

컴퓨터교육학회 논문지 2026년 제29권 제4호
<https://doi.org/10.32431/kace.2026.29.4.010>



디지털 리터러시 검사 수행형 문항에 대한 성별 차별기능문항 분석*

Differential Item Functioning Analysis of Performance-Based Digital Literacy Assessment by Gender

노혜림[†] · 서정희^{††} · 이현숙^{†††}
 Hyerim Noh[†] · Jeonghee Seo^{††} · Hyunsook Yi^{†††}

요약

본 연구는 2023~2024년에 실시된 초·중학생 대상 국가수준 디지털 리터러시 검사 결과를 활용하여 성별에 따른 차별기능문항(DIF)을 분석하고, 평가 영역과 문항 유형·내용 측면에서 그 양상과 특성을 탐색하였다. 분석 결과, 전반적으로 수행 난이도가 높은 기능형 문항은 남학생에게 유리하게 작용한 반면, 복수 조건을 충족하거나 언어적 판단을 요구하는 선다형 및 저작형 문항은 여학생에게 유리하게 나타났다. 또한 사례 기반 윤리 문항이나 생성형 AI 활용과 같은 개념 이해 문항은 선다형임에도 남학생에게 유리하게 기능하였다. 이러한 결과는 DIF가 문항의 주제, 형식, 수행 조건의 상호작용 속에서 발생하며, 공정한 평가를 위해 문항 설계 전반에 대한 다각적 검토가 필요함을 시사한다.

주제어 국가수준 디지털 리터러시 검사, 성차, 수행형 평가, 차별기능문항

ABSTRACT

This study investigated gender-related Differential Item Functioning (DIF) in the National Digital Literacy Assessment administered to elementary and middle school students in 2023 and 2024. Characteristics of DIF were examined across assessment domains, item formats, and content dimensions. The results showed that functional tasks with higher performance demands tended to favor male students, whereas multiple-choice and performance-based production tasks requiring multiple conditions or language-based judgments tended to favor female students. In contrast, concept-oriented multiple-choice items, including those related to digital ethics and generative artificial intelligence, functioned in favor of male students. These findings suggest that DIF arises from interactions among item content, format, and performance conditions, underscoring the need for a comprehensive review of item design to ensure fair assessment.

Keywords National Digital Literacy Assessment, Differential Item Functioning (DIF), Gender Differences, Performance-based Assessment

†정회원	건국대학교 교육학과 심리측정통계전공 석박사통합과정 수료
††정회원	한국교육학술정보원 연구위원
†††정회원	건국대학교 교육학과 교수 (교신저자)
논문투고	2025년 12월 16일
심사완료	2026년 01월 30일
게재확정	2026년 03월 03일
발행일자	2026년 04월 30일

* 본 논문은 2024년 건국대학교 우수연구인력 양성사업 지원을 받아 수행된 연구임.

* 본 논문은 2025년 한국교육학술정보원의 이슈리포트의 내용을 기반으로 작성하였음(RM 2025-08).

1. 서론

인공지능과 디지털 기술의 급속한 발전은 교육 현장에도 커다란 변화를 가져오고 있다. 이러한 변화에 대응하기 위해 2022 개정 교육과정은 디지털 기초 소양을 디지털 전환 시대의 핵심 역량으로 설정하고, 언어 및 수리 소양과 함께 미래사회를 대비하기 위해 학습자가 갖추어야 할 기본 소양으로 명시하였다[1]. 학습자는 단순히 정보를 습득하는 것을 넘어, 이를 비판적으로 평가하고 디지털 환경 속에서 새로운 가치를 능동적으로 창출할 수 있어야 하며, 이러한 맥락에서 디지털 리터러시의 중요성은 더욱 강조되고 있다.

디지털 리터러시가 현대 사회의 필수 역량으로 자리 잡음에 따라, 학생들의 디지털 리터러시 함양을 위한 체계를 구축하고 그 결과를 교육 정책에 반영하려는 노력이 국제적으로 확산되고 있다[2]. 이에 IEA, OECD, EU, UNESCO 등은 디지털 역량을 정의하고, 이를 바탕으로 디지털 리터러시 수준을 측정하고 있다[3-5]. 특히 IEA는 전 세계 중학생들을 대상으로 컴퓨터·정보 소양을 측정하고, 이와 관련한 교육 맥락변인과의 관계를 파악하기 위해 5년 주기로 ICILS 검사를 시행하고 있다[6]. 우리나라 역시 디지털 리터러시 향상을 위한 교육의 필요성이 강조됨에 따라 2013년부터 ICILS에 참여하고 있으며, 국내에서도 한국교육학술정보원, 한국교육과정평가원, 일부 시도 교육청(경기도교육청, 서울특별시교육청, 부산광역시교육청 등)에서 학생들의 디지털 리터러시 수준을 측정하기 위한 연구가 진행되고 있다[7]. 이 중 한국교육학술정보원은 2009년부터 매년 국가수준에서 초·중학생의 디지털 리터러시 수준을 측정해 왔으며, 연도별 심층분석을 통해 그 결과를 교육 정책에 환류할 수 있는 대표적인 평가 체계를 구축하였다[8].

국가수준 초·중학생 디지털 리터러시 수준 측정 연구는 2007년 'ICT 리터러시 검사' 도구 개발을 시작으로 본격화되었으며, 2018년부터 이를 디지털 리터러시의 개념으로 확장하여 기존의 ICT 영역에 CT 영역을 포함한 이원체제로 초·중학생의 디지털 리터러시를 측정해왔다. 2023년에는 기존의 선다형 검사 방식에서 벗어나 실생활에서 접할 수 있는 시나리오를 기반으로 디지털 리터러시 역량을 종합적으로 측정하는 수행형 검사 체제로 전환하였다.

ICT 리터러시 수준 측정 결과를 살펴보면, 여학생이 남학생보다 높은 성취를 보이는 경향이 여러 해에 걸쳐 지속적으로 확인되었다[8-10]. 이러한 결과는 단순한 평균 차이를 넘어, 평가 도구가 특정 성별에게 유리하거나 불리하게 작용하고 있을 가능성에 대한 체계적 검토의 필요성을 제기한다. 실제로 Gebhardt 외(2019)는 디지털 리터러시 평가에서 성별에 따른 응답 행동의 차이가 문항 해석과 정답 확률에 영향을 미칠 수 있음을 지적하였으며[11], Geske와 Ozola(2010)는 평가 도구의 타당한 활용을 위해 문항 수준에서 성별에 따른 차별기능문항 분석을 선행할 필요가 있음을 강조하였다[12]. 또한 Harputlu와 Güray(2024), Jeysankar와 Nachiappan(2021)의 연구에서도 성별 차

이는 단순한 디지털 기술 활용 격차를 넘어, 평가 과정과 결과의 공정성 문제로 확장될 수 있음을 지적하였다[13, 14].

즉, 성별 간 평균 성취도 차이가 실제 능력 차이에 기인한 것인지, 아니면 특정 문항이 특정 성별에게 유리하거나 불리하게 작용한 결과인지 체계적인 검토가 필요하다. 이는 단순히 성차의 원인을 파악하고 설명하는 것을 넘어, 평가 도구의 구조적 편향 가능성을 점검함으로써, 향후 디지털 리터러시 평가의 타당성과 교육적 형평성을 제고하기 위한 기초자료로 기능할 수 있다. 이에 본 연구에서는 디지털 리터러시 검사의 신뢰성과 공정성을 확보하기 위해 성별에 따른 차별기능문항(Differential Item Functioning, DIF) 분석을 실시하고, 차별기능문항이 나타나는 평가영역, 하위영역, 문항 유형 등의 특성과 양상을 함께 탐색하고자 하였다.

이와 같은 연구 목적에 따라 다음과 같은 연구문제를 설정하였다.

첫째, 2023, 2024년도 초·중학생 디지털 리터러시 검사에서 성별에 따른 차별기능문항이 존재하는가?

둘째, 2023, 2024년도 초·중학생 디지털 리터러시 검사에서 성별에 따라 도출된 차별기능문항의 특징은 평가영역, 문항 유형 등에 따라 어떠한 특성을 보이는가?

2. 이론적 배경

2.1 디지털 리터러시의 정의 및 측정

급속한 디지털 대전환 시대에 디지털 리터러시는 사회 구성원이 일상생활을 영위하는 데 필수적으로 갖추어야 하는 기초 소양으로서 그 중요성이 더욱 부각되고 있다[3]. 이와 함께 디지털 리터러시의 개념과 구성요소에 대한 논의도 지속적으로 확장되어 왔다. Gilster(1997)는 디지털 리터러시를 정보에 대한 접근, 이해, 평가, 활용 능력으로 정의하였으며[15], Mills(2010)는 이를 디지털 환경에서의 창의성, 협업, 윤리적 실천까지 아우르는 역량으로 확장하였다[16]. 이후 디지털 리터러시는 단순히 디지털 매체를 사용할 수 있는 능력을 넘어, 정보를 창조·공유하는 능력, 타인과의 협업과 의사소통 역량, 디지털 사회의 시민으로서 요구되는 윤리적 인식과 태도까지 포괄하는 역량으로 발전되었다. 구체적으로 JISC(2015)는 디지털 소양을 6개 영역으로 나누고, ICT 능력을 기반으로 소통, 협력, 문제해결뿐만 아니라 디지털 복지까지 포함하여 개인이 디지털 사회에서 학습하고 일하며 생활하는 데 필요한 역량으로 정의하였다[17]. 유럽 연합(EU)은 디지털 역량 프레임워크 DigComp 2.2를 통해 디지털 정보 기술에 대한 이해와 활용을 기반으로 유럽 시민이 갖추어야 할 디지털 역량을 제시하였다[18]. UNESCO는 DigComp 2.0을 초기 틀로 참고하여 성인 인구의 디지털 리터러시를 모니터링하고 평가할 수 있는 글로벌 역량 체계를 제안하였으며, 디지털 리터러시를 '디지털 기술을 통해 정보를 안전하고 적절하게 접근, 관리, 이해, 통합, 소통, 평

가, 창조할 수 있는 능력'으로 정의하였다.

한편, PISA는 2025년 조사에서 디지털 리터러시 역량을 측정하기 위해 '디지털 세상에서의 학습(Learning in the Digital World, LDW)' 영역을 새롭게 도입하였다. LDW는 컴퓨터 기반 문제해결 역량과 자기조절 학습 역량으로 구성되며, 디지털 환경에서 학습과 관련된 핵심 역량을 진단하는 것을 목적으로 한다[3]. 이에 비해 IEA의 ICILS, 호주의 NAP, 홍콩의 DLA 등은 이미 국제 및 국가 수준에서 디지털 리터러시를 체계적으로 측정·평가하고 있다.

