THE JOURNAL OF KOREAN ASSOCIATION OF COMPUTER EDUCATION Volume 29, Issue 3, 2026

컴퓨터교육학회 논문지 2026년 제29권 제3호
https://doi.org/10.32431/kace.2026.29.3.001

재직자 대상 AI 리터러시 평가를 위한 한국형 SNAIL (K-SNAIL)의 신뢰도 및 타당도 검증*

Validation of the Korean Version of the SNAIL (K-SNAIL) for Assessing AI Literacy in Working Professionals

박진아† · 류호경†† · 김지은†††
Jinah Park† · Hokyoung Ryu†† · Jieun Kim†††

요약

AI 리터러시는 산업 전반의 필수 역량으로 부상했으나, 기존 척도는 학생이나 교사 중심으로 개발되어 재직자의 실제 직무 맥락을 반영하지 못했다. 본 연구는 비전공자 성인을 위한 SNAIL 척도를 한국 재직자 환경에 맞게 번안하고, 그 신뢰도와 타당성을 검증하였다. 번역-역번역과 대규모 언어모델 검토, 전문가 포커스그룹을 통해 문항을 정제하고, 재직자 401명의 응답과 AI 교육 전후 데이터를 실증분석한 결과 모형 적합도(TLI=.93, CFI=.93, RMSEA=.07)와 신뢰도(α≥.97)가 모두 우수했다. 이는 본 척도가 산업 현장의 AI 리터러시를 정밀하게 진단하고 교육 효과를 평가하는 유용한 도구임을 보여준다.

주제어 AI 리터러시, 척도 타당화, 재직자 교육, 요인분석, SNAIL

ABSTRACT

AI literacy has become an essential competency across industries, yet most existing scales were designed for students or teachers and fail to capture the workplace context of professionals. This study adapted the SNAIL (Scale for the Assessment of Non-Experts' AI Literacy) for Korean working professionals and empirically validated its reliability and validity. Through translation–back-translation, large-language-model review, and expert focus groups, items were refined and tested with responses from 401 employees. An additional pre- and post-training empirical analysis confirmed good model fit (TLI=.93, CFI=.93, RMSEA=.07) and high reliability (α≥.97), demonstrating that the K-SNAIL is an effective tool for diagnosing and evaluating AI literacy in workplace learning contexts.

Keywords AI literacy, scale validation, working professionals, factor analysis, Item Response Theory, SNAIL

†정회원 (주)에이블런 대표이사
††정회원 한양대학교 교육혁신처 처장/교수
†††정회원 한양대학교 기술경영학과 교수(교신저자)

논문투고 2025년 10월 16일
심사완료 2025년 11월 06일
게재확정 2025년 11월 19일
발행일자 2026년 03월 15일

*본 연구는 보건복지부의 재원으로 「환자중심 의료기술 최적화 연구사업(Patient-Centered Clinical Research Coordinating Center, PACEN)」 지원을 받았음(과제고유번호: RS-2025-02215037).

1. 서론

인공지능(AI)이 산업과 일상 전반으로 빠르게 확산되면서 AI를 이해하고 활용하는 능력, 즉 AI 리터러시(AI Literacy)를 어떻게 평가할 것인가에 대한 관심이 커지고 있다. 최근 연구들은 단순한 기술 지식 습득을 넘어 비판적 사고, 윤리적 판단, 실제 적용 역량까지 포함하는 다차원적인 개념으로 AI 리터러시를 정의하며[1-3], 이를 측정하기 위한 다양한 척도 개발이 이루어지고 있다. 특히 고등교육 및 성인 학습자 집단을 대상으로 한 연구에서는 기존의 학생이나 교사 중심 평가 도구가 직무 기반 학습자의 실제 역량을 충분히 반영하지 못한다는 점이 반복적으로 지적되어 왔다[4-6]. 이에 따라 비전문가나 산업 재직자를 위한 평가 도구의 필요성이 커졌으며, 대표적으로 Laupichler 등(2023)의 SNAIL(Scale for the Assessment of Non-Experts' AI Literacy) [7, 8]과 Soto-Sanfiel 등(2024)의 SAIL4ALL(Scale of Artificial Intelligence Literacy for All) [9] 등이 개발되었다. 그러나 이러한 국제 척도들은 대부분 영어권 표본을 바탕으로 설계되어, 언어·문화적 적합성과 직무 맥락 측면에서 한계가 있다.

국내의 경우 AI 리터러시를 지식이나 기술, 태도 등의 차원에서 체계적으로 구조화하지 않으면 교육과 평가가 단편화될 수 있다는 점이 지적되고 있다[10]. 기존에는 교사, 예비교사, 대학생 등 집단을 중심으로 AI 리터러시 관련 척도 개발이 이루어졌으나[11-14], 산업 현장 재직자나 비전공 성인을 대상으로 한 표준화된 연구는 여전히 미흡하다. 특히 성인 학습자와 재직자 집단의 AI 활용은 학습 과제 수준이 아닌 직무 수행과 의사결정, 위험 관리와 직접적으로 연결되므로 AI 리터러시의 구성 요소 역시 이러한 맥락을 반영해 평가될 필요가 있다[15]. 최근에는 일반 성인을 대상으로 AI 리터러시를 정량적으로 측정하려는 연구도 확대되고 있으며, 인식, 활용, 평가, 윤리와 같은 하위 요소를 중심으로 타당화된 척도 개발이 보고되고 있다[16]. 이에 따라 국제적으로 개발된 AI 리터러시 척도를 특정 국가와 직무 환경에 맞게 번안하고 재검증하는 연구는 단순한 도구 이전을 넘어 AI 리터러시 개념을 현지 맥락에서 재구성하는 과정으로서 중요한 의미를 가진다. 따라서 검증된 국제 척도를 한국 재직자 맥락에 맞게 번안과 타당화 연구가 필요하다.

본 연구는 비전문가 성인을 대상으로 개발된 영문판 SNAIL 척도를 기반으로, 한국 재직자용 AI 리터러시 평가 도구의 한국형 타당화를 목적으로 한다. 이를 위해 번역-역번역과 전문가 포커스그룹(Focus Group Interview), 대규모 언어모델(Large Language Model, LLM) 보조 점검을 결합한 문항 정제 과정을 수행하고, 국내 재직자 401명을 대상으로 탐색적 요인분석(EFA), 확인적 요인분석(CFA), 문항반응이론(IRT), Rasch 분석을 통해 척도의 신뢰도와 타당도를 검증하였다. 또한 한국형 SNAIL 척도(K-SNAIL)의 민감도를 검증하기 위해 AI 교육 전후 재직자 401명의 리터러시 변화를 실증분석 하였다. 이를 통해 본 연구는 국내 재직자 집단의 직무 기반 AI 활용 역량을 신뢰성 있게 측정할 수 있는 표준 도구를 제시하고, 향후 교육 성과 평가와 디지털 역량 진단 체계의 기초 자료를 제공하고자 한다.

