고유값도 감소하지 않았다. 이는 5번 문항의 부적합이 척도 전체의 측정 정밀도와 응답자 서열에 미치는 영향은 제한적이며 단순 삭제만으로는 잔차에서 확인된 차원성 문제가 개선되지 않음을 의미한다. 따라서 5번 문항을 삭제하기보다 측정 내용과 표현을 수정한 후 그 측정 속성을 재검증할 필요성을 제기하는 결과이다.

표 4. 5번 문항 제외에 따른 민감도 분석 결과
Table 4. Sensitivity analysis of Rasch models with and without item 5

Indicator	8-item model	7-item model (excluding item 5)
Person reliability	.87	.87
Person separation	2.59	2.61
Item reliability	1.00	1.00
Item separation	21.15	21.10
Mean person measure (logit)	1.09	1.30
Variance explained by measures (%)	57.6	60.6
First residual contrast eigenvalue	2.79	2.87
Correlation between person measures	–	.982

*The correlation between person measures was computed for the 2,879 respondents who obtained finite measures in both models.

3-3 실질 측면의 타당도 근거 검토 결과

실질 측면의 타당도 근거를 검토하기 위하여 먼저 평정척도 기능을 분석하였다. 해당 척도는 '보통이다'에 해당하는 중립 범주를 포함하지 않는 6점 척도로 AI 준비도를 측정하고 있으며, 따라서 애초에 의도한 대로 6개 범주가 적절히 기능하는지 확인하였다. 분석 결과는 표 5 및 그림 2와 같다.

먼저 범주별 관찰빈도를 살펴보면, 1점 345회(1%), 2점 1,851회(8%), 3점 3,341회(14%), 4점 7,620회(32%), 5점 7,833회(33%), 6점 2,578회(11%)로 총 23,568회의 응답이 관찰되었다. 즉 모든 범주가 최소 10회 이상의 관찰빈도를 확보하였으며, 4점과 5점에 응답이 집중되는 경향을 보였다. 다음으로 범주별 평균 측정치는 응답 범주가 1점에서 6점으로 증가할수록 -1.57 logit에서 3.53 logit으로 단조증가하였으며, 모형에 의한 기댓값과의 차이는 최대 0.19 logit으로 근접하였다. 범주별 적합도 지수는 내적합도 0.91에서 1.24, 외적합도 0.90에서 1.35의 범위로 모두 2.0 미만이었다. 경계모수치(Andrich threshold) 또한 -3.07, -1.27, -0.71, 1.12, 3.92 logit으로 역전 없이 순증하였으며, 그림 2의 범주확률곡선에서도 6개 범주 모두가 각각 최고점을 형성하였다.

다만 인접 경계모수 간 간격을 확인한 결과, 2점과 3점 사이는 1.80 logit, 4점과 5점 사이는 1.83 logit, 5점과 6점 사이는 2.80 logit으로 기준 범위인 1.4에서 5.0 logit 이내였으나, 3점과 4점 사이는 0.56 logit으로 기준에 미치지 못하였다. 이와 관련하여 범주별 일치도를 살펴보면, 측정치가 해당

범주를 예측하는 비율은 3점 범주에서 39%로 가장 낮았고 4점 범주에서 53%였다. 따라서 6개 범주는 Linacre가 제시한 여섯 가지 조건 중 다섯 가지를 충족하였으나, 3점과 4점 범주 간 경계의 구분은 상대적으로 좁게 형성된 것으로 나타났다.

다음으로 응답자의 능력치와 문항의 난이도 간 정합성을 검토하였다. 응답자의 평균 측정치가 문항의 평균 측정치보다 높거나 문항이 좁은 구간에 밀집되어 있는 경우 해당 구간을 벗어난 응답자를 정밀하게 변별하기 어렵다. 이와 같이 문항이 적절히 분포되어 있지 않은 경우 실질 측면의 타당도 근거가 충분하지 않다고 판단할 수 있다. 분석 결과, 응답자의 AI 준비도 측정치는 전반적으로 넓게 분포한 반면 문항은 좁은 구간에 밀집되어 있었다. 이를 정량적으로 확인하면, 응답자의 평균 측정치는 1.09 logit으로 문항의 평균 측정치인 0.00 logit보다 1.09 logit 높았으며, 응답자 측정치의 표준편차는 1.77 logit인 데 비해 문항 측정치의 표준편차는 0.55 logit이었다. 문항 난이도의 분포 범위는 -0.75에서 0.84 logit으로 1.59 logit에 불과하였다. 특히 가장 어려운 문항인 8번 문항의 위치인 0.84 logit보다 높은 측정치를 보인 응답자는 1,383명으로 전체의 46.9%를 차지하였다. 한편 8개 문항 모두에서 최고 응답범주를 선택한 극단응답자는 55명(1.9%)이었고 최저 응답범주만을 선택한 응답자는 없었다.

다만 평정척도모형에서는 문항 위치와 범주 경계가 결합되어 측정 구간이 확장되므로, 범주별 측정치를 기준으로 한 척도의 측정 범위는 -4.28 logit에서 5.06 logit이었다. 따라서 위의 결과는 0.84 logit을 초과하는 응답자를 측정할 수 없다는 의미가 아니라, 해당 구간에 대응하는 문항이 존재하지 않아 측정의 정밀도와 변별력이 상대적으로 낮아진다는 의미로 해석된다.

표 5. 평정척도 범주 구조 분석 결과
Table 5. Category structure of the rating scale

Category score	Observed count(%)	Measure	Infit MNSQ	Outfit MNSQ	Andrich threshold
1	345(1)	−1.57	1.24	1.35	NONE
2	1,851(8)	−0.99	1.11	1.21	−3.07
3	3,341(14)	−0.37	0.92	0.94	−1.27
4	7,620(32)	0.54	0.91	0.90	−0.71
5	7,833(33)	1.85	0.92	0.94	1.12
6	2,578(11)	3.53	1.08	1.04	3.92

*Total observations = 23,568. Values in parentheses are percentages; percentages may not sum to 100 because of rounding.