이처럼 디지털 리터러시의 중요성에 대한 인식이 증대되면서, 국내에서도 그 정의와 구성 요소에 대한 논의와 함께 이를 측정하기 위한 연구가 활발하게 이루어지고 있다. 초기 연구들은 디지털 리터러시를 정보 탐색, 분석, 활용, 의사소통, 문제해결 역량을 포함하는 개념으로 정의하였으며, 최근에는 비판적 사고, 윤리적 태도, 디지털 시민성 등 사회문화적 역량까지 포괄하는 통합적·복합적 역량으로 개념을 확장하고 있다[18].

이러한 정의에 기반하여 디지털 리터러시를 측정하기 위한 다양한 평가 도구도 개발되어 왔다. 예를 들어, 백순근 외(2009)는 학생들이 실제 산출물을 생성하는 과정을 통해 디지털 역량을 진단할 수 있도록 수행형 ICT 리터러시 검사 도구를 제안하였다[19]. 그러나 대부분의 국내 연구에서는 리커트 척도 기반의 자기보고식 설문을 활용하여 디지털 역량을 측정하고 있으며[20, 21], 이는 실제 수행 능력을 정밀하게 진단하는 데 한계를 지닌다. 이 한계를 보완하기 위해 국가차원에서 보다 타당한 평가 체계가 요구되었고, 한국교육학술정보원은 선다형 문항 중심의 검사 도구를 개발하여 학생들의 디지털 리터러시를 측정하였다. 이후 디지털 리터러시 개념이 점차 복잡적이고 역동적으로 변화함에 따라, 평가의 타당성과 현실 적합성을 제고하기 위해 2023년부터는 새로운 평가 프레임워크에 기반한 시나리오 중심의 수행형 평가 체계를 도입하였다.

해당 도구는 실생활 맥락을 반영한 모듈 기반으로 구성되며, 학교급별로 총 4개의 시나리오를 제공한다. 각 시나리오는 디지털 리터러시 프레임워크의 5개 평가 영역 가운데 3~5개 영역을 통합적으로 포함하도록 설계되었다. 문항 유형은 기능형, 알고리즘형, 연결형, 저작형으로 구분되며, 개인 과제와 협력 과제 맥락에서 실제 생활에 요구되는 디지털 역량을 종합적으로 진단하도록 구성하였다. 각 문항 유형의 정의는 다음과 같다[2, 22]. 먼저 기능형 문항은 디지털 도구나 기능을 직접 조작하여 과제를 수행하는 유형이며, 알고리즘형 문항은 주어진 조건에 따라 논리적 절차를 구성하는 문항이다. 연결형 문항은 둘 이상의 정보 항목을 의미적 관련성에 따라 선으로 연결하거나 분류상자에 배치하는 유형이며, 저작형 문항은 제시된 자료를 바탕으로 도표, 프레젠테이션 슬라이드 등을 직접 작성하거나 수정하는 문항이 해당된다.

2.2 차별기능문항 개념 및 분석 방법

차별기능문항은 검사 문항이 측정하고자 하는 특성

이나 능력 외의 요인에 의해 집단 간 문항 수행(item performance)에 대한 기대치가 달라지는 문항을 말하며, 이는 잠재 능력이 동일함에도 불구하고 응답자가 속한 집단에 따라 문항에 대한 정답 확률이 달라지는 현상을 의미한다[23].

DIF를 분석하기 위해서는 비교 대상이 되는 두 집단을 설정하는 것이 전제되며, 이때, 일반적으로 사례수가 많은 집단이나 기준이 되는 집단을 준거집단(reference group), 상대적으로 사례수가 적거나 연구자가 주요 관심을 두는 집단을 초점집단(focus group)으로 설정한다. 이분 문항을 대상으로 DIF를 분석할 경우, DIF로 추출된 문항은 균일(uniform) 차별기능문항과 비균일(non-uniform) 차별기능문항으로 구분될 수 있다. 균일 DIF는 문항의 난이도가 집단에 따라 일관되게 차이가 나는 경우로, 모든 능력 수준에서 한 집단이 다른 집단보다 항상 유리하거나 불리한 성향을 보이는 문항을 의미한다. 반면, 비균일 DIF는 응답자의 능력 수준과 문항 수행의 상호작용에 따라 특정 구간에서는 한 집단에게 유리하게 작용하다가, 다른 구간에서는 그 관계가 역전되어 다른 집단에게 유리하게 나타나는 문항을 의미한다[24, 25].

DIF 분석을 위한 통계적 접근은 다양한 관점에서 제안되어 왔다. 대표적으로 Mantel-Haenszel(MH), SIBTEST 방법, Wald-test, 로지스틱 회귀분석 그리고 IRT Likelihood Ratio(IRT-LR), Lord의 χ^2 방법 등이 사용된다. 이러한 방법들은 각기 다른 이론적 기반과 통계적 가정에 따라 발전해 왔으며, 각 방법은 고유의 계산 절차와 분석 목적에 따라 장단점이 존재한다. 이 중 MH, SIBTEST 방법은 균일 DIF를 탐지하는 데는 효과적이거나, 비균일 DIF를 식별하는 데에는 한계가 있다. 그러나 실제 평가 상황에서는 피험자의 이질적인 특성이 복합적으로 작용할 경우, 능력 수준에 따른 응답양상이 일관되지 않아 비균일 DIF가 발생할 수 있다. 따라서 본 연구에서는 DIF의 방향이나 크기가 달라지는 비균일 DIF까지 탐지할 수 있는 분석 방법인 로지스틱 회귀분석, Lord의 χ^2 , IRT 우도비 검정 방법을 활용하고자 한다.

2.2.1 로지스틱 회귀분석

로지스틱 회귀분석은 문항점수를 종속변수로, 집단(G), 검사 총점(X), 집단과 검사 총점의 상호작용(XG)을 독립변수로 설정하여 회귀계수를 추정한 후, 단계별 모형 비교를 통해 DIF 여부를 판단하는 방법이다[26]. 이 분석에서 종속변수는 특정 문항에 대해 피험자가 정답을 선택할 확률을 의미하며, 이는 식(1)에 따라 로짓 함수 형태로 표현된다.

$$\text{logit}(P) = \beta_0 + \beta_1 X + \beta_2 G + \beta_3 XG \quad (1)$$

여기서 $\beta_2 \neq 0$ 이고 $\beta_3 = 0$ 일 경우 균일 DIF에 해당하며, $\beta_2 > 0$ 라면 참조집단에게, 반대의 경우 초점집단에게 유리하게 기능하는 문항에 해당한다. 반면, β_2 와 무관하게 $\beta_3 \neq 0$ 라면 비균일 DIF에 해당한다. 비균일 DIF의 경

우 $\beta_3 > 0$ 이면 상위능력 집단에서는 참조집단에게, 하위능력 집단에서는 초점집단에게 유리하게 기능함을 의미하며, $\beta_3 < 0$ 인 경우 반대 상황에 해당한다. 균일과 비균일 DIF를 동시에 검증하기 위해서는 세 가지 위계적 회귀모형을 설정한 후, 순차적으로 적합도를 비교함으로써 DIF의 존재 여부를 판단한다.

$$\begin{aligned} \text{모형1: 완전모형(full model): } & \beta_0 + \beta_1 X + \beta_2 G + \beta_3 XG \\ \text{모형2: 축소된 모형(reduced model): } & \beta_0 + \beta_1 X + \beta_2 G \\ \text{모형3: 영모형(null model): } & \beta_0 + \beta_1 X \end{aligned} \quad (2)$$

모형 간 차이 비교를 식으로 표현하면 식(3)과 같으며, G_1 과 G_2 는 자유도가 1인 χ^2 분포를 따른다. G_1 은 비균일 DIF 여부를 판단하기 위한 지수로 완전모형과 축소된 모형 간 차이를 검정하는데 사용되며, G_2 는 영모형과 축소된 모형을 다루는 균일 DIF 여부를 판단하는데 활용된다.

$$\begin{aligned} G_1 &: [-2\ln(\text{모형2의 최대우도})] - [-2\ln(\text{모형1의 최대우도})] \\ G_2 &: [-2\ln(\text{모형3의 최대우도})] - [-2\ln(\text{모형2의 최대우도})] \end{aligned} \quad (3)$$

로지스틱 회귀분석은 균일과 비균일 DIF를 동시에 검증할 수 있을 뿐만 아니라, 효과크기(effect size)까지 함께 고려할 수 있다는 장점이 있다. 본 연구에서는 Nagelkerke 값 R^2 을 사용하여 효과 크기를 평가하며, 효과 크기에 대한 판단 기준은 Zumbo와 Thomas(1997)의 연구에서 사용한 기준을 참고하였다[27]. 그 판단 기준은 (4)와 같다.

$$\begin{aligned} \Delta R^2 < 0.035: & \text{A 수준(매우 작은 수준)} \\ 0.035 \leq \Delta R^2 < 0.070: & \text{B 수준(중간 수준)} \\ \Delta R^2 \geq 0.070: & \text{C 수준(큰 수준)} \end{aligned} \quad (4)$$

2.2.2 Lord의 χ^2 검정

Lord의 χ^2 검정 방법은 문항반응이론을 기반으로 참조집단과 초점집단에 대해 문항 모수를 각각 추정한 후, 문항 모수 차이를 통계적으로 검정하여 DIF를 추출하는 방법이다[28]. 이때, 집단 간 문항 모수가 동일하다는 귀무가설을 설정하고, 문항 모수 차이에 대한 χ^2 검정 통계량을 산출하여 유의성을 평가한다. 특히, 문항 난이도와 변별도를 동시에 고려할 경우, 두 모형 간 차이를 통합적으로 반영한 검정 통계량은 식(5)를 통해 계산된다[29].

$$\begin{aligned} \chi_j^2 &= V_j^T \sum_j^{-1} V_j, \quad \sum_j = \sum_{F_j} + \sum_{R_j} \quad (5) \\ V_j &= \begin{bmatrix} \hat{a}_{F_j} - \hat{a}_{R_j} \\ \hat{b}_{F_j} - \hat{b}_{R_j} \end{bmatrix}, \quad \sum_{G_j} = \begin{bmatrix} \hat{\sigma}_{aG_j}^2 & \hat{\sigma}_{aG_j, bG_j} \\ \hat{\sigma}_{aG_j, bG_j} & \hat{\sigma}_{bG_j}^2 \end{bmatrix}, \quad G \in \{F, R\} \end{aligned}$$

V_j 는 하위 집단 간 변별도 및 난이도 차이 벡터를 의미하

며, Σ 는 하위 집단의 문항 모수에 대한 분산-공분산 행렬에 해당한다. 이때, 계산된 χ^2 검정 통계량이 통계적으로 유의한 경우, 해당 문항이 균일 또는 비균일 차별기능문항으로 판별될 수 있다.