2. 문헌 연구

2.1 AI 리터러시 평가 도구 개발 및 연구

AI 리터러시의 중요성이 확대되면서 국내외에서는 다양한 집단을 대상으로 한 평가 도구가 개발되어 왔다. 국제적으로 가장 널리 활용되는 척도는 Laupichler 등(2023)이 개발한 SNAIL로, 이는 공식적인 AI 또는 컴퓨터공학 교육을 받지 않은 일반 성인을 뜻하는 비전문가를 대상으로 한 체계적 척도다[7, 8]. 이 연구는 AI 교육 전문가 53인의 델파이 합의를 거쳐 39개 문항을 도출하고, 영어권 성인 415명을 대상으로 탐색적 요인분석을 수행하여 기술적 이해(Technical Understanding), 비판적 평가(Critical Appraisal), 실용적 활용(Practical Application)의 3요인 구조를 제시하였다. 최종 31개 문항으로 구성된 SNAIL은 내적 일관도(Cronbach's α = .85~.93)가 높게 보고되며, 성인 비전공자 연구의 표준 척도로 자리 잡았다.

이후 Soto-Sanfiel 등(2024)은 SNAIL을 기반으로 SAIL4ALL을 개발하여, 다양한 사회와 문화 환경에서도 적용할 수 있는 범용형 척도의 가능성을 제시하였다[9]. 독일의 Carolus 등(2023)은 MAILS(Meta AI Literacy Scale)을 제안하며, AI 이해와 응용, 평가와 창작, 윤리 인식, 그리고 자기효능감을 포함하는 다섯 가지 요인 구조를 검증하였다[17]. Biagini 등(2024)은 대학생 집단을 대상으로 AI에 대한 지식, 실제 운영 능력, 비판적 판단, 윤리 의식으로 구성된 네 가지 요인 구조를 확인하였다[18]. 이러한 연구들은 모두 고등교육 이상의 성인 학습자를 대상으로 하며, AI 리터러시를 지식 이해와 활용 능력, 비판적 사고, 윤리적 책임 의식이 함께 작동하는 통합적 역량으로 측정하려는 공통된 방향을 보여준다.

Lintner(2024)의 체계적 검토 연구에서는 SNAIL이 가장 자주 인용된 성인용 척도로 확인되었으며, 문항의 간결성, 구성 타당도, 문화 간 일관성 측면에서 높은 평가를 받았다[19]. 그러나 대부분의 연구가 영어권 표본에 집중되어 있어, 비영어권 산업 종사자 집단에서의 검증은 부족하다는 한계가 제시되었다. 또한 각 척도별로 요인 수나 항목 경계가 상이하여, AI 리터러시 구성요소가 문화적, 맥락적 요인에 따라 변동될 수 있음이 확인되었다.

국내에서도 교사와 예비교사를 대상으로 한 AI 역량 측정 도구 개발 연구[12], 초등학생 대상 학습 흥미 척도 개발[13], 유아교사 평정척도 개발[14] 등이 이루어지면서 다양한 교육 단계별 도구가 축적되었다. 그러나 이러한 도구는 특정 연령이나 직군 중심으로 개발되어 성인 재직자 집단의 직무 특성과 조직 맥락을 충분히 반영하지 못한다는 한계가 지적된다[11-13]. 이러한 한계는 국외에서 검증된 도구를 국내 재직자 맥락에 맞게 번안하고 타당화하는 연구의 필요성을 제기한다. 따

라서 국내 재직자 맥락을 반영한 표준화된 평가 도구가 요구되며 이를 위해서는 단순한 언어 번역을 넘어 문화적 적합성과 현장 타당성을 확보해야만 기업과 산업 현장에서 실질적으로 활용 가능한 AI 리터러시 평가 체계를 마련할 수 있다.

2.2 평가척도의 타당성 검증 방안

AI 리터러시와 같은 역량 평가 척도의 신뢰도와 타당성을 검증하기 위해서는 통계적 분석과 전문가 검토를 함께 수행하는 절차가 필요하다. 일반적으로 이러한 연구에서는 탐색적 요인분석(Exploratory Factor Analysis, EFA)과 확인적 요인분석(Confirmatory Factor Analysis, CFA)을 활용해 척도의 구성 타당성을 점검한다. EFA 단계에서는 polychoric 상관계수와 사각회전(promax 또는 oblimin) 방식을 적용하여 요인 수와 문항의 분포를 탐색하며, CFA 단계에서는 비교적합지수(Comparative Fit Index, CFI), 터커-루이스지수(Tucker-Lewis Index, TLI), 평균제곱근오차근사값(Root Mean Square Error of Approximation, RMSEA), 표준화잔차평균제곱근(Standardized Root Mean Square Residual, SRMR) 등을 통해 모형의 적합도를 검증한다. 이 과정에서 개념신뢰도(Composite Reliability, CR) 와 평균분산추출값(Average Variance Extracted, AVE)을 산출해 수렴 타당도를 평가하고, Fornell-Larcker 기준이나 이질-수렴비율(Heterotrait-Monotrait Ratio, HTMT)을 통해 판별 타당도를 확인하는 것이 일반적인 절차이다[20-22]

전통적인 요인분석 외에도 문항반응이론(Item Response Theory, IRT) 과 라쉬 모형(Rasch Model) 이 척도의 정밀성과 해석력을 높이기 위한 보완적 방법으로 사용된다 [23]. IRT는 문항별 난이도와 변별도를 추정해 응답자의 능력 수준에 따른 반응 패턴을 세밀하게 분석할 수 있으며, 라쉬 모형은 문항 간의 간격을 동일하게 조정하여 척도의 일관성과 비교 가능성을 높이는 데 유용하다[23]. 또한 차별기능문항(Differential Item Functioning, DIF) 분석을 통해 성별, 연령, 직무 등 집단 간에서 문항이 동일하게 작동하는지를 점검함으로써 평가의 공정성을 확보할 수 있다[23]. Lintner(2024)는 최근의 AI 리터러시 척도 검증 동향을 분석한 체계적 검토에서, 요인분석(EFA, CFA)을 넘어 IRT와 DIF 기반의 검증이 점차 확산되고 있음을 보고하며, 이러한 접근이 향후 척도 연구의 새로운 기준으로 자리 잡고 있음을 제시하였다[17].

아울러 통계적 검증뿐 아니라 정성적 점검을 함께 수행하는 시도도 늘어나고 있다. 그 중 하나로, LLM을 보조 도구로 활용해 번역의 적합성이나 문항 표현의 명확성을 사전에 점검하고, 이후 전문가가 이를 검토 및 수정하는 인간 중심 검토 절차(human-in-the-loop) 방식이 주목받고 있다[24]. 이러한 접근은 언어적 모호성을 줄이고 문화적 적합성을 확보하는 동시에, 문항 개발과 수정 과정을 효율적으로 관리할 수 있다는 장점이 있다.

본 연구 역시 이러한 절차를 실제 분석 과정에 적용하였다. 번역과 역번역을 거친 한국형 SNAIL (K-SNAIL) 문항을 LLM 보조 도구를 활용해 1차 문항을 검토한 뒤, AI 관련 강의 및 연구 경험이 있는 전문가 집단의 검토를 통해 인간-AI 협업 방식으로 문항의 표현, 문화적 적합성, 현장 맥락 등을 정제했다. 이러한 접근은 기존 번안 연구가 주로 전문가 합의에만 의존했던 한계를 보완하는 시도로서, 정성적 타당성 확보의 새로운 방식을 제시한다. 이후 탐색적, 확인적 요인분석(EFA, CFA), 문항반응이론(IRT), 라쉬 모형 분석, 그리고 차별기능문항(DIF) 검증을 단계적으로 수행함으로써 척도의 구조적 타당도, 문항 수준의 정밀성, 그리고 집단 간 공정성을 함께 확인하였다. 이러한 절차를 통해 척도의 신뢰도와 타당성을 확보하고, 재직자 집단의 교육적 맥락에서도 안정적으로 활용할 수 있음을 검증하였다.