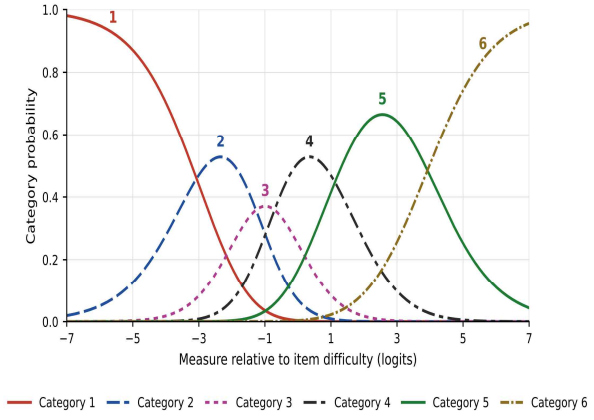


그림 2. 범주확률곡선
Fig. 2. Category probability curves

3-4 구조 측면의 타당도 근거 검토 결과

구조 측면의 타당도 근거를 파악하기 위하여 먼저 표준화 잔차의 주성분분석을 실시하였다. 분석 결과, 전체 원변량 18.88 가운데 Rasch 측정치가 설명한 분산은 10.88로 57.6%를 차지하였으며, 이는 모형에 의한 기댓값인 57.5%와 거의 일치하였다. 이 중 응답자에 의해 설명된 분산은 35.3%, 문항에 의해 설명된 분산은 22.3%였다. 그러나 설명되지 않은 잔차 42.4% 가운데 제1대조의 고유값은 2.79로 판단 기준인 2.0을 상회하였으며, 이는 전체 분산의 14.8%이자 미설명 분산의 34.9%에 해당하였다. 제2대조의 고유값은 1.28로 기준 이내였다. 따라서 자료가 Rasch 모형의 기대에 부합하는 수준으로 설명되었음에도 잔차에는 체계적인 부차 구조가 존재하는 것으로 나타났다.

다음으로 제1대조에 대한 문항별 적재량을 확인하였다. 분석 결과 6번(.56), 7번(.80), 8번(.78) 문항이 정적 방향의 군집을 형성하였고, 1번(-.64), 2번(-.57), 3번(-.49), 4번(-.41), 5번(-.31) 문항이 부적 방향의 군집을 형성하여 두 문항군이 명확히 대비되었다. 이는 1번부터 5번까지의 AI 활용 문항과 6번부터 8번까지의 AI 윤리·규범 이해 문항이라는 내용적 구분과 일치하였다. 다만 5번 문항의 적재량은 -.31로 AI 활용 문항군과 같은 방향에 위치하였으나 적재량의 절댓값은 8개 문항 중 가장 작았다. 이는 해당 문항이 AI 활용 문항군에 포함되면서도 다른 AI 활용 문항에 비해 상대적으로 이질적인 반응 양상을 보였음을 의미한다. 한편 두 군집으로 각각 산출한 응답자 측정치 간 상관은 .524였으며, 측정오차를 보정한 비감쇠 상관은 .635였다.

이어서 문항 간 국소독립성을 확인하기 위하여 표준화 잔차상관을 산출하였다. 그 결과 판단 기준인 절댓값 .30 이상의 정적 상관을 보인 문항쌍은 7번과 8번(.46), 1번과 2번(.30)이었다. 7번과 8번은 각각 AI 윤리적 의무와 법적 책임에 대한 이해를, 1번과 2번은 각각 AI 도구의 선택과 AI 정보의 통합 활용을 측정한다는 점에서 내용적 중복 가능성을 지닌다. 이 밖에 6번과 7번의 잔차상관은 .28로 기준에 근접하였다. 이상에서 단일차원성과 국소독립성이 충분히 확보되지 않은 것으로 나타남에 따라 아래와 같이 후속 분석을 실시하였다.

먼저 두 문항군을 각각 별도의 Rasch 평정척도모형으로 추정하였다. AI 활용 5문항의 경우 응답자 신뢰도는 .85, 분리도는 2.36이었으며, AI 윤리·규범 이해 3문항의 경우 응답자 신뢰도는 .88, 분리도는 2.71로 나타났다. 문항 수가 적음에도 후자의 신뢰도가 더 높게 산출된 것은 7번과 8번 문항 간 국소의존이 내적일치도를 상승시킨 결과로 볼 수 있다. 두 하위척도의 응답자 측정치 간 상관은 .587(95% CI [.563, .610], p<.001)이었고, 두 척도 모두에서 비극단 측정치가 산출된 2,745명을 대상으로 한 민감도 분석에서는 .518(95% CI [.490, .545], p<.001)이었다. 두 하위척도의 응답자 신뢰도를 적용하여 측정오차에 따른 감쇠를 보정한 상관은 .679였다.

다음으로 문항군의 구조를 확인하기 위하여 범주형 확인적 요인분석을 실시한 결과는 표 6과 같다. 8개 문항을 하나의 요인으로 설정한 단일차원모형 적합도는 χ^2(20)=4,056.030, CFI=.919, TLI=.887, RMSEA=.262(90% CI [.255, .269]), SRMR=.085로 나타났다. 반면 AI 활용과 AI 윤리·규범 이해로 구분한 2요인 모형의 적합도는 χ^2(19)=810.871, CFI=.984, TLI=.977, RMSEA=.119(90% CI [.112, .126]), SRMR=.031로 개선되었으며, DIFFTEST 결과에서도 모형 간 차이는 유의하였다($\Delta\chi^2$(1)=978.358, p<.001). 2요인 모형의 표준화 요인부하량은 .659~.920이었으며, 두 요인 간 상관은 .672였다. 이상의 결과는 8개 문항에 대한 응답 구조가 AI 활용과 AI 윤리·규범 이해라는 두 관련 차원으로 구분됨을 지지한다. 다만 2요인 모형은 CFI, TLI 및 SRMR에서 양호한 적합도를 보였으나, RMSEA는 .119로 통상의 수용 기준을 충족하지 못했다. 따라서 2요인 모형이 단일차원 모형보다 상대적으로 우수하다는 결과와 2요인 모형의 전반적인 적합 수준은 구분하여 해석할 필요가 있다.

