

컴퓨터 기반 평가와 지필평가 간 학생 응답 특성 탐색 —컴퓨터 기반 국가수준 학업성취도 평가 병행 시행 결과를 중심으로—

백종호, 이재봉*, 구자욱
한국교육과정평가원

Exploring Differences of Student Response Characteristics between Computer-Based and Paper-Based Tests: Based on the Results of Computer-Based NAEA and Paper-Based NAEA

Jongho Baek, Jaebong Lee*, Jaok Ku
Korea Institute for Curriculum and Evaluation

ARTICLE INFO

Article history:

Received 26 December 2022

Received in revised form

31 January 2023

Accepted 16 February 2023

Keywords:

National Assessment of Educational Achievement (NAEA), Middle school students, Student response characteristics, Computer-based test, Mode effect

ABSTRACT

In line with the entry into the digital-based intelligent information society, the science curriculum emphasizes the cultivation of scientific competencies, and computer-based test (CBT) is drawing attention for assessment of competencies. CBT has advantages to develop items that have high fidelity, and to establish a feedback system by accumulating results into the database. However, it is necessary to solve the problems of improving validity of assessment results, lowering measurement efficiency, and increasing management factors. To examine students' responses to the introduction of the new assessment tools in the process of transitioning from paper-based test (PBT) to CBT, in this study, we analyzed the results of the PBT and the CBT conducted in 2021 National Assessment of Educational Achievement (NAEA). In particular, we sought to find the effects on student achievement when only the mode of assessment was changed without change of items, and the effect on student achievement when the items were composed including technology enhanced features that take advantage of CBT. This study is derived through the analysis of the results of 7,137 third-grade middle school students taking one among the three kinds of assessments, which were the PBT or two kinds of CBT. After the assessment, the percentage of correct answers and the item discriminations were collected for each group, and expert opinions on characteristics of response were collected through the expert council involving 8 science teachers with experience in NAEA. According to the results, there was no significant difference between students' achievement results in the PBT and the CBT-M, which means simple mode conversion type of CBT, so it could be explained that the mode effect did not appear. However, it was confirmed that the percentage of correct answers for the construct response items was somewhat high in the CBT, and this result was analyzed to be related to the convenience of the response. On the other hand, there were the items with a difference of more than 10%p from the correct answer rate of similar items, among the items to which technology enhanced functions were applied following the introduction of CBT. According to the analysis of response rate of options, these results could be explained that the students' level of understanding could be more closely grasped through the innovative items developed through the technology enhanced function. Based on the results, we discussed some guidance to be considered when introducing CBT and developing items through CBT, and presented implications.

1. 서론

디지털 기반 지능 정보 사회로 대변되는 4차 산업혁명 시대로의 진입은 교육 분야에도 큰 영향을 끼치고 변화를 요구하고 있다. 특히, 학생들이 학습을 통해 각 교과 지식에 대한 학습을 넘어 다양한 역량을 함양할 것을 요구하고 있으며(McMillan, 2014), 이러한 변화에 발맞추어 우리나라에서도 2015 개정 교육과정부터 총론에서는 핵심역량을, 과학 교육과정에서는 과학 교과역량 함양을 강조하기 시작했다(MOE, 2015). 역량을 강조하는 방향으로의 변화는 교수학습 영

역뿐 아니라 평가 영역에도 영향을 미쳐 어떻게 역량을 측정할 것인지에 대하여 고민하도록 요구하고 있다. 이에 따라 역량을 평가하는데 기존의 지필평가가 지녔던 한계를 극복하기 위한 방안으로 지필평가를 컴퓨터 기반 평가로 전환하려는 학계와 사회적 요구가 커졌다. 이는 역량을 측정하는데 컴퓨터 기반 평가(Computer-Based Test: 이하 CBT)가 전통적인 지필평가(Paper-Based Test: 이하 PBT)보다 더 적절하고 효과적이라는 점은 이미 많은 선행연구를 통해 보고되고 있기 때문이다(Kozma, 2009; Ripley, 2009; Sireci & Zenisky, 2006; Strain-Seymour, Way, & Dolan, 2009).