3. 연구 방법

3.1 표본 수집

본 연구는 K-SNAIL 척도의 타당성 검증을 위해 AI 교육 전문 기업 ㈜에이블런이 보유한 재직자 패널을 대상으로 진행되었다. 모집은 뉴스레터, 홈페이지 게시 등을 통해 웹 기반 설문 링크를 배포하여 참여자에게는 소정의 음료 쿠폰이 제공되었다. 설문 참여자 중 중복 응답 및 불성실 응답을 제외하고 총 401명(여 206명, 남 195명)의 자료를 사용했다. 모든 참여는 자발적 동의에 근거하며, 응답 자료는 비식별화 및 익명 처리 후 분석했다(Table 1).

Table 1. Sample Profile for Scale Validation (N = 401)

Gender	Frequency (%)	Age	Frequency (%)
Male	195 (48.6)	18 – 24	32 (8.0)
		25 – 29	161 (40.1)
Female	206 (51.4)	30 – 39	128 (31.9)
		40 – 49	43 (10.7)
		50 – 59	25 (6.2)
Company Type	Frequency (%)	Job Role	Frequency (%)
Large	37 (9.2)	IT planning	105 (26.2)
Mid-sized	102 (25.4)	Edu / HR	68 (17.0)
Small	191 (47.6)	Management	53 (13.2)
Startup	44 (11.0)	R&D	34 (8.5)
Public	14 (3.5)	Engineer	29 (7.2)
		Marketing	29 (7.2)
Industry	Frequency (%)	Major	Frequency (%)
IT	163 (40.6)	AI major	126 (31.4)
Manufacturing	64 (16.0)	Non-AI	275 (68.6)
Construction	30 (7.5)	Related field	207 (51.6)
Bio/Chemical	28 (7.0)	No related field	194 (48.4)

Note. Categories with fewer than 10 respondents (< 2.5%) were excluded.

3.2 검증 절차

검증 절차는 4단계로 나눠진다(Figure 1). 첫째, 문항 정교화 및 번역 검증 단계에서는 원 척도의 문항을 번역 및 역번역한 후, LLM 보조도구와 전문가 포커스그룹(FGI)을 연계하여 문항의 언어적, 문화적 적합성과 표현의 명확성을 점검하였다. 둘째, 요인 구조 및 구성 타당도 검증 단계에서는 EFA와 CFA를 통해 척도의 요인 구조를 검증하고, 수렴타당도와 판별타당도를 교차 검증하였다. 셋째, 문항 수준 검증 단계에서는 Rasch/IRT로 문항의 난이도와 변별도를 살펴보고, 이어 DIF 분석으로 성별 등의 집단 간 공정성을 확인했다. 넷째, 문항의 실용 타당성 검증 단계에서는 대응표본 t-검정을 통해 K-SNAIL 척도가 교육 전후 학습자의 AI 리터러시 변화에 민감하게 반응하는지를 검증하였다.

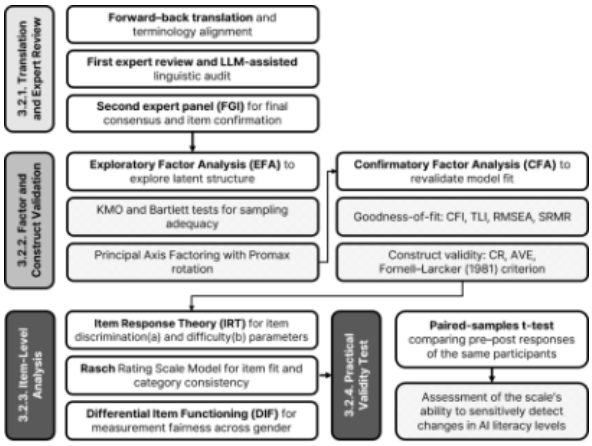

Figure 1. Validation Procedure of the Scale

3.2.1 문항 정교화 및 번역 검증 단계

문항 번역은 2025년 7~8월 약 4주간 진행되었다. 번역팀은 영어 및 교육·기술 분야 전문성의 3인(연구진 2인, 외부 검토자 1인)으로 구성되어, 원문 의미가 자연스럽게 전달되도록 1차 번역-역번역-합의 조정의 3단계 절차를 거쳤다.

이후 Open AI의 Chat GPT 5 모델을 LLM 보조도구로 활용해 문항의 언어적, 문화적 적합성과 모호 표현을 점검하고, AI 교육 경력자 및 언어, 교육 전문가 14인으로 구성된 포커스그룹(FGI)을 통해 표현의 구체성, 전문어 과다, 학습자 경험과의 괴리를 검토하였다.

LLM은 모호 표현, 이중문항, 중복 범위, 문체 불일치를 탐지하도록 설정하였고 (Appendix 1), FGI에서는 제안된 수정안을 수용, 수정 또는 보류로 구분해 최종 반영하였다. 문항 표현은 '~을 설명할 수 있다' 형식으로 통일하고, 중복 개념은 분리하여 요인 간 의미가 겹치지 않도록 조정하였다 (Appendix 2).

3.2.2 요인 구조 및 구성 타당도 검증 단계

AI 리터러시 역량을 측정하는 31개 문항의 잠재 요인 구조를 탐색하고, 문항이 이론적으로 일관된 하위 요인으로 구분되는지를 확인하기 위해 EFA를 실시하였다. 요인 추출은 주축요인분석(principal axis factoring, PAF)을 사용하고, 요인 간 상관을 허용하는 프로맥스(promax) 사각회전을 적용하였다. 분석 전, Kaiser-Meyer-Olkin(KMO) 지수와 Bartlett의 구형성 검정으로 표본의 요인분석 적합성을 확인하였다. 요인 수 결정은 고유값 1 이상 기준과 스크리 플롯, 이론적 해석 가능성을 종합해 판단하였다.

EFA로 도출된 구조는 Amos 21.0을 활용한 CFA로 재검증하였다. CFA는 이론적으로 설정된 요인 구조의 자료 적합성을 평가하는 절차로[21, 22], 각 문항이 사전 정의된 요인에 적절히 부하되는지를 확인하였다. 모형 적합도 평가는 χ^2 통계 외에 TLI, CFI, RMSEA, SRMR을 함께 보고하였으며, 해석 준거는 TLI/CFI ≥ .90, RMSEA/SRMR ≤ .08 기준을 따랐다[21, 22].

구성개념 타당도는 수렴타당도와 판별타당도로 검증하였다. 수렴타당도는 각 요인의 평균분산추출(Average Variance Extracted, AVE)과 개념신뢰도(Composite Reliability, CR)를 산출해 평가하였고, 판별타당도는 Fornell-Larcker(1981) 기준에 따라 각 요인의 $\sqrt{AVE}$가 상호 상관계수보다 큰지를 확인하였다[20]. 이를 통해 도출된 요인들이 통계적으로 구별되면서도 개념적으로 일관됨을 검증하였다.