2.2.3 IRT 우도비 검정

IRT 우도비 검정(Likelihood Ratio Test)은 문항반응이론에 기반하여 모수를 추정한 후, 집단 간 차이를 검정하는 방식으로 Thissen, Steinberg, Wainer(1988)에 의해 제안되었다[30]. 이 방법은 주변최대우도추정(Marginal Maximum Likelihood Estimation) 알고리즘을 활용하여 참조집단과 초점집단 각각의 문항 모수를 추정된 뒤, 두 집단 간 문항특성곡선(Item Characteristic Curve; ICC)이 동일하다는 귀무가설 하에, 동일성 제약을 적용한 축소 모형(compact model, C)과 이를 해제한 확장 모형(augmented model, A)을 구성한다. 이후 두 모형 간의 차이를 χ^2 검정으로 평가하며, 이때 산출되는 LRT 통계치인 G^2 값을 바탕으로 DIF를 판별한다. 이를 통해 균일 DIF와 비균일 DIF를 모두 탐지할 수 있으며, 우도비 검정 통계량을 식으로 표현하면 (6)과 같다[31].

$$G^2 = -2\ln \frac{L_c}{L_A} = -2(\ln L_c - \ln L_A) \sim \chi_{df}^2 \quad (6)$$

식(6)에서 $\ln L$ 은 각 모형에서 산출된 최대 우도 추정치 상에서의 로그 우도 값을 의미하며, $-2\ln L$ 은 이탈도(deviance)로 정의된다[32]. 두 이탈도의 차이는 영가설 하에서 χ^2 분포를 따르며, 이때, 단일 문항에 사용되는 문항 모수의 수가 자유도로 설정된다. 본 연구에서는 2PL IRT 모형을 활용하였으므로 자유도는 2로 설정하였다. 검정 통계량 G^2 값이 모수에 통계적으로 유의한 차이가 있다고 판단되면, 개별 모수에 대한 검정을 실시하여 차별기능문항의 유형을 파악할 수 있다[33].

3. 연구방법

3.1 분석 대상

본 연구에서는 2023년, 2024년 국가수준 디지털 리터러시 검사에 응시한 초등학교 4~6학년, 중학교 1~3학년 학생들의 응답자료를 분석자료로 활용하였다. 해당 검사는 전국 초등학교 4~6학년, 중학교 1~3학년 학생 수의 약 1%를 대상으로 유층 무선 표집을 통해 표본을 구성하였으며, 2023년은 32,804명, 2024년은 37,957명이 참여하였다. 먼저, 초등학교의 경우 2023년에는 남학생 7,841명(51.32%), 여학생 7,437명(48.68%)으로 총 15,278명이 분석에 포함되었으며, 2024년에는 남학생 9,711명(51.57%), 여학생 9,120명(48.43%)으로 총 18,831명이 포함되었다. 중학교는 2023년에 남학생 8,830명(50.38%), 여학생 8,696명(49.62%)으

로 총 17,526명이 포함되었으며, 2024년에는 남학생 9,511명(49.73%), 여학생 9,615명(50.27%)으로 총 19,126명이 분석에 활용되었다. 초등학교와 중학교 모두 성별 비율이 유사하였고, 연도별로도 큰 차이는 나타나지 않았다.

Table 1. Number of Participants by School Level and Gender

※ Unit: n (%)

Gender	Elementary school		Middle School	
	2023	2024	2023	2024
Male (F)	7,841 (51.32)	9,711 (51.57)	8,830 (50.38)	9,511 (49.73)
Female (R)	7,437 (48.68)	9,120 (48.43)	8,696 (49.62)	9,615 (50.27)
Total	15,278 (100.00)	18,831 (100.00)	17,526 (100.00)	19,126 (100.00)

※ F: Focal group, R: Reference group

3.2 측정 도구

국가수준 디지털 리터러시 검사는 컴퓨터 기반의 수행형 평가 도구로 5개의 평가영역 및 11개의 하위요소로 구성되어 있다. 평가영역별 문항 구성을 살펴보면, 초등학교는 ‘A. 디지털 도구’ 영역은 4문항(15.38%), ‘B. 디지털 정보·데이터’ 6문항(23.08%), ‘C. 디지털 의사소통 및 협력’ 4문항(15.38%), ‘D. 디지털 자원 생산’ 4문항(15.38%), ‘E. 디지털 안전과 건강’ 8문항(30.77%)으로 구성되어 있으며, ‘E’ 영역의 문항 비중이 가장 높았다. 중학교는 ‘A. 디지털 도구’ 영역은 6문항(23.08%), ‘B. 디지털 정보·데이터’ 6문항(23.08%), ‘C. 디지털 의사소통 및 협력’ 4문항(15.38%), ‘D. 디지털 자원 생산’ 5문항(19.23%), ‘E. 디지털 안전과 건강’ 5문항(19.23%)으로 구성되어 있으며, ‘A’ 및 ‘B’ 영역의 문항 비중이 가장 높았다(부록 Table 2 참고).

문항유형별 분포를 살펴보면, 초등학교는 기능형 15문항(57.69%), 선다형 8문항(30.77%), 알고리즘형 1문항(3.85%), 저작형 2문항(7.69%)으로 기능형의 비율이 가장 높았다. 중학교는 기능형 10문항(38.46%), 선다형 9문항(34.62%), 알고리즘형 3문항(11.54%), 연결형 1문항(3.85%), 저작형 3문항(11.54%)으로 기능형과 선다형의 비율이 유사한 비중으로 높게 나타났다(부록 Table 3 참고).

3.3 분석 모형 및 방법

국가수준 디지털 리터러시 검사에서는 남학생의 성취가 여학생에 비해 낮게 나타나는 경향이 여러 해에 걸쳐 일관되게 확인되었다. 이에 본 연구에서는 성취 수준이 상대적으로 낮은 남학생을 초점집단(focal group), 여학생을 참조집단(reference group)으로 설정하여 분석을 실시하였다.

성별에 따라 집단을 구분할 경우 초등학교와 중학교 모두 집단별 사례수가 7,000명 이상으로 표본의 크기가 매우 크다. 이처럼 집단별 사례수가 클 경우 집단 간 차이가 존재하지 않아도 차이가 있다고 판별하는 제1종 오류가 발생할 가능성이 높다[32, 34]. 이에 본 연구에서는 여러 선행연구에서 제안된 기준에 따라 두 개 이상의 DIF 검정 방법에서 동시에 유의하게 나타난 문항을 차별기능문항으로 선정하였으며, 이때, p값이 .01 미만인 문항을 차별기능문항으로 간주하였다[9, 31]. 성별에 따른 차별기능문항을 추출하기 위해 균일 DIF와 비균일 DIF를 모두 탐지할 수 있는 로지스틱 회귀분석과 Lord의 χ^2 검정 방법, 우도비 검정 방법을 적용하였다. 로지스틱 회귀분석, Lord의 검정은 R의 difR 패키지를[35], IRT 우도비 χ^2 검정은 mirt 패키지를 사용하여 분석하였다[36].

4. 연구결과

본 장의 표에서 F는 여학생(참조집단), M은 남학생(초점집단)을 의미하며, 2023년, 2024년 디지털 리터러시 검사의 성별 기술통계 결과는 부록 Table 4에 제시하였다.

4.1 차별기능문항 추출 결과

4.1.1 초등학교

초등학생을 대상으로 2023년, 2024년 디지털 리터러시 검사에 대한 DIF 문항을 탐색하기 위해 로지스틱 회귀분석, χ^2 검정 방법, IRT 우도비 검정 방법을 적용하였으며, 분석 결과는 Table 2와 같다. 먼저 2023년의 경우 26문항 중 12문항이 세 분석 방법에서 공통적으로 DIF 문항으로 추출되었으며, 이 중 5개 문항(3-3, 4-4, 4-6, 4-7, 4-8)은 비균일 DIF, 7개 문항(1-4, 2-2, 2-4, 3-4, 4-1, 4-2, 4-9)은 균일 DIF로 추출되었다. 구체적으로 비균일 DIF 문항은 모두 상위 집단에서 여학생에게 유리하게 나타났으며, 균일 DIF 문항 역시 7개 문항 중 5개 문항이 여학생에게 유리하였다. 한

Table 2. Results of Differential Item Functioning Analysis by Gender (Elementary School)

Items	Item Type	2023					2024						
		Logistic Regression		χ^2 test	IRT LR	DIF direction	DIF Items	Logistic Regression		χ^2 test	IRT LR	DIF direction	DIF Items
		χ^2	ΔR^2	χ^2	χ^2			χ^2	ΔR^2	χ^2	χ^2		
1-1	F	0.58	A	6.29	30.36***	1-1	F	1.63	A	1.50	26.84***		
1-2	F	0.70	A	7.02	32.96***	1-2	F	81.10***	A	63.79***	0.15		
1-3	F	22.96***	A	6.89	11.28***	1-3	F	24.21***	A	30.77***	107.59***	(+)	O
1-4	MC	92.54***	A	100.89***	208.03***	F	O	47.54***	A	49.16***	157.63***	F	O

2023							2024								
Items	Item Type	Logistic Regression		χ^2 test	IRT LR	DIF direction	DIF Items	Items	Item Type	Logistic Regression		χ^2 test	IRT LR	DIF direction	DIF Items
		χ^2	ΔR^2	χ^2	χ^2					χ^2	ΔR^2	χ^2	χ^2		
2-1	MC	2.49	A	10.70**	39.40***			2-1	MC	8.25	A	6.68	73.36***		
2-2	F	137.90***	A	138.79***	241.94***	F	O	2-2	F	2.89	A	5.07	81.72***		
2-3	MC	1.78	A	3.96	35.37***			2-3	MC	15.81***	A	18.64***	93.77***	F	O
2-4	F	135.89***	A	59.24***	22.95***	M	O	2-4	F	85.48***	A	46.52***	3.96		
2-5	A	152.88***	A	51.84***	7.66			2-5	A	114.78***	A	60.97***	2.70		
2-6	MC	2.63	A	10.94**	51.81***			2-6	MC	0.29	A	1.46	37.97***		
3-1	F	15.56***	A	8.29	34.28***			3-1	F	37.24***	A	23.19***	39.66***	M	O
3-2	MC	6.40	A	1.69	73.36***			3-2	MC	6.24	A	0.87	108.38***		
3-3	F	16.87***	A	25.34***	24.15***	(+)	O	3-3	F	5.66	A	10.80	85.10***		
3-4	P	41.58***	A	61.81***	176.42***	F	O	3-4	P	10.91***	A	17.10***	89.66***	F	O
3-5	F	6.80	A	14.40***	72.70***			3-5	F	21.35***	A	22.26***	189.53***	(+)	O
3-6	MC	6.88	A	19.39***	106.04***			3-6	MC	32.09***	A	28.93***	173.13***	(+)	O
4-1	F	41.76***	A	36.65***	17.11***	M	O	4-1	F	70.75***	A	65.61***	13.91**	M	O
4-2	F	15.23***	A	34.21***	146.50***	F	O	4-2	F	93.48***	A	84.06***	274.62***	F	O
4-3	F	8.57	A	20.67***	55.09***			4-3	F	5.76	A	10.19	90.22***		
4-4	MC	21.63***	A	11.23**	27.96***	(+)	O	4-4	MC	25.80***	A	19.68***	58.35***	(+)	O
4-5	F	3.46	A	15.98***	112.73***			4-5	F	11.41**	A	22.06***	130.32***	(+)	O
4-6	F	11.65**	A	29.76***	106.19***	(+)	O	4-6	F	3.68	A	5.88	152.41***		
4-7	MC	11.64**	A	34.51***	134.11***	(+)	O	4-7	MC	9.07	A	14.58**	136.13***		
4-8	F	13.08**	A	22.02***	65.48***	(+)	O	4-8	F	32.32***	A	22.02***	33.71***	M	O
4-9	P	67.44***	A	78.68***	247.79***	F	O	4-9	P	135.75***	A	90.46***	359.62***	F	O
4-10	F	1.63	A	9.34**	125.14***			4-10	F	13.99***	A	10.33**	85.86***	(-)	O

** $p < .01$, *** $p < .001$

※ Note 1, (+): a non-uniform DIF item that functions in favor of male students (focal group) in the lower ability group and in favor of female students (reference group) in the higher ability group.