이어서 국소의존이 확인된 문항쌍의 영향을 확인하기 위하여 잔차공분산을 자유추정한 모형을 추가로 검토하였다. 그 결과 7번과 8번의 잔차공분산을 자유추정한 모형은 χ^2(18)=633.952, CFI=.988, TLI=.981, RMSEA=.108, SRMR=.026이었으며, 해당 잔차공분산은 .472로 유의하였다($\Delta\chi^2$(1)=120.559, p<.001). 1번과 2번의 잔차공분산까지 함께 자유추정한 모형은 χ^2(17)=426.565, CFI=.992, TLI=.986, RMSEA=.090(90% CI [.083, .098]), SRMR=.021로 나타났으며, 두 잔차공분산은 각각 .365와 .475였다(모두 p<.001). 한편 잔차공분산을 순차적으로 자유추정할수록 두 요인 간 상관은 .672, .698, .706으로 다소 상승하였는데, 이는 두 문항쌍의 국소의존이 두 영역을 실제보다 이질적으로 나타나게 했을 가능성을 시사한다. 다만 이들 모형은 수정지수에 근거한 사후 탐색적 검토이므로 본 연구에서는 구조에 관한 결론의 근거가 아니라 문항 내용의 중복 여부를 확인하는 보조 자료로 사용하였다.

마지막으로 확인된 이차원 구조가 8개 문항 총점의 사용을 지지하는지 판단하기 위하여 이중요인 모형을 추정하였으며, 결과는 표 7과 같다. 모형 적합도는 χ^2(12)=304.781, CFI=.994, TLI=.986, RMSEA=.091(90% CI [.082, .100]), SRMR=.015였다. 일반요인에 대한 표준화 요인부하량은 .603에서 .838의 범위로 모든 문항에서 유의하였다(p<.001). 반면 특수요인의 부하량은 AI 활용의 경우 .170에서 .709, AI 윤리·규범 이해의 경우 .462에서 .723의 범위로 문항 간 편차가 컸으며, 일반요인보다 높은 부하량을 보인 문항은 1번과 7번이었다. 지수별로 살펴보면 일반요인의 ECV는 .675, 두 특수요인의 ECV는 각각 .140과 .185였으며, PUC는 .536이었다. 총점의 ωH는 .788, omega total은 .947이었고, 하위척도의 ωHS는 AI 활용 .177, AI 윤리·규범 이해 .415였다. 이는 PUC가 .80 미만인 조건에서 ECV .60과 ωH .70이라는 판단 기준을 모두 충족하는 값이다.

표 6. 범주형 확인적 요인분석 모형 적합도

Table 6. Fit indices for the categorical confirmatory factor analysis models

Model	CFI	TLI	RMSEA	SRMR
One-dimensional	.919	.887	.262	.085
Correlated two-factor	.984	.977	.119	.031
Two-factor + AIR7↔AIR8	.988	.981	.108	.026
Two-factor + AIR1↔AIR2, AIR7↔AIR8	.992	.986	.090	.021
Bifactor	.994	.986	.091	.015

*WLSMV estimation. χ^2(df) = 4,056.030(20), 810.871(19), 633.952(18), 426.565(17), 304.781(12) in the order listed; all p<.001. Factor correlations were .672, .698, and .706 for the three two-factor models. The models with residual covariances were examined post hoc. Residual covariance estimates are standardized (STDYX) values.

표 7. 이중요인 모형의 요인부하량 및 지수

Table 7. Standardized loadings and bifactor statistical indices

	General	AI use	AI ethical-normative understanding
AIR1_SEL	.603	.709	
AIR2_INT	.796	.406	
AIR3_PRC	.767	.296	
AIR4_PRF	.838	.199	
AIR5_CRT	.635	.170	
AIR6_LIM	.689		.462
AIR7_ETH	.638		.723
AIR8_LEG	.630		.595
ECV	.675	.140	.185
ω	.947		
ωH/ωHS	.788	.177	.415

*PUC = .536. All loadings were significant at p<.001. ECV = explained common variance; PUC = percentage of uncontaminated correlations; ωH = omega hierarchical; ωHS = omega hierarchical subscale.

3-5 일반화 측면의 타당도 근거 검토 결과

일반화 측면의 타당도 근거를 파악하기 위하여 성별에 따른 차별기능문항 분석을 실시하였다. 분석은 유한한 측정치가 산출된 남성 1,551명과 여성 1,340명을 대상으로 하였으며 최고점 극단응답자 55명(남성 40명, 여성 15명)은 추정에서 제외되었다. 먼저 실질적 크기를 살펴보면, 8개 문항의 성별에 따른 측정값 차이인 DIF contrast의 절댓값은 0.06에서 0.29 logit의 범위로 나타났다. 가장 큰 차이를 보인 문항은 3번 문항(“나는 AI가 어떤 절차를 거쳐 문제에 대한 답을 제공하는지 설명할 수 있다.”)으로, 남성의 문항 곤란도는 -0.07 logit, 여성은 0.22 logit이었으며 DIF contrast는 -0.29 logit이었다. 이는 해당 문항이 여성보다 남성에게 상대적으로 쉽게 기능하였음을 의미한다. 반면 7번(“나는 AI 시대에 가져야 할 윤리적 의무를 알고 있다.”)과 8번(“나는 AI 기술을 사용하는 과정에서 지켜야 할 법적 책임을 알고 있다.”) 문항의 DIF contrast는 각각 0.22와 0.18 logit으로 나타나 남성보다 여성에게 상대적으로 쉽게 기능하는 경향을 보였다. 8개 문항 모두 실질적 차별기능의 판단 기준인 0.43 logit에 미치지 못하여 무시할 수 있는 수준으로 분류되었다. 다음으로 통계적 유의성을 확인하였다. 모든 차별기능문항 대비값의 공통 표준오차는 .05였으며, Rasch-Welch t 검정 결과 3번(t=-5.73), 7번(t=4.53), 8번(t=3.83) 문항이 Bonferroni 보정 유의수준인 .00625에서 통계적으로 유의하였다. 1번 문항은 p=.027로 보정 전 유의수준에서는 유의하였으나 보정 후에는 유의하지 않았다. 나머지 문항은 보정 전후 모두 유의하지 않았다.