3.2.3 문항 수준 검증 단계

문항의 정밀성과 집단 간 공정성을 검증하기 위해 IRT, 라쉬(Rasch) 모형, DIF 분석을 수행하였다. IRT 분석은 등급반응모형(graded response model; GRM)을 기반으로 2요인 구조에 적합하였으며, 각 문항의 변별도(a)와 난이도(b_1-b_4)를 추정하였다. 변별도는 응답자의 능력 수준에 따른 문항의 판별력을, 난이도는 5점 리커트 척도 각 등급의 경계값을 의미한다. 추정은 R 패키지 mirt를 활용한 한계우도법(marginal maximum likelihood, MML)으로 산출하였다.

라쉬 모형은 평정척도모형(Rating Scale Model, RSM)을 적용해 문항의 적합도와 난이도를 추정하였다. 모든 문항은 동일 기울기(a=1)를 가정하고 등급 전환점(b_1-b_4)을 추정하여 범주가 순차적으로 작동하는지, 척도의 일관성이 유지되는지를 검증하였다.

성별(남, 여)에 따른 집단 간 공정성은 Mantel-Haenszel 검정 기반 DIF 분석으로 확인하였다. 동일 능력 수준에서 문항 응답이 특정 집단에 편향되지 않는지를 점검하였으며, difR 패키지를 활용하고 유의수준은 FDR 5% 보정 기준으로 판정하였다.

3.2.4 문항의 실용 타당성 검증 단계

K-SNAIL 척도의 구조적 타당화 과정 이후, 학습자의 AI 리터러시 변화를 반영하는지 후속 검증으로 확인하기 위해,

사전-사후 응답 결과를 비교하였다. 이 분석은 교육 효과를 직접적으로 평가하기보다는, 척도가 교육 전후 변화를 민감하게 포착하는지 검증하기 위해 동일 응답자의 사전-사후 점수를 대응표본 t-검정(paired-samples t-test)으로 분석하였다. 이 방법은 두 시점 간 평균 차이를 검증하여 척도의 변화 반응성을 평가하는 절차로, 선행 연구[24]에서도 척도의 민감도 검증에 활용되어 왔다.

AI 리터러시 점수는 요인별 문항 평균을 기준으로 산출하였다. '기술적 이해'는 14문항(Q1-Q14), '응용과 성찰'은 17문항(Q15-Q31)으로 구성되며, 전체 점수는 31개 문항의 평균값으로 계산하였다. 모든 문항은 5점 리커트 척도로 측정되었고, 점수가 높을수록 AI 리터러시 수준이 높은 것을 의미한다.

분석은 전체 점수와 두 하위 요인 점수가 교육 프로그램 참여 전후에 유의한 변화를 보이는지를 확인하기 위한 것으로, 양측 검정에 유의수준 $p < .05$를 적용하였다. 정규성 가정은 사전 점검하였고, 필요 시 비모수 검정을 병행하였다. 결과는 평균(Mean), 표준편차(SD), t값, p값으로 제시하였다.

4. 연구 결과

4.1 문항 번역 및 전문가 검증

전문가 검토 결과 대부분의 문항은 적절하다고 평가되었으나 일부 문항에서는 수정이 필요하다는 의견이 제시되었다. 특히 '어떻게', '일반적인' 등 추상적인 표현은 보다 구체화할 필요가 있었으며(Q2, Q4, Q7, Q9, Q12), '기술적 이해' 영역의 일부 문항은 세부 내용이 부족하다는 지적이 있었다(Q7, Q8, Q10, Q12, Q13, Q14). 또한 '응용과 성찰' 영역에서는 의미가 유사하거나 중복되는 문항이 확인되었고(Q7, Q15, Q16, Q23), 일상 사례와 직무 사례를 구분해 제시할 필요가 있다는 의견도 제시되었다(Q25, Q26). 1차 검토에서 확인된 사항을 바탕으로, 문항의 표현을 정밀하게 다듬기 위해 LLM을 활용한 교열 및 감수 절차를 추가로 수행하였다.

LLM 분석 결과, 전문가 의견과 대체로 일치하는 네 가지 주요 개선 사항이 도출되었다. 첫째, '어떻게', '일반적인' 등 추상적 표현의 반복 사용으로 인한 문화적, 언어적 모호성이 확인되어 구체적 서술이 요구되었다. 둘째, 한 문장에서 두 가지 행동을 동시에 묻는 이중 문항이 발견되었으며(Q7, Q15, Q16, Q23), 셋째, 일상생활과 직무 사례가 혼합된 중복 문항이 확인되었다(Q25, Q26). 넷째, AGI, XAI, NLP 등 약어 표기와 문장 종결 표현이 일관되지 않아 용어 및 문체 통일이 필요했다.

전문가들은 LLM이 제안한 수정 사항을 검토하여 각 항목을 수용, 수정, 유지로 구분하였다. 표현의 구체화와 용어 통일 제안은 대부분 반영되었고, 예를 들어 '어떻게 작동하는지 설명할 수 있다'는 '작동 원리를 설명할 수 있다'로, '개발과 활용'을 묻는 문항은 '개발 과정 설명'과 '활용 사례 제시'로 분리되었다. 반면, Q24("AI가 무엇인지 설명할 수 있다")는 LLM이 '기술적 이해'로 재분류를 제안했으나, 전문가들은 학습자의 경험적 서술이 '응용과 성찰'에 더 가깝다고 판단해 기존 분류를 유지하였다.

영문 약어는 첫 등장 시에만 기재하고 이후에는 국문 표기로 통일하였으며, 문장 종결 표현은 '~을 설명할 수 있다'로 일원화하였다. 이 과정을 통해 문화적 부적합이나 의미 왜곡은 발견되지 않았고, 중복되거나 모호한 문항은 분리 또는 구체화되었다. 최종 문항은 원문 의미를 유지하면서 한국 재직자의 언어와 업무 맥락에 적합하게 확정되었으며, 수정안은 원저자 Laupichler 등(2023)에게 공유되어 번안과 변환의 적절성에 대한 동의를 받았다.

4.2 AI 리터러시 척도의 잠재 요인 구조 탐색

원 척도의 개발 연구에서는 EFA를 통해 기술적 이해, 비판적 평가, 실용적 활용의 3요인 구조를 제시한 바 있다. 그러나 K-SANIL 척도에서는 문항 번역과 문화적 맥락 조정 과정을 거치며, 일부 문항이 중복 부하되거나 요인 간 상관이 과도하게 높게 나타났다. 이에 본 연구는 한국 재직자 표본의 응답 데이터를 바탕으로 새로운 잠재 구조를 탐색하였다.

먼저 표본이 요인분석에 적합한지를 확인하기 위하여 Kaiser-Meyer-Olkin(KMO) 표본 적합도 지수와 Bartlett의 구형성 검정을 실시하였다. 그 결과, KMO 값은 .979로 나타나 기준치인 .60을 크게 상회하였으며, Bartlett의 구형성 검정은 χ^2=12401.397, df=435, $p < .001$로 유의하였다(Table 2).