※ Note 2, (-): a non-uniform DIF item that functions in favor of female students (reference group) in the lower ability group and in favor of male students (focal group) in the higher ability group.

편, DIF로 추출된 12문항에 대한 차별기능문항 효과의 크기는 'A 수준'으로 매우 작은 크기로 나타났다.

2024년 분석 결과를 살펴보면, 26문항 중 14문항이 DIF로 추출되었으며, 이 중 6개 문항(1-3, 3-5, 3-6, 4-4, 4-5, 4-10)은 비균일 DIF, 8개(1-4, 2-3, 3-1, 3-4, 4-1, 4-2, 4-8, 4-9)는 균일 DIF로 추출되었다. 비균일 DIF는 4-10번 문항을 제외한 모든 문항이 상위집단으로 갈수록 여학생에게 유리하게 기능하였고, 균일 DIF 문항 역시 8문항 중 5문항이 여학생들에게 유리한 방향으로 나타났다. DIF 효과 크기 분석 결과, 모든 문항이 'A 수준'으로 확인되었다. 전반적으로 효과 크기는 작으나 2023년과 2024년 검사에서 공통적으로 여학생에게 유리한 방향으로 작용하는 DIF 문항이 다수 존재하였다.

4.1.2 중학교

중학교 학생을 대상으로 한 2023, 2024년 디지털 리터러시 검사에서의 성별 DIF 분석 결과는 Table 3에 제시하였다. 전체 26문항 중 2023년은 18문항, 2024년은 20문항에서 DIF가 존재하였으며, 효과 크기는 모두 'A수준'으로 작은 수준이었다. 구체적으로 2023년은 비균일 DIF 9문항, 균

일 DIF 9문항이 추출되었고, 비균일 DIF는 모두 하위 능력 집단에서 남학생이, 상위 능력 집단에서 여학생이 유리하게 나타났다. 균일 DIF 문항 중 1-4, 2-1, 2-8번 문항은 남학생에게 유리하게 기능하였으며, 1-6, 2-2, 2-7, 2-9, 3-4, 3-5번 문항은 여학생에게 유리하게 작용하여 전반적으로 여학생들에게 유리한 문항이 더 많은 것을 확인할 수 있다.

2024년은 비균일 DIF가 8문항, 균일 DIF가 12문항으로 추출되었으며, 비균일 DIF는 2023년과 동일하게 모든 문항이 상위 능력 집단으로 갈수록 여학생에게 유리하게 작용하였다. 균일 DIF는 12개 문항 중 5문항(2-5, 2-7, 2-8, 3-1, 3-3)은 남학생에게, 7문항(1-5, 1-6, 1-7, 2-1, 2-6, 3-2, 3-5)은 여학생에게 유리하였다. 중학교 2024년의 경우 2023년에 비해 남학생에게 유리한 문항이 상대적으로 많이 추출되었으나, 전체적으로 여학생에게 유리하게 작용하는 문항이 여전히 다수 존재하였다. 이러한 경향은 초등학교와 중학교에서 공통적으로 확인되었으며, 성별 DIF가 전반적으로 여학생에게 유리한 방향으로 나타남을 시사한다.

4.2 차별기능문항의 문항유형별 특성 분석 결과

분석을 통해 추출된 문항을 평가영역 및 하위요소, 문항 유형, 문항 내용을 토대로 차별기능문항의 특성을 체계적으로 분석하였다. DIF로 추출된 문항은 수행기반 문항과 지식기반 문항으로 구분할 수 있으며, 이를 토대로 문항의 특

성을 비교·분석하였다. 수행기반 문항은 문제해결 과정에서 실제로 디지털 기술을 활용하여 역량을 평가하는 기능형, 저작형, 알고리즘형, 연결형이 포함된다. 반면, 지식기반 문항은 지식 및 이론적 개념에 대해 평가하는 선다형이 포함된

Table 3. Results of Differential Item Functioning Analysis by Gender (Middle School)

		2023						2024									
Items	Item Type	Logistic Regression		χ^2 test		IRT LR χ^2	DIF direction	DIF Items	Items	Item Type	Logistic Regression		χ^2 test		IRT LR χ^2	DIF direction	DIF Items
		χ^2	ΔR^2	χ^2	ΔR^2						χ^2	ΔR^2	χ^2	ΔR^2			
1-1	F	15.29***	A	10.87**	41.64***	(+)	O	1-1	F	46.73***	A	11.96**	19.06***	(+)	O		
1-2	F	39.13***	C	16.53***	8.97			1-2	F	57.68***	A	36.10***	104.97***	(+)	O		
1-3	MC	30.98***	A	46.28***	254.03***	(+)	O	1-3	MC	55.36***	A	54.90***	437.87***	(+)	O		
1-4	F	21.39***	A	12.19**	24.19***	M	O	1-4	F	36.90***	A	10.47**	90.37***	(+)	O		
1-5	F	15.00***	A	6.93	129.82***			1-5	F	54.27***	A	49.26***	439.21***	F	O		
1-6	P	90.26***	A	91.78***	448.78***	F	O	1-6	P	30.04***	A	18.62***	498.47***	F	O		
1-7	MC	2.40	A	8.43	117.95***			1-7	MC	45.65***	A	13.01**	630.70***	F	O		
2-1	MC	137.53***	A	102.97***	34.01***	M	O	2-1	MC	18.74***	A	11.52**	479.89***	F	O		
2-2	MC	16.52***	A	22.72***	328.98***	F	O	2-2	MC	64.09***	A	48.23***	580.91***	(+)	O		
2-3	MC	51.93***	A	57.85***	429.26***	(+)	O	2-3	MC	6.37	A	8.99	347.64***				
2-4	F	20.21***	A	7.33	278.49***			2-4	F	21.55***	A	6.58	507.72***				
2-5	MC	30.98***	A	21.77***	385.64***	(+)	O	2-5	MC	15.02***	A	45.17***	339.55***	M	O		
2-6	F	16.66***	A	14.44***	129.96***	(+)	O	2-6	F	120.50***	A	75.29***	723.58***	F	O		
2-7	P	156.79***	A	130.94***	619.10***	F	O	2-7	P	76.06***	A	34.78***	10.36***	M	O		
2-8	F	156.97***	A	82.06***	23.90***	M	O	2-8	F	129.14***	A	78.96***	100.48***	M	O		
2-9	MC	30.08***	A	31.89***	439.07***	F	O	2-9	MC	13.17**	A	4.50	470.99***				
3-1	F	2.59	A	0.05	259.64***			3-1	F	10.05**	A	21.62***	333.31***	M	O		
3-2	M	6.42	A	6.33	120.12***			3-2	M	32.08***	A	23.55***	486.62***	F	O		
3-3	MC	21.97***	A	25.91***	169.85***	(+)	O	3-3	MC	49.30***	A	27.37***	79.76***	M	O		
3-4	F	12.91**	A	18.74***	129.38***	F	O	3-4	F	4.61	A	2.74	319.85***				
3-5	P	94.28***	A	66.04***	623.93***	F	O	3-5	P	86.92***	A	42.23***	742.19***	F	O		
3-6	F	17.82***	A	31.39***	309.34***	(+)	O	3-6	F	7.39	A	13.52**	330.83***				
4-1	MC	51.21***	A	49.66***	177.10***	(+)	O	4-1	MC	3.55	A	4.10	247.04***				
4-2	A	49.50***	A	19.05***	17.34***	(+)	O	4-2	A	134.56***	A	100.95***	94.59***	(+)	O		
4-3	A	32.52***	A	7.98	81.82***			4-3	A	153.68***	A	110.52***	58.79***	(+)	O		
4-4	A	163.81***	A	74.72***	0.13			4-4	A	82.55***	A	62.16***	61.99***	(+)	O		

** $p < .01$, *** $p < .001$

※ Note 1, (+): a non-uniform DIF item that functions in favor of male students (focal group) in the lower ability group and in favor of female students (reference group) in the higher ability group.

※ Note 2, (-): a non-uniform DIF item that functions in favor of female students (reference group) in the lower ability group and in favor of male students (focal group) in the higher ability group.

다.

4.2.1 수행기반 문항(Performance-based items)

1) 초등학교

먼저, 초등학교 분석결과(부록 Table 5 참고)를 살펴보면, 2023년, 2024년 모두 기능형, 저작형 문항에서만 DIF 문항이 추출되었다. 기능형 문항 중 디지털 도구의 기술 조작 및 기능 활용을 측정하는 문항은 대부분 남학생에게, 디지털 협업 및 창의적 콘텐츠 제작과 관련된 문항은 여학생에게 유리한 경향을 보였다.

구체적으로 QR 코드 인식(2023년, 2024년: 4-1), 파일 저장(2023년: 4-6), 이미지 URL 복사(2024년: 4-8), 코딩 및 알고리즘 설계(2023년: 2-4)와 같이 수행 난이도가 높은 기능형 문항에서는 남학생이 상대적으로 더 유리한 것으로 나타났다. 반면, 수행 난이도가 비교적 낮고, 온라인 의사소통 및 협업 활동과 관련된 문항에서는 여학생이 더 유리한 양상을 보였다. 이러한 결과는 남학생이 기술적 활용에서, 여학생이 정보 검색 및 소통 활용에서 강점을 보인다고 한 Hargittai(2010)의 연구와 일치한다[37].

한편, 일부 기능형 문항(2023년 3-3, 4-6, 4-8; 2024년

1-3, 3-5, 4-5)은 하위 성취 집단에서는 남학생에게, 상위 성취 집단에서는 여학생에게 유리하게 작용하는 비균일(+) DIF로 확인되었다. 이러한 양상은 여러 단계를 거쳐 절차적으로 문제를 해결해야 하는 기능형 문항에서 특히 두드러지게 나타났다.