이상의 결과를 종합하면, 3개 문항에서 성별에 따른 문항 기능의 차이가 통계적으로 확인되었으나 그 크기는 모두 실질적으로 무시할 수 있는 수준이었다. 이는 표본 규모가 2,891명으로 커 미세한 차이까지 통계적으로 검출되었을 가능성이 있다. 한편 통계적으로 유의한 3개 문항 중 3번은 AI 활용 문항이고 7번과 8번은 AI 윤리·규범 이해 문항으로, 두 문항군의 차별기능 방향이 서로 반대로 나타났다.

집단별 적합도 분석에서 5번 문항의 내적합도는 남성 1.39, 여성 1.42로 유사하게 나타나, 3-2절에서 확인된 해당 문항의 부적합은 특정 성별에 국한된 것이 아님을 확인하였다. 다만 본 분석은 성별에 한정된 것이며, 연령, 학력, 전공계열, 취업상태 및 AI 사용빈도 등 다른 집단 구분에 따른 문항 기능의 동등성은 검토되지 않았다.

Ⅳ. 논의 및 결론

본 연구는 8차년도 KEEP Ⅱ 조사에서 사용된 AI 준비도 8개 문항에서 산출되는 점수의 타당도 근거를 검토하였다. 이를 위해 Rasch 평정척도모형을 적용하였으며, Wolfe와 Smith[28]가 제안한 내용 측면, 실질 측면, 구조 측면 및 일반화 측면에서 타당도 근거를 검토하였다. 아울러 확인된 구조가 8개 문항 총점의 사용을 지지하는지를 함께 검토하였다. 주요 결과에 대한 논의는 다음과 같다.

첫째, 본 연구의 핵심 결과는 8개 문항이 AI 활용과 AI 윤리·규범 이해라는 두 관련 영역을 포함함에도 일반요인이 상당한 비중을 설명하여 8개 문항 합산점수를 전반적인 자기보고 AI 준비도의 지표로 제한적으로 활용할 수 있다는 점이다. 실제로 표준화 잔차의 주성분분석에서 제1대조의 고유값은 2.79로 단일차원성의 판단 기준을 상회하였으며, AI 활용 문항군과 AI 윤리·규범 이해 문항군이 뚜렷하게 대비되었다. 이러한 이차원 구조는 동일한 자료에 적용한 서로 다른 세 가지 분석방법에서 일관되게 확인되었다. 잔차 군집 간 비감쇠 상관은 .635, 범주형 확인적 요인분석의 잠재요인 상관은 .672, 하위척도별 Rasch 측정치의 감쇠보정 상관은 .679로 유사한 범위에 수렴하였다. 또한 단일차원 모형과 상관 2요인 모형의 비교에서도 2요인 모형이 유의하게 더 적합하였다($\chi^2\Delta$(1)=978.358, p<.001). 다만 이중요인 분석에서 일반요인의 설명공통분산은 .675, 총점의 ωH는 .788로 나타났으며, 이는 PUC의 비율이 .80 미만인 조건에서 다차원성이 존재하더라도 본질적 단일차원으로 해석할 수 있다는 기준[42]을 충족하는 값이다. 요컨대 8개 문항은 두 하위영역을 포함하지만 공통분산의 상당 부분이 단일한 일반요인으로 설명되었다. 이는 해당 척도를 총점 형태로 활용해 온 선행연구[10],[21]의 관행에 내적 구조 측면의 측정학적 근거를 제공한다. 현재 자료에서는 8개 문항의 총점을 전반적인 자기보고 AI 준비 수준을 나타내는 연구용 지표로 제한적으로 사용할 가능성이 있으나, 두 관련 차원과 국소의존이 확인된 점을 고려하여 총점의 단일차원적 해석에는 주의가 필요하다. 또한 개인 수준의 진단, 절단점에 근거한 수준 판정 및 하위영역별 진단점수로의 분할 사용은 권장하기 어려우며, 성별 이외의 집단을 비교하려면 해당 집단에 대한 차별기능문항 또는 측정동등성 검토가 추가로 요구된다.

추가적으로 AI 활용과 AI 윤리·규범 이해의 하위점수를 서로 독립적인 진단 점수로 사용하는 것은 지지되지 않았다. 일반요인의 영향을 통제한 이후 특수요인에 남는 ωHS는 AI 활용 .177, AI 윤리·규범 이해 .415로 낮았다. 이는 하위척도 점수의 분산 대부분이 두 영역에 공통된 일반요인에서 비롯되며, 각 영역에 고유한 신뢰할 만한 정보는 제한적임을 의미한다. 이러한 결과는 특수요인의 부하량 양상에서도 확인된다. 특수요인의 부하량이 일반요인의 부하량을 상회한 문항은 1번과 7번이었으며, 두 문항은 각각 국소의존이 확인된 문항쌍에 포함되어 있었다. 이러한 부하량 양상과 낮은 ωHS를 함께 고려하면, 현재의 하위척도 점수가 각 영역에 고유한 안정적 정보를 충분히 제공한다고 보기는 어렵다. 따라서 후속 연구에서 하위영역별 진단이 필요한 경우에는 현재의 8개 문항을 단순히 분할하기보다 각 영역의 문항 수를 확충한 뒤 그 구조와 하위점수의 신뢰도를 재검증하는 접근이 요구된다.