고유값 기준 1 이상을 적용한 결과, 총 2개의 요인이 추출되었다. 첫 번째 요인은 AI 개념과 기술적 작동 원리에 대한 이해를 포함해 '기술적 이해' 요인으로 명명하였으며, 두 번째 요인은 학습된 개념을 실제 업무나 일상에 적용하고 그 결과를 성찰하는 역량으로 '응용과 성찰'로 명명하였다. 두 요인의 전체 누적 설명분산은 71.326%로 나타나, 문항들이 비교적 높은 설명력을 지니고 있음을 보여준다. 내적 일관성을 검토하기 위해 Cronbach's α 계수를 산출한 결과, 요인별 α 값은 각각 .970와 .971로 모두 매우 높은 수준을 보였다.

이러한 결과는 원 척도에서 분리되었던 비판적 평가와 실용적 활용이 한국 재직자 집단에서는 하나의 연속적 인식 과정으로 통합되어 나타난다는 점을 시사한다. 이에 따라 번역 및 문화 적합성 조정 이후 문항 간 중복 부하와 요인 간 과도한 상관을 해소하고, 해석 가능성을 높이기 위해 본 연구는 K-SNAIL 척도의 구조를 2요인 모델로 수정하여 이후 분석에 활용하였다.

Table 2. Main Results of Factor Analysis

Item	Factor Loading	
	Technical Understanding	Application and Reflection
Eigenvalue	18.601	2.797
Variance (%)	62.002	9.324
Cumulative Variance (%)	62.002	71.326
Cronbach's α	.970	.971
KMO Measure of Sampling Adequacy	.979	
Bartlett's Test of Sphericity	χ^2	12401.397
	df	435
	p	.000

4.3 요인 구조의 타당성 및 신뢰도 검증

4.3.1 확인적 요인분석을 통한 구조 적합성 확인

FEA에서 확인된 2요인 구조가 실제 표본 데이터에 적합한지를 검증하기 위해 CFA를 실시한 결과, 본 연구의 측정모형은 TLI=.927, CFI=.932, RMSEA=.072, SRMR=.047로 나타났다(Table 3). 일반적으로 TLI와 CFI 값이 .90 이상, RMSEA와 SRMR 값이 .08 이하일 경우 모형이 수용 가능한 수준으로 평가된다[21, 22]. 따라서 이 결과는 2요인 구조가 실제 표본(401명)의 응답 자료에 잘 부합하며, 한국 재직자 집단의 AI 리터러시를 설명하는 구조로서 적합함을 의미한다.

Table 3. Results of Confirmatory Factor Analysis (CFA)

χ^2	df	TLI	CFI	RMSEA	SRMR
1243.874	404	.927	.932	.072	.047

4.3.2 신뢰도 및 수렴 타당도 검증

개념신뢰도(CR)는 각 요인별 문항의 요인부하량을 기반으로 계산하였다(Table 4). CR 값이 .70 이상이면 요인을 구성하는 문항들이 동일 개념을 안정적으로 반영한다고 판단한다[23]. 분석 결과, 기술적 이해 요인의 CR은 .970, 응용과 성찰 요인의 CR은 .971로 모두 높은 수준의 신뢰도를 확보했다.

수렴타당도(AVE)는 각 요인이 설명하는 공통 분산의 비율을 산출하여 검증하였다. AVE가 .50 이상이면 요인이 오차보다 공통 요인을 더 많이 설명한다는 의미로 수렴타당도가 확보된 것으로 본다[23]. 분석 결과, '기술적 이해' 요인의 AVE는 .699, '응용과 성찰' 요인의 AVE는 .677로 모두 기준치를 충족했다. 이 결과는 두 요인에 속한 문항들이 각자의 개념을 안정적으로 측정하고 있으며, 요인 내부의 문항들이 동일 개념으로 잘 수렴하고 있음을 보여준다.

Table 4. Results of AVE and CR

Factor	AVE	CR
Technical Understanding (Q1 – Q14)	.699	.970
Application and Reflection (Q15 – Q31)	.677	.971

4.3.3 판별 타당도 검증

서로 다른 요인들이 실제로 구분되는 독립된 개념임을 확인하기 위한 절차로 판별타당도 분석 결과, 기술적 이해와 응용과 성찰 요인의 상관계수 제곱값은 .584로 두 요인의 AVE 값보다 작게 나타났다(Table 5). 이에 따라 본 연구의 측정도구는 판별타당도 또한 충족하는 것으로 확인되었다. 즉, 두 요인은 서로 독립된 개념임을 나타낸다.

Table 5. Discriminant Validity Test

Factor	1	2
1. Technical Understanding	**.699**	
2. Application and Reflection	.584	**.677**

Note. Bold values on the diagonal represent AVE. The value below the diagonal indicates the squared correlation between factors.

종합하면 본 연구에 사용된 한국 재직자의 실제 인식 구조를 반영한 2요인 K-SNAIL 모델은 수렴타당도와 판별타당도를 모두 확보하였으며, 따라서 통계적으로 타당한 측정도구임을 확인하였다.

4.4 문항의 수준 분석

4.4.1 문항 변별도 및 난이도 분석

등급반응모형(Graded Response Model)을 적용한 결과, 각 문항의 변별도(a)는 0.48~1.53 범위로 나타나 응답자의 능력 수준에 따라 점수를 안정적으로 구분할 수 있었다. 문항의 난이도(b_1–b_4)는 -2.3에서 +5.3 범위로, 대부분 문항이 평균 능력 수준($\theta \approx 0$) 근처에 분포하여 재직자 집단의 전반적 역량 수준에 적절하게 맞춰져 있었다. 요인 간 공분산은 0.243으로, 두 요인(기술적 이해, 응용과 성찰)이 완전히 독립적이지는 않지만 약한 상관을 보였다.

4.4.2 문항 일관성과 응답 범주의 적합성

라쉬(Rasch) 적합도 지수(INFIT, OUTFIT)는 허용 기준인 0.7~1.3 범위 내에 포함되었다.

다만 Q11과 Q14는 OUTFIT이 1.5 이상으로 나타나 응답자의 선택이 일정한 패턴에서 벗어나는 경향이 있었다. 이는 일부 응답자가 중간 점수(2~3점)를 명확히 구분하지 못했거나, 문항 서술이 다소 추상적으로 느껴졌기 때문으로 해석된다. 전반적으로 문항 일관성은 허용 기준을 충족했고, 응답 범주 라벨 정교화 같은 미세 조정만으로 충분히 개선 가능한 수준이었다.

4.4.3 집단 간 공정성

Mantel-Haenszel 방식의 DIF 분석에서 대부분의 문항은 성별에 따른 유의한 차이가 나타나지 않아 공정성이 확보되었다. Q1에서만 유의한 차이가 관찰되었고($p < .001$), 여성은 낮은 점수 선택 비율이, 남성은 중간 이상 점수 선택 비율이 상대적으로 높았다.

4.5 교육 전후 비교를 통한 실용 타당성 검증

K-SNAIL 척도의 변화 반응성을 확인하기 위해 동일 응답자 401명의 사전-사후 점수를 대응표본 t-검정으로 분석하였다(Table 6). 전체 AI 리터러시 평균은 교육 전 M=2.33(SD=.83)에서 교육 후 M=3.70(SD=.76)으로 유의한 차이를 보였다(t=-38.175, p<.001).