저작형 문항은 2023년과 2024년 모두 여학생들에게 유리하게 나타났다. 이 문항은 소프트웨어 프로그램을 활용하여 배경 색상, 제목 글씨, 본문 글씨를 주어진 조건에 맞게 편집하는 수행 난이도가 높은 문항으로 정답을 위해 7개의 세부 조건을 모두 충족해야 하는 복잡한 구조를 가진다. 일반적으로 수행 난이도가 높은 문항은 대체로 남학생들에게 유리한 경향을 보였으나, 이처럼 세밀한 조건을 정확하게 수행해야 하는 유형의 문항에서는 여학생이 상대적으로 더 유리하게 기능하는 경향을 보였다.

또한 오픈 대화방에서 이름을 등록하고 친구를 추가하는 문항(4-2)은 SNS 어플리케이션의 활용 능력을 측정하는 수행 난이도가 높은 문항임에도 불구하고 여학생에게 유리하게 기능하였다. 이는 대화창 내 프로필 사진과 ‘(모르는 사용자)’로 지정된 프로필들의 대화를 분석하여 적절한 친구를 판단하는 과정에서, 단순한 버튼 클릭을 넘어 텍스트와 프로필 정보를 종합적으로 해석하는 능력이 요구되기 때문이다. 즉, 수행 난이도가 높으면서도 종합적인 해석을 요구하는 문항 유형에서는 여학생이 상대적으로 강점을 보이며, 이는 여러 선행 연구 결과와도 일치한다[38, 39].

2) 중학교

중학교 수행기반 문항에 대한 분석 결과, 2023년에는 기능형, 저작형, 알고리즘형 문항에서, 2024년에는 기능형, 저작형, 연결형, 알고리즘형 문항에서 DIF가 추출되었다(부록 Table 6 참고). 중학교 역시 디지털 도구를 조작하거나 활용하는 문항들은 남학생에게, 디지털 협업 및 창의적 콘텐츠를 제작하는 문항은 여학생들에게 유리하게 작용하였다. 예를 들어, 공유 드라이브 및 학교 홈페이지에 이미지를 업로드하는 문항(2023년: 2-8; 2024년: 2-5, 2-8), 무선 와이파이 연결 문항(2024년: 3-1)은 지문 해석보다 소프트웨어의 인터페이스 요소를 직관적으로 이해하고 활용하는 능력이 요구되는 문항으로, 남학생에게 더 유리한 경향을 보였다.

한편, 중학교 문항은 초등학교에 비해 전반적으로 수행 난이도가 높았으며, 언어적 이해력, 절차적 사고력, 세심한 실행력이 복합적으로 요구되는 문항이 다수 포함되었다. 이러한 문항에서는 하위 능력 집단에서는 남학생이, 상위 능력 집단에서는 여학생이 유리하게 반응하는 비균일(+) DIF로 분류되는 경향이 확인되었으며, 이와 같은 양상은 초등학교에서도 유사하게 나타났다. 특히, ‘D1. 창의적 사고를 활용한 디지털 콘텐츠 생산’ 요소를 측정하는 저작형 문항은 초등학교와 중학교 모두에서 여학생에게 유리한 DIF 문항으로 나타났다(부록 Figure 1 참조).

다음으로 알고리즘형 문항의 경우 초등학교에서는 DIF로 추출되지 않았으나, 중학교에서는 2023년, 2024년 모두 비균일 DIF 문항으로 확인되었다. 이는 알고리즘 문항이 요

구하는 문제해결 전략이 성취 수준에 따라 다르게 작용했을 가능성을 시사하며, 개념과 원리에 기반하여 문제해결 절차를 구성하는 사고에서 여학생이 높은 성취를 보인다는 기존 연구의 결과와도 일치한다[40, 41].

4.2.2 지식기반 문항(Knowledge-based items)

1) 초등학교

지식기반 문항(선다형 문항)에서는 초등학교와 중학교 모두 여학생에게 유리하게 작용하는 문항이 많은 것으로 확인되었다(부록 Table 7 참고).

문항 내용을 기준으로 살펴보면, 크게 디지털 건강과 윤리 관련 문항과 정보 검색 및 활용 관련 문항으로 구분할 수 있다. 선행연구에 따르면, 정보 출처의 신뢰성 평가, 표절 방지와 같은 윤리적 정보 활용 능력을 측정하는 문항에서 여학생이 상대적으로 우수한 수행을 보인다고 보고하였는데 [42, 43], 본 연구에서도 이와 유사한 양상이 관찰되었다.

한편, 2023년의 4-4번과 4-7번 문항, 2024년의 3-6번과 4-4번 문항은 비균일(+) DIF 문항으로 추출되었다. 해당 문항들은 일반적인 텍스트 기반의 객관식 응답 형식이 아니라, 스마트폰 환경을 모사한 UI 상호작용을 포함하거나, 실제 사례 기반의 판단 상황을 제시하여 정보 판단 및 해석 및 문제해결 과정을 복합적으로 요구하였다. 이러한 유형의 문항은 하위 성취 집단에서는 남학생에게, 상위 성취 집단에서는 여학생에게 유리하게 작용하여 비균일 DIF로 분류되었다.

2) 중학교

중학생 대상 지식기반 문항의 성별 DIF 분석 결과(부록 Table 8 참고), 디지털 건강과 윤리(2023년: 2-9; 2024년: 1-7, 2-1)와 정보 검색(2023년: 2-2) 문항은 여학생에게, 디지털 도구 활용과 이해를 묻는 문항은 남학생에게 유리하게 나타났다. 예컨대, 포트 식별(부록 Figure 2 참고), 공유 문서 및 설문조사 플랫폼 활용, 파일 공유 방식, 생성형 AI 개념 활용 문항에서는 남학생에게 유리하게 작용하였다. 이 중 2024년 2-7번 문항은 텍스트 기반 선다형 문항임에도 남학생에게 유리하게 기능하였는데, 이는 설문조사에서 나타난 남학생의 높은 생성형 AI 활용 효능감과 관련될 가능성이 있다. 이러한 경향은 디지털 도구의 이해·활용을 요구하는 문항에서 남학생이 유리하다는 앞선 분석 결과와 일관되며, 하위 성취 집단에서 더욱 뚜렷하게 나타났다.

한편, 상위 성취 집단으로 갈수록 여학생에게 유리한 방향으로 전환되는 비균일 DIF도 다수 확인되었다. 이 중 2023년 2-5번 문항은 디지털 윤리와 관련된 문항으로, 성취 수준에 따른 DIF 방향의 전환이 보다 뚜렷하게 나타났다. 해당 문항은 다양한 이미지 사용 조건을 바탕으로 올바른 출처 표기 방식을 판단하도록 구성되어 있어 단순한 저작권 개념 이해를 넘어 실제 적용 경험과 사례 해석 능력을 요구한다. 이는 일부 고성취 집단에서 규칙 추론과 정보 구조 해석 측면에서 남학생이 상대적으로 유리했을 가능성을 시사하며, 복잡한 문제 상황에서 남학생의 수행이 높게 나타난다

는 선행연구 결과와도 맥락을 같이한다[44].

아울러 자동차와 관련된 문항(2023년: 4-1)은 주제 친숙성과 시각 정보 처리 능력이 요구되는 특성으로 인해 남학생에게 유리한 비균일 DIF로 나타났다(부록 Figure 2 참고). 이는 남학생이 시각적 정보 구조 파악과 변수 간 관계 분석에서 우수하다는 선행연구와[44], 자동차라는 주제 친숙성이 남학생에게 유리하게 작용할 수 있다는 기존 연구 결과와도 맥락을 같이한다[45, 46].

5. 결론 및 제언

검사 결과는 교육적 의사결정과 정책 수립의 핵심 근거로 활용되기 때문에, 검사 도구는 응시자가 속한 집단에 관계없이 검사 목적에 부합하는 결과를 일관되게 산출할 수 있어야 하며, 이를 위해 집단 간 타당성, 특히 문항 수준에서의 측정 공정성에 대한 검증은 필수적이다[47, 48]. 차별기능문항 분석은 검사 도구의 공정성과 타당성을 점검하기 위한 중요한 절차에 해당한다. 실제로 본 연구에서 활용된 국가수준 디지털 리터러시 검사에서는 성별에 따른 디지털 리터러시 성취 격차가 여러 해에 걸쳐 반복적으로 확인되었고, 이러한 차이가 실제 능력 차이인지, 혹은 문항 특성에 기인한 것인지 검토할 필요성이 제기되었다. 따라서 본 연구에서는 2023년, 2024년 국가수준 초·중학생 디지털 리터러시 검사를 대상으로, 성별에 따른 차별기능문항의 존재 여부와 그 특성을 평가영역 및 문항 유형을 중심으로 분석하였다.

첫째, 성별에 따른 DIF 문항은 모든 평가영역에서 확인되었으며, 초등학교와 중학교 공통적으로 'C. 디지털 의사소통 및 협력', 'D. 디지털 자원 생산' 영역에서 비교적 높은 비율로 나타났다. 특히, 'D1. 창의적 사고를 활용한 디지털 콘텐츠 생산'은 모든 연도와 학교급에서 전 문항이 DIF로 추출되었으며, 중학교의 'B1. 정보·데이터 탐색 및 수집' 하위요소 또한 2023년과 2024년 모두 모든 문항에서 DIF가 나타났다. 그러나 DIF 문항은 특정 평가영역이나 하위요소, 문항 유형에 국한되지 않고, 평가영역 전반에 분포하였다. 이는 문항의 주제나 친숙도, 응답 방식, 수행 조건, UI, UX를 포함한 정보 제시 방식 등 문항 설계의 구조적 특성과 성별 간 상호작용에 따라 DIF가 발생했을 가능성을 시사한다. 특히 특정 문항 유형, 문항 수행 맥락, 표현 방식 등이 성별에 따라 응답자에게 인지적 부담이나 전략 선택의 차이를 유발했을 가능성이 있으므로, 향후 DIF 발생의 원인을 다면적으로 검토할 필요가 있다.

둘째, 문항의 주제와 수행 방식에 따라 성별 간 유불리가 다르게 나타나는 양상이 확인되었다. 예를 들어, 디지털 도구 조작이나 소프트웨어 기능 활용을 요구하는 기능형 문항은 남학생에게, 협업, 디지털 윤리, 정보 탐색과 같은 디지털 안전과 건강에 관한 주제를 다룬 선다형 문항은 여학생에게 유리하게 작용하는 경향이 확인되었다. 이러한 경향은 선행연구에서도 확인된 바 있다[42, 43]. 그러나 실제 사례나 조

건에 기반한 디지털 윤리 문항은 여학생들에게 불리하게 작용하였고, 포트 위치 식별이나 생성형 AI 활용 방안과 같이 디지털 도구의 개념적 이해를 묻는 문항은 선다형임에도 불구하고 남학생에게 더 유리하게 나타났다. 이는 동일 영역 내에서도 문항이 요구하는 인지적 부담이나 표현 방식에 따라 성별 간 유불리가 다르게 작용할 수 있음을 보여준다.