둘째, 문항 수준에서는 두 유형의 개선 과제가 확인되었다. 먼저 5번 문항(“나는 오락이나 취미활동을 위해 AI 앱(프로

그램)을 활용할 수 있다")은 내적합도와 외적합도가 기준치를 초과하였다. 선행연구에 따르면 문항 내용의 표현이 모호하거나[24] 해당 문항에 대하여 응답자들이 일관되지 않는 반응 패턴을 보이는 경우[26] 적합도 지수 차원에서 부적절한 결과가 나타난다. 5번 문항은 문제 해결, 성과 향상 등 과업 지향적 활용을 묻는 다른 문항과 달리 여가 장면에서의 활용을 묻고 있어, 측정하는 활용 맥락이 달라 구인의 범위를 이탈했을 가능성이 있다. 특히 해당 문항은 8개 문항 중 유일하게 이미지 제작, 영상 편집, 음악 작곡이라는 구체적 예시를 제시하고 있어 응답자가 생성형 창작 도구의 사용 경험이라는 협소한 준거로 반응했을 가능성이 있다. 실제로 해당 문항은 잔차 제1대조에서 AI 활용 문항군과 같은 방향에 위치하였으나 적재량의 절댓값이 .31로 가장 작아 상대적으로 이질적인 반응 양상을 보였으며, 성별 집단별 적합도가 유사하게 나타나 부적합이 특정 집단에 국한된 현상도 아니었다. 다만 민감도 분석 결과 5번 문항을 제외하더라도 응답자 측정치 간 상관은 .982로 순위 변화가 미미하였고 신뢰도 역시 동일하였던 반면 잔차 제1대조의 고유값은 오히려 증가하였다. 따라서 5번 문항은 단순히 삭제하기보다 여가 장면으로 한정된 표현을 제거하고 산출물의 생성·편집 능력에 초점을 두어 수정한 뒤 재검증할 필요가 있다. 예컨대 "나는 AI 앱(프로그램)을 활용하여 이미지·영상·문서 등 필요한 콘텐츠를 직접 생성하거나 편집할 수 있다"와 같은 형태이다. 다만 이는 예시이며 수정된 문항의 실제 측정 속성은 본 연구에서 검증되지 않았다.

다음으로 문항 간 국소의존이 확인되었다. 7번과 8번 문항의 표준화 잔차상관은 .46, 1번과 2번 문항은 .30으로 판단 기준을 상회하였다. 7번과 8번은 각각 AI 시대의 윤리적 의무와 AI 사용 과정의 법적 책임을, 1번과 2번은 각각 문제 해결에 적합한 AI 도구의 선택과 AI 정보의 통합 활용을 측정한다는 점에서 내용적 중복 가능성을 지닌다. 2번 문항이 판정 기준 내에서 보인 과적합 경향(내적합도 0.71, 외적합도 0.72) 또한 이러한 맥락에서 해석할 수 있다. 즉, 이는 문항의 질이 낮음을 의미하는 것이 아니라, 응답 일치도가 관찰값 63.0%로 기댓값 55.5%를 크게 상회한 데서 나타나듯 인접 문항과 중복된 정보를 측정하여 반응이 상대적으로 예측 가능해진 결과로 볼 수 있다. 이와 관련하여 잔차공분산을 순차적으로 자유추정할수록 RMSEA는 .119에서 .108과 .090으로 감소하였고, 요인 간 상관은 .672에서 .698과 .706으로 다소 상승하였다. 이는 문항 간 내용 중복이 2요인 모형의 부적합을 일부 설명하는 동시에 두 영역을 실제보다 이질적으로 보이게 했을 가능성을 시사한다. 다만 이러한 결과는 요인구조에 관한 확정적 근거가 아니라 문항 간 차별화와 개정의 필요성을 뒷받침하는 보조 근거로 해석하였다.

이러한 결과를 고려할 때, 후속 척도 개정에서는 각 문항쌍이 서로 구별되는 측면을 측정하도록 재진술할 필요가 있다. 구체적으로 1번은 과업의 목적과 조건에 맞는 AI 도구를 비교·선택하는 능력에, 2번은 AI가 제공한 정보의 정확성과 적절성을 검토하고 이를 기존 지식과 통합하는 능력에 초점을 둘 수 있다. 7번은 공정성·투명성·인간 존중 등 AI 활용에서 고려해야 할 윤리적 판단 원칙의 이해에, 8번은 저작권·개인정보보호 등 준수해야 할 법적 규정의 이해에 초점을 둘 수 있다. 다만 이러한 수정 방향은 예시이며, 인지면접과 전문가 내용타당도 검토 및 예비조사를 통해 그 적절성을 검증할 필요가 있다.

셋째, 문항의 곤란도 분포는 응답자의 준비도 수준에 비해 낮은 곳에 형성되어 있었다. 응답자의 평균 측정치는 1.09 logit으로 문항의 평균 측정치보다 1.09 logit 높았으며, 문항 곤란도의 분포 범위는 1.59 logit에 그쳤다. 특히 가장 어려운 문항인 8번 문항보다 높은 측정치를 보인 응답자가 46.9%에 달하였다. 이는 현재 문항이 준비도 수준이 평균 이하인 응답자에게는 적절히 기능하나, 중상위 응답자를 정교하게 변별하기에는 대응하는 문항이 부족함을 의미한다. 한편 평정척도기능 분석에서 6개 응답범주는 관찰빈도, 평균 측정치의 단조증가, 적합도, 경계모수의 순서 등 여섯 조건 중 다섯을 충족하였으나, 3점과 4점 사이의 경계모수 간격이 0.56 logit으로 기준에 미치지 못하였고 3점 범주의 일치도가 39%로 가장 낮았다. 이는 3점과 4점이 각각 '약간 그렇지 않다'와 '약간 그렇다'로 방향만 다른 대칭적 표현이어서, 중립 범주가 없는 조건에서 응답자가 두 범주를 명확히 구분하기 어려웠을 가능성을 시사한다. 따라서 중상위 구간을 측정할 수 있는 곤란도 높은 문항의 추가와 함께, 중간 두 범주의 표현을 보다 변별적으로 재진술하는 방안을 검토할 필요가 있다.