하위 요인별로는 기술적 이해 점수가 사전 M=2.16(SD=.89)에서 사후 M=3.52(SD=.94)로(t=-38.076, p<.001), 응용과 성찰 점수가 사전 M=2.48(SD=.89)에서 사후 M=3.86(SD=.71)로(t=-36.727, p<.001) 모두 유의하게 상승하였다(Table 6). 이 결과는 본 척도가 교육 전후 상황에서 학습자의 AI 리터러시 수준 변화를 정확하고 민감하게 탐지할 수 있음을 보여준다.

Table 6. Results of Paired-Samples t-Test

Variable	Group	Mean	SD	t	p
Overall AI Literacy	Pre-test	2.33	.83	−38.175***	.000
	Post-test	3.70	.76		
Technical Understanding	Pre-test	2.16	.89	−38.076***	.000
	Post-test	3.52	.94		
Application and Reflection	Pre-test	2.48	.89	−36.727***	.000
	Post-test	3.86	.71		

Note. N = 401. *p < .05, **p < .01, ***p < .001.

5. 결론 및 시사점

본 연구는 재직자 대상 AI 리터러시 척도의 한국형 타당화를 목적으로 수행되었다. 이를 위해 탐색적·확인적 요인분석, 문항반응이론, Rasch 및 DIF 분석을 통해 척도의 구조적·문항 수준의 타당성과 공정성을 모두 검증하였다.

그 결과, 기존 SNAIL 원 척도의 세 요인(기술적 이해, 비판적 평가, 실용적 활용)이 한국 재직자 집단에서는 기술적 이해와 응용과 성찰의 2요인 구조로 수렴되었으며, 모형 적합도(TLI=.93, CFI=.93, RMSEA=.07)와 신뢰도($\alpha \geq .97$)가 모두 우수하게 나타났다. 또한 IRT, Rasch, DIF 분석을 통해 개별 문항 수준에서도 변별도와 공정성이 확보되었고, 대응표본 t-검정 결과($p < .001$)는 K-SNAIL 척도가 교육 전후의 리터러시 변화에 민감하게 반응함을 확인하였다. 이러한 결과는 본 척도가 재직자 교육 맥락에서 신뢰도와 실용성을 모두 갖춘 평가 도구임을 보여준다.

5.1 시사점

본 연구의 시사점은 첫째, 한국 재직자 맥락에서의 개념 정교화와 구조 검증이다. 본 연구는 기존 SNAIL의 세 요인(기술적 이해, 비판적 평가, 실용적 활용)[7, 8]이 한국 재직자에게는 '기술적 이해'와 '응용과 성찰'의 두 요인으로 통합된다는 점을 확인하였다. 이는 재직자에게 직장 환경에서의 AI 리터러시는 기술 이해, 업무 적용, 성찰과 판단이 결합된 형태로 나타난다는 선행 연구의 논의에 따라[15], 응답자들이 기술적 이해와 실제 적용을 별개로 인식하기보다 하나의 연속적인 학습과 실천, 성찰의 과정으로 받아들이기 때문으로 해석된다.

이러한 결과는 AI 리터러시가 단순한 도구 활용 능력이 아니라, AI의 가능성과 한계를 이해하고 결과를 비판적으로 해석하며 책임 있게 활용하는 다차원적 역량이라는 선행 개념 정의와도 부합한다[2, 3]. 또한 맥락과 집단에 따라 요인 구조가 달라질 수 있음을 보고한 Carolus 등(2023)의 MAILS[17], Biagini 등(2024)의 4요인 모델[18]에서 나타난 맥락별 구조 차이와도 일치하며, AI 리터러시 구성요소가 고정된 구조가 아니라 현지 맥락에 따라 연구가 필요하다는 점[10, 19]에 대해 국내 재직자 집단에서 실증적으로 그 결과를 확장한 것으로 볼 수 있다.

둘째, 문항 정교화 과정에서 LLM을 검토 보조 도구로 활용하여 인간과 AI의 협업 절차를 통합적으로 적용한 점이다. 기존 연구들이 전문가 합의에만 의존했다면[12, 14], 본 연구는 LLM을 교열 보조 도구로 활용하고 전문가 그룹이 제안 내용을 수용, 수정, 유지로 판정하는 절차를 거쳤다. 이는 Lintner(2024)[19]가 강조한 문화와 언어의 반응성 확보가 필요하다는 관점과 관련하여 인간 전문가의 판단을 보조하는 AI 기반 검토의 실제 사례를 보여준다.

셋째, 척도의 구조와 문항의 정밀성, 평가의 공정성까지 검토한 다차원 검증 과정이다. 국내 선행연구들이 주로 탐색적 요인분석과 확인적 요인분석에 집중해 왔다면[12, 14], 본 연구는 이를 넘어 IRT로 각 문항의 난이도와 변별도를 분석하고[23], Rasch 모형을 통해 응답 범주의 일관성을 확인했다. 또한 DIF 분석을 활용해 성별 집단 간 공정성도 점검했다[23]. 이 과정은 Lintner(2024)가 제시한 최신 연구 흐름인 요인분석을 넘어 문항 수준의 정밀성과 공정성 검증으로 확장이라는 방향과 부합한다[19]. 그 결과 K-SNAIL 척도는 요인 구조가 안정적일 뿐만 아니라, 개별 문항 단위에서도 신뢰성과 공정성을 확보한 평가 도구로 검증되었다.

넷째, 교육 전후 비교를 통해 실용 타당성을 확인했다. 대응표본 t-검정 결과(총점 및 각 요인 모두 $p < .001$)는 본 척도가 학습자의 수준 변화를 유의하게 포착할 수 있다는 것을 뜻한다. 이는 K-SNAIL이 단순한 교육 효과 측정 도구를 넘어 AI 리터러시를 교육 성과의 핵심 지표로 활용해야 한다는 최근 논의[4]와 같이 재직자 학습의 변화를 진단할 수 있는 실질적 평가 도구로서 활용할 수 있음을 시사한다.

종합하면, 본 연구는 한국 재직자라는 실제 현장 맥락에서 AI 리터러시의 개념을 정교화하고, 척도의 구조와 문항

수준을 모두 검증하여 신뢰성과 실용성을 입증했다. 이를 통해 한국형 AI 리터러시 평가 도구의 표준화를 위한 기반을 마련하였으며, 기존 연구가 제시했던 교육 단계 중심의 한계를 넘어 국제 척도 개발 및 표준화 필요성에 관한 논의[9, 19]와 학습 참여 및 동기 등 성과 지표와의 연계 논의[4], 고등교육 맥락에서의 척도 연구[17, 18]을 확장했다.

5.2 한계 및 제언

연구의 한계와 제언은 다음과 같다. 첫째, 표본의 구성에 따른 일반화 가능성의 한계다. 본 연구의 표본은 20~30대 응답자가 중심이었고, 직무 역시 IT기획과 교육·인사 분야에 상대적으로 집중되어 있었다. 그럼에도 제조, 건설, 바이오 등 비ICT 산업군 응답자도 약 59% 포함되어 표본의 다양성을 확보하려 노력했다. 향후 연구에서는 산업과 직무, 연령, 경력별로 표본을 층화하여 수집하고, 다집단 확인적 요인분석(CFA) 등을 검증한다면 척도의 일반화 가능성과 산업별 적용 타당성을 보다 명확히 확인할 수 있을 것이다.