또한 저작형 문항은 수행 난이도가 높은 문항임에도, 복수의 세부 조건을 충실히 이행해야 하는 특성이 여학생의 성실한 수행 태도 및 언어적 능력과 맞물리면서 상대적으로 높은 수행 결과로 이어진 것으로 해석될 수 있다. 알고리즘형 문항은 중학교에서만 비균일(+) DIF로 추출되었다. 이는 선행연구에서 알고리즘 영역은 이론적 측면과 개념 및 원리에 따른 문제해결 능력을 요구하며, 해당 영역에서는 여학생의 성취가 높게 나타나는 반면 실제 프로그래밍 수행에서는 남학생의 성취가 상대적으로 높게 나타났다고 보고한 결과와 관련될 수 있다[41, 49]. 즉, 상위 성취 집단에서는 조건을 체계적으로 해석하고 절차화하는 과정이 중요하게 작용하여 여학생에게 유리하게 나타난 반면, 하위 성취 집단에서는 실제 수행능력이 상대적으로 더 중요하게 작용하여 남학생에게 유리하게 작용했을 가능성이 있다. 이러한 결과는 문항의 주제, 유형, 수행 방식에 따라 성별 간 문제해결 전략이 달라질 수 있음을 보여주며, 향후 DIF 해석에서는 문항 특성과 성별 특성 간 상호작용을 보다 정교하게 분석할 필요가 있음을 시사한다.

나아가 이러한 경향이 초등학교와 중학교 모두에서 공통적으로 확인되었다는 점은 교육 현장에서 성별 특성을 반영한 교수·학습 전략 수립에 중요한 함의를 제공한다. 구체적으로 남학생의 경우 협업, 디지털 윤리, 정보 탐색 및 평가 역량을 강화할 수 있도록 사례 기반 토의 활동이나 디지털 시민성 함양을 위한 교육 프로그램을 구성할 필요가 있다. 반면 여학생은 디지털 도구의 개념적 이해와 적용 능력을 향상시킬 수 있도록 실습 중심 수업이나 문제해결 중심의 프로젝트형 활동 등을 통해 실제 활용 능력을 강화할 수 있는 수업 설계가 요구된다.

셋째, 두 연도 모두에서 성별에 따른 DIF가 다수 확인되었다는 점은 문항 전반에 대한 체계적이고 다각적인 검토의 필요성을 제기한다. 예를 들어, 이문수(2018)는 ICILS 2013 분석에서 여학생에게 유리한 문항이 언어 능력과 관련되었을 가능성을 지적하였는데[6], 본 연구에서도 선다형뿐 아니라 기능형 문항에서도 언어적 이해와 추론 능력이 수행에 영향을 미쳤을 가능성이 관찰되었다. 이러한 맥락에서 인터페이스 요소의 위치 파악이나 메뉴 구조 탐색이 정답 도출의 핵심 요인이 되지 않도록 문항이 설계되었는지, 수행 과정에서 요구되는 클릭 단계나 절차의 수가 과도한 인지적 부담으로 작용하지는 않는지에 대한 검토가 요구된다. 아울러 동일한 문항 유형이라 하더라도 텍스트와 시각 정보를 동시에 처리해야 하는 구조가 성별에 따라 상이한 문제해결 전략을 유발하지 않는지 여부를 예비검사 단계에서 확인하는 과정이 필요하다. 따라서 향후 평가에서는 특정 인지 능력이나 배경 지식에 과도하게 의존하지 않도록 문항의 주제와 유

형뿐 아니라 과제 수행 구조, 정보 제시 방식, 서술의 복잡성 등을 보다 정교화하고, 잠재적 편향을 사전에 식별하여 검사의 형평성과 타당성을 확보할 필요가 있다.

다만 본 연구는 차별기능문항의 존재 여부와 그 양상을 탐색적으로 분석하는 데 목적을 두고 있어, 관찰된 DIF가 문항 설계 특성에 기인한 것인지, 혹은 성별 간 실제 디지털 리터러시 역량이나 학습 경험의 차이를 반영한 결과인지를 인과적으로 구분하는 데에는 한계가 있다.

이상의 한계점과 연구결과를 바탕으로 후속 연구를 제안 하면 다음과 같다. 첫째, 학년과 성취수준을 고려한 집단 세분화를 통해 차별기능분석을 실시할 필요가 있다. 현재 검사는 학교급 내에서 학년에 관계없이 동일한 문항을 제시하고 있으나, 학년에 따라 교육과정, 학습 경험, 디지털 활용 역량 등이 상이할 수 있고, 성취수준에 따라 문항 반응 양상도 달라질 가능성이 있다. 따라서 향후 분석에서는 학년과 성취수준을 함께 고려한 집단 세분화를 통해 문항 반응을 분석함으로써 평가 결과에 대한 해석을 보다 정교화할 필요가 있다.

둘째, 동일 문항이라도 문제해결 과정이나 전략이 학생별로 상이하게 나타날 수 있다. 이는 단순 정답률이나 통계적 수치만으로는 충분히 파악하기 어려우므로, 문항 내용 및 교육과정의 질적 분석과 더불어 로그 데이터 기반의 수행 과정 분석이 병행될 필요가 있다. 특히, 클릭 순서, 응답 시간, 기능 탐색 방식 등 로그 데이터 분석은 수행 결과 이면의 문제해결 과정을 정밀하게 파악할 수 있게 하며, 검사의 타당성과 공정성을 검증하는 실증적 근거로 활용될 수 있을 것이다.

셋째, 검사 도구 개발 과정에서 성별 편향을 완화하기 위해서는 문항 설계 가이드라인 마련을 뒷받침할 실증적 근거가 필요하다. 이를 위해 차별기능문항과 비차별기능문항의 설계 특성을 비교, 분석하여, 텍스트 양, 조작 단계의 수, 과제 수행의 복잡성이 어느 수준까지 DIF를 유발하지 않는지, 반대로 어느 수준 이상에서 측정하고자 하는 역량 이외의 능력 요인이 개입되어 DIF가 발생하는지 검토할 필요가 있다. 이는 성별에 따른 차별기능을 최소화할 수 있는 문항 설계 범위를 규명하고, 향후 문항 개발 시 합리적인 설계 판단 기준을 마련하는데 기여할 수 있을 것이다.

본 연구는 국가수준 디지털 리터러시 검사에서 성별에 따른 차별기능문항의 특성을 평가영역, 문항 유형, 문항 내용 차원에서 분석하고, 이를 초등학교와 중학교를 대상으로 그 양상을 비교하였다. 이를 통해 성별 편향 가능성을 조명하고, 학교급에 따른 성별 DIF 양상의 공통점과 차이를 파악함으로써, 이에 대한 교육적 지원 방향과 평가 문항 개선을 위한 기초 자료를 제시하였다. 디지털 리터러시가 미래 핵심 역량으로 강조되는 오늘날, 본 연구는 공정하고 타당한 평가 체제 구축 및 도구 개발을 위한 실증적 근거를 제공한다는 점에서 의의를 가진다. 또한 본 연구 결과는 디지털 리터러시 교육 정책 수립과 교육의 질 향상을 위한 기초 자료로 활용될 수 있을 것으로 기대된다.

참고문헌

- [1] Ministry of Education. (2021). *Major points of the 2022 revised national curriculum general guidelines (draft)*. <https://www.moe.go.kr/boardCnts/viewRenew.do?boardID=294&boardSeq=89671&dev=0>
- [2] Kim, H., Lee, H., Yoo, S., Park, J., Lim, Y., Seo, J., & Gong, H. (2024). *A study on measuring students' digital literacy levels and strategies for advancement in 2024*. Korea Education and Research Information Service, KR 2024-04. <https://www.keris.or.kr/main/ad/pblcte/selectPblcteRRInfo.do?mi=1138&pblcteSeq=13834>
- [3] OECD. (2023). *PISA 2025 learning in the digital world framework (second draft)*.
- [4] UNESCO. (2018). *A Global Framework of Reference on Digital Literacy Skills for Indicator 4.4.2*. UNESCO Institute for Statistics, Information Paper No. 51.
- [5] Riina, V., Stefano, K., & Yves, P. (2022). *DigComp 2.2: The Digital Competence Framework for Citizens-With new examples of knowledge, skills and attitudes (No. JRC128415)*. Joint Research Centre. <https://publications.jrc.ec.europa.eu/repository/handle/JRC128415>
- [6] Lee, M. (2018). Differential Item Functioning in the aspects of Country and Gender Differences in the Computer and Information Literacy Test: Based on ICILS 2013. *Journal of Educational Evaluation*, 31(3), 653-679. <https://doi.org/10.31158/JEEV.2018.31.3.653>
- [7] Cho, Y., Lee, E., Kim, M., & Lee, J. (2024). *Restructuring the Gyeonggi-style digital competency framework and strategies for curriculum application*. Gyeonggi Institute of Education, 2024-04. <https://www.gie.re.kr/publication/stdreportDetail.do?id=141091125&currRow=1>
- [8] Kim, S., Lee, H., Kim, H., Park, J., Choi, M., & Han, N. (2023). *A study on the improvement of performance-based digital literacy assessment tools and level measurement for elementary and middle school students in 2023*. Korea Education and Research Information Service, KR 2023-13. <https://www.keris.or.kr/main/ad/pblcte/selectPblcteRRInfo.do?mi=1138&pblcteSeq=13751>
- [9] Kim, C., Lee, M., & Ahn, S. (2016). The Analysis of Gender Differential Item Functioning on the Information and Communications Technology Literacy Tests, 2015. *Journal of Educational Evaluation*, 29(2), 301-324.
- [10] Lee, H., Lee, W., Cha, H., Kim, S., Na, W., Kye, B., & Han, N. (2022). *A study on measuring the national-level digital literacy of elementary and middle school students in 2022*. Korea Education and Research Information Service, KR 2022-2. <https://www.keris.or.kr/main/ad/pblcte/selectPblcteRRInfo.do?mi=1138&pblcteSeq=13645>
- [11] Gebhardt, E., Thomson, S., Ainley, J., & Hillman, K. (2019). *Gender differences in computer and information literacy: An in-depth analysis of data from ICILS*. Springer Nature. <https://doi.org/10.1007/978-3-030-26203-7>
- [12] Geske, A., & Ozola, A. (2010). Differential item functioning in the aspect of gender differences in reading literacy. *In The 4th IEA International Research Conference*.