넷째, 성별에 따른 문항 기능의 차이는 실질적으로 무시할 수 있는 수준이었다. 8개 문항의 DIF contrast 절댓값은 0.29 logit 이하로 모두 실질적 판단 기준인 0.43 logit에 미치지 못하였다. 다만 다중비교를 보정한 이후에도 3번, 7번, 8번 문항에서 통계적 유의성이 확인되었는데, 이는 표본 규모가 큰 조건에서 미세한 차이도 검출된 결과로 해석된다. 즉 이들 세 문항 중 3번은 AI 활용 문항, 7번과 8번은 AI 윤리·규범 이해 문항으로 차별기능의 방향이 서로 반대였으나, 효과크기가 모두 미미하여 이를 두 문항군의 구조적 구분에 관한 근거로 확대해석하지는 않았다. 아울러 이러한 결과는 성별에 한정된 측정 불변성의 근거일 뿐이며 척도 전반의 일반화 타당도가 확보되었음을 의미하지는 않는다.

이상의 결과를 타당도 근거의 관점에서 정리하면 다음과 같다. 내용 측면에서는 대부분의 문항이 적합도 기준을 충족하고 문항–측정치 상관도 모두 정적 방향으로 나타났으나, 5번 문항의 부적합이 제약으로 남는다. 실질 측면에서는 6개 응답범주가 대체로 의도한 순서로 기능하였으나, 중상위 구간을 측정할 문항의 부족과 중간 범주의 좁은 경계 간격이 확인되었다. 구조 측면에서는 두 관련 하위영역과 일반요인의 존재가 확인되어 총점의 제한적 활용은 지지되었으나, 하위점수의 독립적 사용은 지지되지 않았으며 두 문항쌍의 국소의존이 남아 있다. 일반화 측면에서는 성별에 따른 문항 기능의 차이가 실질적으로 무시할 수준이었으나, 성별 이외의 집단에 관한 근거는 확보되지 않았다. 따라서 본 연구는 네 측면의

근거를 검토했다는 사실만으로 척도 전반의 타당도가 확보되었다고 결론짓기보다, 각 측면에서 확보된 근거와 남아 있는 제약을 구분하여 제시하는 데 의의를 둔다.

이상의 연구결과와 시사점에도 불구하고 본 연구에는 몇 가지 한계점이 존재한다. 첫째, 패널조사 특성상 Wolfe와 Smith가 제시한 타당성 검증 차원 중 외적 기준에 근거한 타당성과 결과에 기초한 타당성을 검증하는 데 한계가 있었다. 본 연구가 확인한 것은 척도 내부의 측정 속성에 관한 근거이며, 해당 점수가 외부 준거와 어떤 관계를 갖는지, 그리고 점수의 활용이 어떠한 결과를 초래하는지에 대해서는 검증되지 않았다. 아울러 본 연구가 내용 측면의 근거로 활용한 문항 적합도와 문항–측정치 상관은 문항이 전체 측정구조 안에서 기대되는 방식으로 기능하는지에 관한 경험적 근거이며, 전문가 평정을 통해 문항 내용의 관련성, 대표성, 포괄성을 확인하는 절차를 대체하지는 않는다. 둘째, 성별을 제외한 연령, 학력, 전공계열, 취업상태, AI 사용빈도 등 다양한 인구통계학적 변인에 따른 차별기능문항 분석이 이루어지지 못하였다. 측정의 불변성에 대한 견고한 근거를 확보하기 위해서는 추후 연구에서 다양한 배경과 맥락에 따른 차별기능문항의 확인이 지속적으로 필요하다. 셋째, 본 연구의 분석대상은 AI 앱 또는 프로그램 사용 경험이 있으며 AI 준비도 8개 문항에 모두 응답한 청년으로 제한되었다. 특히 선택편향 검토 결과 분석 표본에는 학사학위 취득자가 상대적으로 많이 포함되어 있었으므로, 연구결과를 AI 사용 경험이 없는 청년을 포함한 전체 청년집단에 일반화하는 데에는 한계가 있다. 넷째, 문항모수 추정의 불변성 전제를 고려하여 조사 가중치를 적용하지 않았으므로, 본 연구의 결과는 문항과 척도의 측정 속성에 관한 것이며 모집단의 AI 준비도 수준에 대한 추정치로 직접 일반화될 수 없다. 다섯째, 본 연구에서 제안한 문항 개선 방향은 현재 자료에 근거한 예시 수준이며, 수정된 문항이 실제로 개선된 측정 속성을 보이는지는 확인되지 않았다. 따라서 제안된 개선 방향은 인지면접, 전문가 내용타당도 검토, 예비조사 및 재교정 절차를 거쳐 검증되어야 한다.

이러한 한계에도 불구하고 본 연구는 국가 수준 패널조사에 수록되어 이미 다수의 실증연구에 활용되고 있는 AI 준비도 문항에 대하여 문항 수준의 측정 근거를 제시하였다는 점에서 의의를 갖는다. 종합하면, KEEPⅡ AI 준비도 8개 문항은 AI 활용과 AI 윤리·규범 이해라는 두 관련 하위영역을 포함하지만 일반요인이 공통분산의 상당 부분을 설명하여 합산점수는 전반적인 자기보고 준비 수준을 나타내는 연구용 지표로 제한적으로 활용할 수 있다. 반면 두 하위영역의 고유신뢰도는 낮아 하위점수를 독립적인 진단점수로 사용하는 것은 권장되지 않는다. 척도의 측정 정밀도를 높이기 위해서는 5번 문항의 활용 맥락을 재검토하고 1번과 2번 및 7번과 8번 문항의 내용 중복을 줄이는 동시에 준비도 상위 구간을 변별할 수 있는 고난도 문항을 추가할 필요가 있다. 이러한 결론은 AI 사용 경험이 있는 KEEPⅡ 청년 표본에 한정되며, 수정 문항과 다른 집단에서는 재검증이 요구된다.