둘째, 응답 범주와 문항 표현의 세밀한 보완이 점진적으로 요구되는 점이다. 대부분 문항은 Rasch 모형의 INFIT/OUTFIT 기준(0.7~1.3)을 충족했지만, 일부 문항(Q11, Q14)에서는 중간 점수(2~3점)의 해석이 모호한 것으로 나타났다. 이는 응답자가 범주 간 의미를 명확히 구분하지 못한 결과로 보인다. 본 연구에서는 범주 라벨을 보완했지만, 추후 인지면접 단계 등을 추가하여 응답 과정 타당도를 점검하거나 4점 척도로 단순화하는 대안적 설계를 시험한다면 측정 안정성을 높이는 방법도 있다.

셋째, LLM을 문항 검토에 활용한 점은 유용했으나, 프롬프트 설계나 검토 절차의 표준화는 아직 미흡한 실정이다. 앞으로는 LLM을 활용한 번역, 검토, 전문가 합의, 파일럿 검증 절차를 체계화하고, AI 분석과 자동 피드백 기능을 결합한 한국형 척도 개발 모델을 구체화할 필요가 있다. 특히 LXP(Learning Experience Platform) 기반 교육에서 데이터 피드백과 자기성찰이 학습 지속성과 만족도를 높였다는 선행연구[25]는, AI 리터러시 척도 또한 AI 기반 분석과 실시간 피드백을 접목한 적응형 진단 시스템으로 발전할 가능성을 보여준다.

본 연구는 K-SNAIL이 향후 다양한 언어와 산업 직군으로 확장될 가능성을 보여주었다. 기관이나 기업 단위의 교육 성과 관리 체계에 적용함으로써 척도의 활용성을 높일 수 있으며, 국제적 표준화 논의의 기반 자료로도 활용될 수 있을 것이다.

참고문헌

[1] Choi, S. (2024). Core competency framework and education plan for future talent in the era of generative AI. *Journal of The Korean Association of Computer Education, 27*(9), 23-33. https://doi.org/10.32431/kace.2024.27.9.003

[2] Long, D., & Magerko, B. (2020, April). What is AI literacy? Competencies and design considerations. *In Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-16). https:// doi. org/ 10. 1145/ 3313831.3376727

[3] Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. Computers and Education: *Artificial Intelligence, 2*, 100041. https://doi.org/10.1016/j.caeai.2021.100041

[4] Zhao, J. (2025). The role of learners' AI literacy and resilience in boosting their engagement and motivation in AI-based settings: From an achievement goal theory perspective. *Learning and Motivation, 91*, 102152. https://doi.org/10.1016/j.lmot.2025.102152

[5] Choi, S. (2022). A Study on the AI Literacy Framework . *Journal of The Korean Association of Computer Education, 25*(5), 73-84. https://doi.org/10.32431/kace.2022.25.5.007

[6] Jung, S., & Park, J. (2024). Determinants of AI Literacy : Focusing on AI Usage Experience and Innovativeness. *Korean Journal of Broadcasting & Telecommunications Research*, 137-168. https://doi.org/10.22876/kjbtr.2024..128.005

[7] Laupichler, M. C., Aster, A., & Raupach, T. (2023). Delphi study for the development and preliminary validation of an item set for the assessment of non- experts' AI literacy. *Computers and Education: Artificial Intelligence, 4*, 100126. https://doi. org/10.1016/j.caeai.2023.100126

[8] Laupichler, M. C., Aster, A., Haverkamp, N., & Raupach, T. (2023). Development of the "Scale for the assessment of non-experts' AI literacy"–An exploratory factor analysis. *Computers in Human Behavior Reports, 12*, 100338. https://doi.org/10.1016/j.chbr.2023.100338

[9] Soto-Sanfiel, M. T., Angulo-Brunet, A., & Lutz, C. (2024). The scale of artificial intelligence literacy for all (SAIL4ALL): a tool for assessing knowledge on artificial intelligence in all adult populations and settings. Preprint at arXiv https://osf.io/rgv36

[10] Hwang, H., & Hwang, Y. (2023). A study on Conceptual Constructs of AI literacy with a Focus on AI literacy competence. *Journal of Cybercommunication Academic Society, 40*(2), 89-148. https://doi.org/10.36494/ JCAS.2023.06.40.2.89

[11] Cha, H. (2025). Designing a AI Literacy Course for Non-IT Major Undergraduates in Higher Education : based on Backward Design. *Journal of The Korean Association of Computer Education, 28*(7), 29-42. https://doi. org/10.32431/kace.2025.28.7.003

[12] Kim, S., Kim, S., Park, C., Hong, J., & Park, J. (2023). Development of an AI competency measurement tool for pre-service teachers to enhance expertise in digital education. *Journal of The Korean Association of Computer Education, 26*(4), 21-32. http://dx.doi. org/10.32431/kace.2023.26.4.003

[13] Kim, T., Go, H., & Lee, Y. (2024). Development of Artificial Intelligence Learning Interest Scale for Elementary School Students. *Journal of The Korean Association of Computer Education, 27*(4), 1-11. http://dx.doi. org/10.32431/kace.2024.27.4.001

[14] Kim, S.,& Yoo, G. (2024). Development and Validation of an AI Literacy Rating Scale for Early Childhood Teachers. *Early Childhood Education Research & Review, 28*(4), 31-60. https://doi.org/10.32349/ ECERR.2024.8.28.4.31

[15] Cetindamar, D., Kitto, K., Wu, M., Zhang, Y., Abedin, B., & Knight, S. (2022). Explicating AI literacy of employees at digital workplaces. *IEEE transactions on engineering management, 71*, 810-823. https://doi.org/10.1109/ TEM.2021.3138503

[16] Wang, B., Rau, P. L. P., & Yuan, T. (2023). Measuring user competence in using artificial intelligence: validity and reliability of artificial intelligence literacy scale. *Behaviour & information technology, 42*(9), 1324-1337. https://doi.org/10.1080/0144929X.2022.2072768

[17] Carolus, A., Koch, M. J., Straka, S., Latoschik, M. E., & Wienrich, C. (2023). MAILS-Meta AI literacy scale: Development and testing of an AI literacy questionnaire based on well-founded competency models and psychological change-and meta-competencies. *Computers in Human Behavior: Artificial Humans, 1*(2), 100014. https://doi.org/10.1016/j.chbah.2023.100014

[18] Biagini, G., Cuomo, S., & Ranieri, M. (2024). Developing and validating a multidimensional AI literacy questionnaire: Operationalizing AI literacy for higher education. *In CEUR Workshop Proceedings* (Vol. 3605, pp. 1-15). Edited by Daniele Schicchi, Davide Taibi, Marco Temperini. https://hdl.handle.net/2158/1349235

[19] Lintner, T. (2024). A systematic review of AI literacy scales. *npj Science of Learning, 9*(1), 50. https://doi.org/10.1038/s41539-024-00264-4

[20] Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39-50. https://doi. org/10.1177/002224378101800104

[21] Hair Jnr, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis*. https://digitalcommons.kennesaw.edu/facpubs/2925/

[22] Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal, 6*(1), 1-55. https:// doi.org/10.1080/10705519909540118

[23] Baylor, C., Hula, W., Donovan, N. J., Doyle, P. J., Kendall, D., & Yorkston, K. (2011). An introduction to item response theory and Rasch models for speech-language pathologists. https://doi.org/10.1044/1058-0360(2011/10-0079)