- [13] Harputlu, L., & Güray, B. (2024). Gender differences in digital literacy: An analysis. *Bati Anadolu Eğitim Bilimleri Dergisi*, 15(3), 3286-3300. <https://doi.org/10.51460/baebd.1562102>
- [14] Jeyshankar, R., & Nachiappan, N. (2021). Study on gender differences in information literacy skills among research scholars of Alagappa University, Karaikudi, Tamil Nadu. *Library Philosophy and Practice*, 1-28.
- [15] Gilster, P. (1997). Digital literacy. Wiley Computer.
- [16] Mills, K. A. (2010). A review of the “digital turn” in the new literacy studies. *Review of Educational Research*, 80(2), 246-271. <https://doi.org/10.3102/0034654310364401>
- [17] JISC. (2015). Building digital capabilities: The six elements defined.
- [18] Jung, J., & Shin, Y. (2022). Reconceptualizing digital literacy after COVID-19: A focus on elementary education. *The Korean Journal of Literacy Research*, 13(2), 75-106.
- [19] Baek, S., Lim, C., Kim, H., Yoo, Y., & Kim, M. (2009). *Development of a performance-based ICT literacy assessment tool*. Korea Education and Research Information Service, KR 2009-4. <https://www.keris.or.kr/main/ad/pblcte/selectPblcteRRInfo.do?mi=1138&pblcteSeq=11851>
- [20] Shin, S., & Lee, S. (2019). A Study on Development and Validity Verification of a Measurement Tool for Digital Literacy for University Students. *The Journal of Learner-Centered Curriculum and Instruction*, 19(7), 749-768. <https://doi.org/10.22251/jlcci.2019.19.7.749>
- [21] Chung, M., Kim, J., & Hwang, H. (2021). A study on development and validation of digital literacy measurement tool. *Journal of Internet Computing and Services*, 22(4), 51-63. <https://doi.org/10.7472/jksii.2021.22.4.51>
- [22] Care, E., Griffin, P., & McGaw, B. (2012). *Assessment and teaching of 21st century skills*. Springer. <https://doi.org/10.1007/978-94-017-9395-7>
- [23] Dorans, N. J., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 35-66). Lawrence Erlbaum Associates.
- [24] Seong, T., & Song, M. (1999). A distinction between uniform DIF and parallel DIF among unidirectional DIFs. *Journal of Educational Evaluation*, 12(1), 1-16.
- [25] Yun, S., & Lee, S. (2013). Detection of differential item functioning using structural equation modeling: A comparison of MACS and MIMIC. *Korean Journal of Psychology: General*, 32(4), 1023-1052.
- [26] Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27(4), 361-370. <https://doi.org/10.1111/j.1745-3984.1990.tb00754.x>
- [27] Zumbo, B. D., & Thomas, D. R. (1997). *A measure of effect size for a model-based approach for studying DIF*. University of Northern British Columbia, Edgeworth Laboratory for Quantitative Behavioral Science.
- [28] Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Lawrence Erlbaum Associates.
- [29] Lee, S., & Suh, Y. (2018). Lord's Wald test for detecting DIF in multidimensional IRT models A comparison of two estimation approaches. *Journal of Educational Measurement*, 55(2), 328-353. <https://doi.org/10.1111/jedm.12178>
- [30] Thissen, D., Steinberg, L., & Wainer, H. (1988). *Use of item response theory in the study of group difference in trace lines*. In Test validity. Routledge.
- [31] Seo, K., Ahn, H., & Choi, Y. (2024). Analysis of differential item functioning according to digital device familiarity in computer-based tests: Focusing on eTIMSS 2019 fourth-grade mathematics. *The Journal of Curriculum and Evaluation*, 27(1), 209-243. <https://doi.org/10.29221/jce.2024.27.1.209>
- [32] Kang, T. (2015). Effects of sample sizes and unequal examinee ratios on the type I error rates of DIF detection. *Journal of Educational Evaluation*, 28(2), 577-600.
- [33] No, U., Kim, J., & Kim, S. (2010). Detecting differential item functioning by gender on the problem-solving literacy of PISA 2003. *The Korean Journal of Educational Methodology Studies*, 22(4), 165-194. <https://doi.org/10.17927/tkjems.2010.22.4.165>
- [34] DeMars, C. E. (2009). Modification of the Mantel-Haenszel and Logistic Regression DIF Procedures to Incorporate the SIBTEST Regression Correction. *Journal of Educational and Behavioral Statistics*, 34(2), 149-170. <https://doi.org/10.3102/1076998607313923>
- [35] Magis, D., Beland, S., Tuerlinckx, F., & De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods*, 42, 847-862. <https://doi.org/10.3758/BRM.42.3.847>
- [36] Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1-29. <https://doi.org/10.18637/jss.v048.i06>
- [37] Hargittai, E. (2010). Digital natives? Variation in internet skills and uses among members of the “net generation”. *Sociological Inquiry*, 80(1), 92-113. <https://doi.org/10.1111/j.1475-682X.2009.00317.x>
- [38] Kwon, S. (2018). Exploring differential item functioning of TIMSS 2015 mathematics items by gender, based on cognitive diagnosis models. *Journal of Educational Studies*, 49(1), 61-82. <https://doi.org/10.15854/jes.2018.3.49.1.61>
- [39] Martiniello, M. (2009). Linguistic complexity, schematic representations, and differential item functioning for english language learners in math tests. *Educational Assessment*, 14(3), 160-179. <https://doi.org/10.1080/10627190903422906>
- [40] Yu, B., Kim, J., & Lee, W. (2012). Analysis on the relation between programming achievement and problem solving according to gender. *The Journal of the Korean*

Association of Computer Education, 15(6), 1–10. <https://doi.org/10.32431/kace.2012.15.6.001>

- [41] Yang, H., Kim, J., & Lee, W. (2020). Analysis of algorithm and programming level according to gender of high school students. *The Journal of Learner-Centered Curriculum and Instruction*, 20(2), 185–202. <https://doi.org/10.22251/jlcci.2020.20.2.185>
- [42] Jones, S. L., & Procter, R. (2023). Young peoples' perceptions of digital, media and information literacies across Europe: gender differences, and the gaps between attitudes and abilities. *Technology, Pedagogy and Education*, 32(4), 435–456. <https://doi.org/10.1080/1475939X.2023.2210152>
- [43] Campos, D. G., & Scherer, R. (2024). Digital gender gaps in Students' knowledge, attitudes and skills an integrative data analysis across 32 Countries. *Education and Information Technologies*, 29(1), 655–693. <https://doi.org/10.1007/s10639-023-12272-9>
- [44] Funke, J. (2014). Analysis of minimal complex systems and complex problem solving require different forms of causal cognition. *Frontiers in psychology*, 5, 739. <https://doi.org/10.3389/fpsyg.2014.00739>
- [45] Stoet, G., & Geary, D. C. (2015). Sex differences in academic achievement are not related to political, economic, or social equality. *Intelligence*, 48, 137–151. <https://doi.org/10.1016/j.intell.2014.11.006>
- [46] Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and things, women and people: A meta-analysis of sex differences in interests. *Psychological Bulletin*, 135(6), 859–884. <https://doi.org/10.1037/a0017364>
- [47] Chin, S., & Seong, T. (2004). An exploratory study on item-format related DIF with MH and SIBTEST techniques. *Journal of Educational Evaluation*, 17(2), 215–236.
- [48] Yoon, J., & Lee, Y. (2013). A study of differential item functioning (DIF) analyses using the TIMSS (2007) mathematics test across South Korea, the United States, and Singapore. *Journal of Educational Evaluation*, 26(2), 415–439.
- [49] Lau, W. W., & Yuen, A. H. (2010). Promoting conceptual change of learning sorting algorithm through the diagnosis of mental models: The effects of gender and learning styles. *Computers & Education*, 54(1), 275–288. <https://doi.org/10.1016/j.compedu.2009.08.014>



노혜림

· 2015년 신라대학교 교육학과 (학사)
· 2020년~현재 건국대학교 교육학과 심리측정통계 전공 석박사통합과정 수료

✦ 관심분야 : 맞춤형 교육, 검사의 타당성, 측정평가 등
✉ ahdcl5@naver.com



서정희

· 1993년 서울대학교 생물교육과 (이학사)
· 1995년 서울대학교 대학원 과학교육과 (교육학석사)
· 2002년 서울대학교 대학원 과학교육과 (교육학박사)
· 1999년~현재 한국교육학술정보원 연구위원

✦ 관심분야 : 디지털 교육, 교원·학생 디지털 역량 모델 및 측정
✉ jhseo@keris.or.kr



이현숙

· 1996년 서울대학교 지구과학교육과 (이학사)
· 2000년 서울대학교 대학원 과학교육과 (이학석사)
· 2005년 University of Iowa Educational Measurement & Statistics 전공 (Ph.D.)
· 2008년~현재 건국대학교 교육학과 교수

✦ 관심분야 : 학습분석학, 지식추적, 측정평가, AI in Education 등
✉ hyunsi@konkuk.ac.kr

부록

Table 1. Performance-Based Digital Literacy Assessment Scenarios and Tasks (2023–2024)

School Level	No.	Module	Content	Task Context (Task Type)	Number of Items
Elementary School	1	Using school public PC safely	Performing a problem-solving task to use school shared PCs safely	Individual (Production task)	4
	2	Food serving robots	Performing a problem-solving task to design a robot that delivers food	Individual (Production task)	6
	3	School safety and accident prevention pictogram	Performing a content-creation task to prevent school safety accidents	Individual (Algorithm design)	6
	4	Experiential learning report	Completing a report-writing task from planning to final production through team collaboration	Collaborative (Production task)	10
Middle School	1	Introducing the job, "YouTube Creator"	Performing a data-creation task to introduce YouTube creator career paths through online team collaboration	Collaborative (Production task)	7
	2	Making a public relations poster for a student council election candidate	Performing a task to create a promotional poster for a student council election	Individual (Production task)	9
	3	"Our Future School" Contest	Performing a task to create a promotional plan and SNS content for "Our Future School"	Collaborative (Production task)	6
	4	Self-driving cars	Performing a task to program an autonomous driving system	Individual (Algorithm design)	4

Table 2. Number and Percentage of Test Items by Digital Literacy Assessment Domains (2023–2024)

Domain	Subdomain	Elementary School				Middle School			
		2023		2024		2023		2024	
		Number of Items	Percentage(%)	Number of Items	Percentage(%)	Number of Items	Percentage(%)	Number of Items	Percentage(%)
A. Digital Tools	A1. Understanding digital tools	2	7.69	2	7.69	2	7.69	1	3.85
	A2. Using digital tools	2	7.69	2	7.69	4	15.38	5	19.23
B. Digital Information and Data	B1. Searching for and collecting information and data	2	7.69	2	7.69	2	7.69	2	7.69
	B2. Analyzing and evaluating information and data	2	7.69	2	7.69	2	7.69	2	7.69
	B3. Managing information and data	2	7.69	2	7.69	2	7.69	2	7.69
C. Digital Communication and Collaboration	C1. Digital communication	2	7.69	2	7.69	2	7.69	2	7.69
	C2. Digital collaboration	2	7.69	2	7.69	2	7.69	2	7.69
D. Digital Resource Production	D1. Creating digital resources using creative thinking	2	7.69	2	7.69	2	7.69	2	7.69
	D2. Producing programs using computational thinking	2	7.69	2	7.69	3	11.54	3	11.54
E. Digital safety and well-being	E1. Digital safety	5	19.23	5	19.23	2	7.69	2	7.69
	E2. Digital well-being	3	11.54	3	11.54	3	11.54	3	11.54
	Total	26	100.00	26	100.00	26	100.00	26	100.00