참고문헌

[1] OECD, *OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market*, Paris, France: OECD Publishing, 2023. https://doi.org/10.1787/08785bba-en

[2] C. Lutz, “Digital Inequalities in the Age of Artificial Intelligence and Big Data,” *Human Behavior and Emerging Technologies,* Vol. 1, No. 2, pp. 141-148, April 2019. https://doi.org/10.1002/hbe2.140

[3] X. Chen, Y. Chen, Y. Xie, and L. Cheng, “Translation and Psychometric Validation of the Medical Artificial Intelligence Readiness Scale (MAIRS-MS) for Chinese Medical Students,” *BMC Nursing,* Vol. 24, 1210, September 2025. https://doi.org/10.1186/s12912-025-0385 2-w

[4] X. Wang, L. Li, S. C. Tan, L. Yang, and J. Lei, “Preparing for AI-Enhanced Education: Conceptualizing and Empirically Examining Teachers’ AI Readiness,” *Computers in Human Behavior,* Vol. 146, 107798, September 2023. https://doi.org/10.1016/j.chb.2023.107798

[5] D. Long and B. Magerko, “What Is AI Literacy? Competencies and Design Considerations,” in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu: HI, pp. 1-16, April 2020. https://doi.org/10.1145/3313831.3376727

[6] T. K. F. Chiu, Z. Ahmad, M. Ismailov, and I. T. Sanusi, “What Are Artificial Intelligence Literacy and Competency? A Comprehensive Framework to Support Them,” *Computers and Education Open,* Vol. 6, 100171, June 2024. https://doi.org/10.1016/j.caeo.2024.100171

[7] D. T. K. Ng, J. K. L. Leung, S. K. W. Chu, and M. S. Qiao, “Conceptualizing AI Literacy: An Exploratory Review,” *Computers and Education: Artificial Intelligence*, Vol. 2, 100041, 2021. https://doi.org/10.1016/j.caeai.2021.100041

[8] O. Karaca, S. A. Çalışkan, and K. Demir, “Medical Artificial Intelligence Readiness Scale for Medical Students (MAIRS-MS): Development, Validity and Reliability Study,” *BMC Medical Education,* Vol. 21, 112, February 2021. https://doi.org/10.1186/s12909-021-02546-6

[9] J. Jöhnk, M. Weißert, and K. Wyrtki, “Ready or Not, AI Comes—An Interview Study of Organizational AI Readiness Factors,” *Business & Information Systems Engineering,* Vol. 63, No. 1, pp. 5-20, February 2021. https://doi.org/10.1007/s12599-020-00676-7

[10] J. Y. Jeong and N. O. Kim, “The Mediating Effect of AI Readiness on the Relationship Between Digital Utilization and Job Satisfaction Among Employed College Graduates,” in *Proceedings of the 2025 KRIVET Panel Conference,* Seoul, pp. 199-218, September 2025.

[11] N. Khajeali, N. Kohan, S. Rezaei, and A. Saberi, “Psychometric Assessment of the Persian Translated Version of the ‘Medical Artificial Intelligence Readiness Scale for Medical Students’,” *PLoS ONE,* Vol. 20, No. 5, e0323543, May 2025. https://doi.org/10.1371/journal.pone.0323543

[12] S. M. Kim, A Study on the Clinical Artificial Intelligence Readiness of Nurses, Master’s Thesis, Ewha Womans University, Seoul, August 2024.

[13] S. Chang and Y. Kim, “Exploratory Study of Readiness for Nursing College Students to Utilize Artificial Intelligence,” *The Journal of Yeolin Education,* Vol. 31, No. 3, pp. 71-91, May 2023. https://doi.org/10.18230/tjye.2023.31.3.71

[14] Y. Lee, “Application and Effectiveness Analysis of AI-Based Teaching Methods for Improving AI Preparedness in Healthcare Students,” *The Journal of Learner-Centered Curriculum and Instruction*, Vol. 25, No. 11, pp. 655-664, June 2025. https://doi.org/10.22251/jlcci.2025.25.11.655

[15] B. Y. Lee, An Analysis of Factors Influencing Teachers’ Intention to Use Generative AI: Focusing on the Technology Readiness and Acceptance Model, Ph.D. Dissertation, Chung-Ang University, Seoul, 2024.

[16] S. S. Baeck and J. S. Lee, “The Structural Impact of Technology Readiness on Call Center Counselors' Intention to Use in the Introduction of Artificial Intelligence Systems: Focusing on AICC (Artificial Intelligence Contact Center),” *Journal of Information Technology Services*, Vol. 22, No. 4, pp. 1-19, August 2023. https://doi.org/10.9716/KITS.2023.22.4.001

[17] W. C. Jin and S. I. Baek, “Meta-Research for Identifying Key Factors in AI Readiness: An Integrated Approach Using Systematic Literature Review and Text Analysis,” *Journal of Service Research and Studies,* Vol. 15, No. 4, pp. 167-187, December 2025. https://doi.org/10.18807/jsrs.2025.15.4.167

[18] Y.-S. Moon and B.-J. Park, “Impact of National AI Readiness on Digital Services Trade Performance,” *Journal of Digital Contents Society,* Vol. 26, No. 11, pp. 3127-3138, November 2025. https://doi.org/10.9728/dcs.2025.26.11.3127

[19] T. G. Bond and C. M. Fox, *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*, 3rd ed., New York, NY: Routledge, 2015. https://doi.org/10.4324/9781315814698

[20] W. J. Boone, J. R. Staver, and M. S. Yale, *Rasch Analysis in the Human Sciences*, Dordrecht, Netherlands: Springer, 2014. https://doi.org/10.1007/978-94-007-6857-4

[21] Y. Lee and T. Im, “Latent Profile Analysis of AI Readiness, Digital Utilization at Work, and Soft Skills among University Graduates: Associations with Job Satisfaction and Well-Being,” *Journal of the Korea Society of Computer and Information,* Vol. 31, No. 4, pp. 257-266, April 2026. https://doi.org/10.9708/jksci.2026.31.04.257

[22] T. J. Seong, *Understanding and Application of Item Response Theory,* 2nd ed. Seoul, Kyoyookbook, 2016.