[24] Aubin Le Quéré, M., Schroeder, H., Randazzo, C., Gao, J., Epstein, Z., Perrault, S. T., ... & Li, H. (2024, May). LLMs as research tools: Applications and evaluations in HCI data work. *In Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (pp. 1-7). https://doi.org/10.1145/3613905.3636301

[25] Park, J. (2025). From Learning Management Systems (LMS) to Learning Experience Platforms (LXP): An Empirical Study on Online Learner Performance and Reflective Journaling [Doctoral dissertation]. *Hanyang University*. https:// www. riss. kr/ link? id=T17291940

박진아

· 2012년 경희대학교 한국어학·언론정보학과(학사)
· 2022년 서강대학교 기술경영학과(석사)
· 2025년 한양대학교 기술경영학과(박사)
· 2019년~현재 ㈜에이블런 대표이사

관심분야 : AI 리터러시, 재직자교육, 성인교육, 교육평가

jina@ablearn.kr

류호경

· 1997년 한양대학교 산업공학과(학사)
· 1999년 KAIST 산업공학과(석사)
· 2004년 University of York(U.K) 심리학 (박사)
· 2011~현재 한양대학교 기술경영학과 교수
· 2023~현재 한양대학교 교육혁신 처장

관심분야 : 온라인 러닝, 게임기반학습, 인간-인공지능 협업 의사결정

hryu@hanyang.ac.kr

김지은

· 2007년 KAIST 산업디자인학과(학사)
· 2008년 Arts et Metiers ParisTech (Paris, France) 산업공학과(석사)
· 2011년 Arts et Metiers ParisTech (Paris, France) 산업공학과(박사)
· 2013년~현재 한양대학교 기술경영학과 교수

관심분야 : 에듀테크, UI/UX, 학습동기

jkim2@hanyang.ac.kr

부록

〈Appendix 1〉 The Prompt used for LLM-assisted item review

[LLM Prompt Structure]

1. Context
This study aims to adapt and refine the AI literacy scale (SNAIL), originally developed for non-expert adults, to fit the Korean workplace context.
The table below contains the results of translation and back-translation for 31 items. Each item should be reviewed for linguistic naturalness and contextual appropriateness for Korean-speaking working professionals.
The LLM functions as a language editing and content review assistant, and its output will subsequently be re-examined by an expert focus group (FGI).

2. Task
For each item, evaluate the following four aspects and present structured revision suggestions where necessary.
1) Ambiguity: Detect ambiguous or abstract expressions within the item and suggest clearer alternatives.
2) Double-barreled: Identify whether the item asks for more than one behavior or concept simultaneously.
3) Redundancy: Check for semantic overlap or repetition with other items.
4) Style Consistency: Ensure consistency in the phrasing pattern ("can explain …"), abbreviation usage (English – Korean balance), and difficulty level across items.

Output Format
Present results in tabular form, including the following columns:
1) Item number, Original item, Summary of review findings, Suggested revision (one-sentence or example-based), and Notes (optional).
2) Provide revision suggestions in clear, concise, and descriptive sentences whenever possible.
3) Avoid overly technical jargon or translationese expressions that may sound unnatural to Korean professionals.

〈**Appendix 2**〉Stepwise Item Review Results

Domain	No.	Original translated item	LLM review output	Expert panel decision	Status***
Technical Understanding	1	Explain how a machine learning model is trained, validated, and tested.	Clarify and summarize the procedures and purposes of model training, validation, and testing.	Explain the procedures and purposes of training, validating, and testing a machine learning model.	Revised
	2	Explain how deep learning is related to machine learning.	Specify that deep learning is a subfield of machine learning and describe its structural features and differences.	Explain how deep learning is related to machine learning.	Kept
	3	Explain the differences between rule-based program systems and machine-learning systems.	Describe the differences in decision mechanisms between rule-based and statistical/neural-network-based systems.	Explain the difference between rule-based and machine-learning systems.	Accepted
	4	Explain how an AI application makes decisions.	Describe the principles of decision logic based on rule-, statistical-, and learning-based approaches.	Explain how AI makes decisions.	Kept
	5	Explain the basic principles of reinforcement learning in the context of machine learning.	Describe the reward, policy, and exploration principles of reinforcement learning with examples.	Explain the principles and examples of reinforcement learning in the context of machine learning.	Revised
	6	Explain the differences between strong (general) AI and weak (narrow) AI.	Distinguish between Artificial General Intelligence (AGI) and Narrow AI.	Explain the difference between general AI and narrow AI.	Revised
	7	Explain how computer sensors collect data and how the data are used for AI purposes.	Describe the data flow from sensing → preprocessing → model input.	Explain how data for AI are collected and used.	Revised
	8	Explain what an artificial neural network is.	Describe its components (layers, weights, activation functions) and learning principles.	Explain what an artificial neural network is.	Accepted
	9	Explain how machine learning generally works.	Explain the learning–generalization mechanism through loss minimization and parameter optimization.	Explain how machine learning works.	Kept
	10	Explain the difference between supervised and unsupervised learning in machine learning.	Contrast the two approaches in terms of input–label availability and representative algorithms.	Explain the difference between supervised and unsupervised learning in machine learning.	Kept
	11	Explain the concept of explainable AI (XAI).	Clarify the objectives of explainable AI (transparency, accountability) and describe common methods.	Explain the concept of explainable AI (XAI).	Kept
	12	Explain how some AI systems perceive their environment and act accordingly.	Describe the intelligent system pipeline consisting of perception–decision–action.	Explain how AI perceives its environment and acts accordingly.	Revised
	13	Explain the concept of big data.	Add the five characteristics (5Vs) and provide example applications.	Explain the concept of big data.	Kept
	14	Judge whether the depiction of AI in media such as films or video games is exaggerated beyond reality.	Explain the differences between media portrayals of AI and real AI capabilities or limitations.	Explain the differences between AI in media and real AI.	Accepted
Application and Reflection	15	Explain why privacy protection is important when developing and using AI applications.	Provide legal and ethical grounds and describe associated risks in AI contexts.	Explain why privacy protection is important when developing or using AI.	Kept
	16	Explain why data security must be considered when developing and using AI applications.	Describe the types of security threats and the CIA principles (confidentiality, integrity, availability).	Explain why data security is necessary when developing and using AI.	Revised
	17	Identify ethical concerns related to artificial intelligence.	Distinguish and provide examples of major ethical issues such as bias, discrimination, copyright, and privacy.	Identify and explain ethical issues in AI, including bias, discrimination, copyright, and privacy, with examples.	Accepted

Notes.
*Column labels: Original translated item = item wording after forward–back translation; LLM review output = issues/suggestions flagged by an LLM as a proofreading aid; Expert panel decision = outcome after the FGI (accept / revise / retain), shown here as the finalized wording intent.
**Domain names follow the validated two-factor model: Technical Understanding and Application and Reflection.
***Accepted – The LLM's suggestion was almost fully adopted, with only minor stylistic refinements made by the expert panel.
Revised – The expert panel substantially modified or restructured the item based on the LLM's suggestion, adjusting clarity or difficulty.
Kept – The expert panel reviewed but decided that the original item wording was more appropriate, retaining it without change.