Table 3. Number and Percentage of Test Items by Item Type in the Digital Literacy Assessment (2023–2024)

Item Type		Elementary School				Middle School			
		2023		2024		2023		2024	
		Number of Items	Percentage(%)	Number of Items	Percentage(%)	Number of Items	Percentage(%)	Number of Items	Percentage(%)
Knowledge-based items	Multiple-choice item (MC)	8	30.77	8	30.77	9	34.62	9	34.62
	Functional task (F)	15	57.69	15	57.69	10	38.46	10	38.46
Performance-based items	Algorithmic item (A)	1	3.85	1	3.85	3	11.54	3	11.54
	Matching item (M)	–	–	–	–	1	3.85	1	3.85
	Production item (P)	2	7.69	2	7.69	3	11.54	3	11.54
	Total	26	100.00	26	100.00	26	100.00	26	100.00

Table 4. Descriptive Statistics of Digital Literacy Assessment Domains by Gender (2023–2024)

School Level	Year	Group	A. Digital Tools		B. Digital Information and Data		C. Digital Communication and Collaboration		D. Digital Resource Production		E. Digital safety and well-being	
			M	SD	M	SD	M	SD	M	SD	M	SD
Elementary School	2023	Total	71.78	28.60	62.57	26.95	65.92	30.83	43.50	32.68	71.33	23.90
		M	69.73	29.33	60.07	27.21	61.76	31.53	41.20	33.13	68.59	24.65
		F	73.95	27.64	65.21	26.42	70.31	29.44	45.93	32.04	74.23	22.74
		t	9.15***		11.84***		17.32***		8.98***		14.73***	
	2024	Total	76.30	29.28	61.49	32.58	72.58	28.65	41.76	31.58	66.61	25.32
		M	74.31	30.64	58.04	33.11	69.23	29.47	39.50	31.42	64.02	25.99
		F	78.43	27.59	65.17	31.60	76.15	27.31	44.17	31.57	69.37	24.30
		t	9.71***		15.11***		16.74***		10.17***		14.59***	
Middle School	2023	Total	60.03	24.29	66.98	27.24	51.24	25.36	29.34	27.16	69.08	29.72
		M	57.00	25.59	61.93	28.28	48.62	25.61	26.20	27.27	63.13	31.66
		F	63.10	22.49	72.10	25.13	53.90	24.82	32.52	26.67	75.12	26.26
		t	16.77***		25.17***		13.86***		15.51***		27.31***	
	2024	Total	65.03	27.21	53.19	27.94	47.14	28.62	44.26	31.61	75.53	29.94
		M	60.15	28.95	47.80	29.19	43.12	29.11	39.30	31.92	69.09	33.03
		F	69.86	24.44	58.52	25.56	51.12	27.56	49.16	30.52	81.89	24.95
		t	25.04***		27.01***		19.52***		21.84***		30.23***	

*** $P < .001$

※ Scores for each assessment domain were calculated based on a 100-point scale.

Table 5. Results of DIF Analysis by Performance-Based Item Type (Elementary School)

Item No.	Domain Main	Sub	Content	Type	DIF Direction
1) 2023					
2-2	C	C1	Taking and sharing photos of food-serving robots	F	F
2-4	D	D2	Completing a robot movement path using an algorithm	F	M
3-3	B	B1	Searching for appropriate safety sign images online	F	(+)
3-4	B	D1	Creating safety signs using image editing tools	P	F
4-1	A	A2	Joining a chat room by scanning a QR code	F	M
4-2	C	C2	Adding friends and registering a user name in an open chat room	F	F
4-6	B	B3	Saving files for an experiential learning report	F	(+)
4-8	E	E1	Identifying and copying an image's original URL	F	(+)
4-9	D	D1	Preparing a report using a specified formatting template	EC	F
1) 2024					
1-3	E	E1	Deleting stored account information from a shared computer	F	(+)
3-1	A	A1	Connecting to a wireless internet network	F	M
3-4	D	D1	Searching for appropriate images of safety signs online	P	F
3-5	B	B3	Downloading and saving files	F	(+)

Table 6. Results of DIF Analysis by Performance-Based Item Type (Middle School)

Item No.	Domain Main	Sub	Content	Type	DIF Direction
1) 2023					
1-1	C	C2	Selecting appropriate software for real-time discussion with team members in a team project	F	(+)
1-4	B	B1	Applying filters on a search engine based on specified conditions	F	M
1-6	D	D1	Completing an online collaborative document according to a given format	P	F
2-6	A	A2	Creating a folder in a shared drive and uploading images	F	(-)
2-7	D	D1	Designing a promotional poster following a specified format	P	F
2-8	C	C1	Uploading images from local storage to a school website bulletin board	F	M
3-4	C	C2	Providing feedback on survey results using document comments	F	F
3-5	B	B3	Preparing a production plan and diagrams using a specified template	P	F
3-6	E	E2	Copying a collaborative document link to request survey participation	F	(+)
4-2	D	D2	Coding a path that satisfies given movement conditions	A	(+)
1) 2024					
1-1	C	C2	Adding a section and assigning editing permissions in a collaborative document	F	(+)
1-2	A	A2	Adding new ideas to one's section in a collaborative document	F	(+)

Item No.	Domain		Content	Type	DIF Direction
	Main	Sub			
4-1	A	A2	Joining a chat room by scanning a QR code	F	M
4-2	C	C2	Adding friends and registering a user name in an open chat room	F	F
4-5	C	C1	Saving files for an experiential learning report	F	(+)
4-8	E	E1	Identifying and copying an image's original URL	F	M
4-9	D	D1	Preparing a report using a specified formatting template	P	F
4-10	C	C2	Sharing a completed report in a chat room	P	(-)

Table 7. Results of DIF Analysis by Knowledge-Based Item Type (Elementary School)

Item No.	Domain		Content	Type	DIF Direction
	Main	Sub			
1) 2023					
1-4	E	E2	Items related to the prevention of internet and gaming overuse	MC	F
4-4	B	B2	Analyzing search results to select the most appropriate information	MC	(+)
4-7	E	E1	Items related to copyright protection and source attribution	MC	
1) 2024					
1-4	E	E2	Items related to the prevention of internet and gaming overuse	MC	F
2-3	B	B2	Analyzing movement data of a food-serving robot to select an optimal path	MC	F
3-6	E	E2	Selecting an appropriate method for sharing a completed safety sign	MC	(+)
4-4	B	B2	Analyzing search results to select the most appropriate information	MC	(+)

Item No.	Domain		Content	Type	DIF Direction
	Main	Sub			
1-4	B	B1	Applying filters on a search engine based on specified conditions	F	(+)
1-5	E	E1	Selecting an appropriate explanation for citing a news source	F	F
1-6	D	D1	Completing an online collaborative document according to a given format	P	F
2-5	A	A2	Creating a folder in a shared drive and uploading images	F	M
2-6	D	D1	Designing a promotional poster following a specified format	P	F
2-8	C	C1	Uploading images from local storage to a school website bulletin board	F	M
3-1	A	A1	Connecting to a wireless internet network	F	M
3-2	A	A2	Organizing tasks required for a future school design competition using a mind map	M	F
3-5	B	B3	Preparing a production plan and diagrams using a specified template	F	F
4-2	D	D2	Coding a path that satisfies given movement conditions	A	(+)
4-3	D	D2	Coding an optimal path that satisfies movement conditions using loops	A	(+)
4-4	D	D2	Coding an optimal path that outputs a message upon reaching a table using loops	A	(+)

Table 8. Results of DIF Analysis by Knowledge-Based Item Type (Middle School)

Item No.	Domain		Content	Type	DIF Direction
	Main	Sub			
1) 2023					
1-3	C	C1	Selecting an appropriate method for assigning team roles using a shared document	MC	(+)
2-1	A	A1	Connecting a USB device to a computer	MC	M
2-2	B	B2	Evaluating the reliability of web-based test results	MC	F
2-3	A	A2	Safely sharing files with specific recipients	MC	(+)
2-5	E	E1	Using online images in compliance with copyright regulations	MC	(-)
2-9	E	E2	Identifying inappropriate behaviors in digital-tool-based promotional activities	MC	F
3-3	B	B1	Selecting an appropriate survey item format	MC	(+)
4-1	B	B2	Analyzing search results to select accurate information	MC	(+)
1) 2024					
1-3	C	C1	Selecting an appropriate method for assigning team roles using a shared document	MC	(+)
1-7	E	E2	Selecting privacy protection principles when posting on social media	MC	F
2-1	B	B2	Evaluating the reliability of web-based test results	MC	F
2-2	A	A2	Safely sharing files with specific recipients	MC	(+)
2-7	A	A2	Identifying copyright and ethical issues in poster creation using generative AI	MC	M
3-3	B	B1	Selecting an appropriate survey item format	MC	M

2-7 동훈이는 친구들과 의논하며 온라인 협업 문서로 학생자치회장 선거 포스터를 최종 제작하기로 하였습니다. 온라인 협업 문서를 실행한 후, 다음 선거 포스터 양식에 따라 학생자치회장 선거 홍보용 포스터를 만들어보세요. (다음을 클릭하세요.)

3-6 설문조사 결과, 교사와 학생은 대부분 참여하였으나 학부모의 참여가 저조한 상황입니다. 그래서 학부모 밴드에 설문조사 참여 요청을 하려고 합니다. 아래 설문지를 보내기 위한 링크를 복사하세요. (해당하는 버튼을 클릭하세요.)

Figure 1. Middle school items from the 2023 assessment: 2-7 (EC: uniform(F)), 3-6(F: non-uniform(+))

2-1 선생님은 동훈이가 포스터 제작에 참고할 수 있도록 작년 학생자치회장 선거에 출마한 선배의 홍보 포스터 파일을 USB 저장 장치에 저장해서 동훈이에게 주셨습니다. 이 파일을 노트북에서 확인하려면 USB 저장 장치를 어디에 꽂아야 할까요. (다음을 클릭하세요.)

4-1 자연이네 모뎀은 자율주행 자동차에 대해 다음과 같이 검색 결과를 정리하였다. 검색 결과를 보고 올바른 정보를 고르세요. (다음을 클릭하세요.)

- ① 카메라는 날씨 등 외부 환경에 강하다.
- ② 레이더는 렌즈를 통한 시각 정보로 거리, 속도 측정을 한다.
- ③ 레이더는 직선 거리만 측정이 가능하다.
- ④ 라이다는 센서 중 가장 낮은 해상도를 보인다.
- ⑤ 라이다는 가격이 매우 저렴하다.

Figure 2. Middle school items from the 2023 assessment: 2-1(MC: uniform(M)), 4-1(MC: non-uniform(+))