[23] Y. Jo, A. J. Hong, and C. Jeong, “A Psychometric Investigation of the Utrecht Work Engagement Scale-9 Using the Rasch Measurement Model,” *Journal of Competency Development & Learning*, Vol. 15, No. 1, pp. 111-139, March 2020. https://doi.org/10.21329/khrd.2020.15.1.111

[24] A. J. Hong, Y. Jo, and S. Lee, “An Investigation of the Validity of Employee Engagement Scale Applying the Item Response Theory: Based on Rasch Measurement Model,” *Korean Journal of Business Administration*, Vol. 33, No. 4, pp. 577-598, April 2020. https://doi.org/10.18032/kaaba.2020.33.4.577

[25] K. W. Lee, “Validation of the CESD-11 Using Item Response Theory,” *Mental Health & Social Work,* Vol. 49, No. 2, pp. 5-26, June 2021. https://doi.org/10.24301/MHSW.2021.3.49.2.5

[26] D. Yeum, “Validation of the Disability Acceptance Scale Using Item Response Theory,” *Disability & Employment,* Vol. 35, No. 1, pp. 5-26, February 2025. https://doi.org/10.15707/disem.2025.35.1.001

[27] J. Kim, M. Son, and D. Yeum, “Validation of an Adolescent Self-Esteem Scale Using the Rasch Model,” *Children’s Light,* Vol. 120, No. 1, pp. 69-92, 2025.

[28] E. W. Wolfe and E. V. Smith Jr., “Instrument Development Tools and Activities for Measure Validation Using Rasch Models: Part II—Validation Activities,” *Journal of Applied Measurement,* Vol. 8, No. 2, pp. 204-234, 2007.

[29] Korea Research Institute for Vocational Education and Training, Korean Education and Employment Panel Survey II (2024), Korea Research Institute for Vocational Education and Training, Sejong, Report No. 2024-06, December 2024.

[30] S. M. Yates, Rasch and Attitude Scales: Explanatory Style, in *Applied Rasch Measurement: A Book of Exemplars,* Dordrecht, Netherlands: Springer, pp. 207-225, 2005. https://doi.org/10.1007/1-4020-3076-2_12

[31] D. Andrich, “A Rating Formulation for Ordered Response Categories,” *Psychometrika*, Vol. 43, No. 4, pp. 561-573, December 1978. https://doi.org/10.1007/BF02293814

[32] S. Messick, “Validity of Psychological Assessment:

Validation of Inferences from Persons' Responses and Performances as Scientific Inquiry into Score Meaning," *American Psychologist,* Vol. 50, No. 9, pp. 741-749, September 1995. https://doi.org/10.1037/0003-066X.50.9.741

[33] American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, *Standards for Educational and Psychological Testing*, Washington, DC: American Educational Research Association, 2014.

[34] H.-S. Seol, "A Study for Validation of Scores for Instruction Evaluation Questionnaires Using Rasch Measurement Model," *Journal of Educational Evaluation,* Vol. 20, No. 4, pp. 31-51, December 2007.

[35] B. D. Wright and J. M. Linacre, "Reasonable Mean-Square Fit Values," *Rasch Measurement Transactions*, Vol. 8, No. 3, p. 370, 1994.

[36] J. M. Linacre, "Optimizing Rating Scale Category Effectiveness," *Journal of Applied Measurement,* Vol. 3, No. 1, pp. 85-106, 2002.

[37] A. Gayef and G. Ilgaz, "Medical Students' Readiness for Medical Artificial Intelligence (AI)," *INQUIRY: The Journal of Health Care Organization, Provision, and Financing,* Vol. 63, 00469580261418133, February 2026. https://doi.org/10.1177/00469580261418133

[38] A. Ziapour, F. Darabi, P. Janjani, M. A. Amani, M. Yıldırım, and S. Motevaseli, "Factors Affecting Medical Artificial Intelligence (AI) Readiness Among Medical Students: Taking Stock and Looking Forward," *BMC Medical Education,* Vol. 25, No. 1, 264, February 2025. https://doi.org/10.1186/s12909-025-06852-1

[39] M. J. Zieky, Practical Questions in the Use of DIF Statistics in Test Development, in *Differential Item Functioning*, Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 337-347, 1993.

[40] L. K. Muthén and B. O. Muthén, *Mplus User's Guide*, 8th ed., Los Angeles, CA: Muthén & Muthén, 2017.

[41] A. Rodriguez, S. P. Reise, and M. G. Haviland, "Evaluating Bifactor Models: Calculating and Interpreting Statistical Indices," *Psychological Methods,* Vol. 21, No. 2, pp. 137-150, June 2016. https://doi.org/10.1037/met0000045

[42] S. P. Reise, R. Scheines, K. F. Widaman, and M. G. Haviland, "Multidimensionality and Structural Coefficient Bias in Structural Equation Modeling: A Bifactor Perspective," *Educational and Psychological Measurement,* Vol. 73, No. 1, pp. 5-26, February 2013. https://doi.org/10.1177/0013164412449831



조윤성(Yunseong Jo)

2012년 : 중앙대학교 대학원 (교육학 석사)
2018년 : 중앙대학교 대학원 (인적자원개발학 박사)

2012년～2014년: 경기도평생교육진흥원 사원
2018년～2025년: 중앙대학교 연구처 연구전담교수
2025년～현 재: 현대해상화재보험 인재개발파트 책임연구원
※관심분야 : 측정 및 평가, 조직개발 및 변화, 리더십, 연구방법론 등