* 교신저자 : 이재봉 (jblee@kice.re.kr)

본 연구는 컴퓨터 기반 학업성취도 평가 전환을 위한 연구보고서인 '컴퓨터 기반 국가수준 학업성취도 평가(eNAEA) 도입을 위한 출제 방안 연구(RRE 2020-5)', '컴퓨터 기반 국가수준 학업성취도 평가(eNAEA)의 검사 도구 정교화 연구(RRE 2021-2)', '맞춤형 학업성취도 자율평가 지원 시스템(i성취)을 위한 컴퓨터 기반 출제 및 시행 체제 구축(RRE 2022-2)'의 내용을 기반으로 재구성 한 것임.

<http://dx.doi.org/10.14697/jkase.2023.43.1.17>

다양한 선행연구들은 컴퓨터 기반 평가에서는 기존의 지필평가에서 측정이 어려웠던 특성들을 평가할 수 있음을 강조하는데, 문항의 맥락 측면에서 우리 주변의 삶에서 발생하는 문제를 현실적으로 재현하는 데 효과적이며, 5지 선다형 방식에서 어려웠던 고차원적 사고 능력을 측정하는 데 도움이 된다고 보고하고 있다(Sireci & Zenisky, 2006; de Klerk, 2012; Strain-Seymour, Way, & Dolan, 2009; Bryant, 2017). 특히, 시뮬레이션과 같이 상호작용이 일어나는 과정을 문항에 구현하여 문제 해결 과정을 보다 정확하게 확인할 수 있다는 이점은 역량 측정에 더욱 도움을 준다고 볼 수 있다(Strain-Seymour, Way, & Dolan, 2009). 아울러 평가의 결과가 자동적으로 데이터베이스로 축적될 수 있다는 점은 평가의 환류 체계를 구축함에 있어서도 이점이 있다고 볼 수 있다. Lee et al.(2020)에서는 대규모 평가의 컴퓨터 기반 평가 전환에 따른 효과를 논의한 선행연구를 종합하여, 컴퓨터 기반 평가 전환의 기회 요소와 위험 요소를 정리하였는데, 혁신적 문항 도입에 따른 평가 구인 개선, 학생의 평가 참여 의욕 및 편의 개선, 시스템 전환화에 따른 평가 시행 효율성 개선을 기회 요소로, 평가 타당도 개선 입증 문제, 측정 효율성 저하 가능성, 시행 비용 상승 및 관리 요소 증가를 위험 요소로 제시한 바 있다.

이러한 맥락에서 평가 체계를 지필평가에서 컴퓨터 기반 평가로 전환시키는 것은 전세계적인 추세이다. 예를 들어, 국제 학업성취도 평가인 PISA와 TIMSS 같은 경우에는 컴퓨터 기반 평가 체제의 도입을 과거부터 준비해왔다. PISA의 경우에는 2006년에 과학 영역을 시작으로 3개 주기에 걸쳐 컴퓨터 기반 평가 도입을 위한 준비를 점진적으로 진행해 왔고, 2015년부터는 전 영역에 걸쳐 컴퓨터 기반 평가를 도입하였다(Song et al., 2014). TIMSS에서도 2019년부터 디지털 기반 평가인 eTIMSS로 명명한 컴퓨터 기반 평가를 도입, 전환하기 시작하였으며, TIMSS 2023에서 전면적인 도입을 예고하고 있다(Cotter, Centurino, & Mullis, 2020). 이외에도 미국의 대규모 학업성취도 평가인 NAEP, 호주의 학업성취도 평가인 NAPLAN에서도 컴퓨터 기반 평가가 도입되는 변화를 가져왔다(ACARA, 2018, Jewsbury et al., 2020), 이러한 국외의 대규모 평가들은 모두 전면적인 도입에 앞서 안정적인 전환을 위해 장기간에 걸쳐 평가나 결과 보고 방식, 절차 등을 준비하고, 시범 적용을 실시하였다.

한편, 우리나라에서도 컴퓨터 기반 평가에 대한 관심이 높아지고 있으며, 해외의 대규모 평가와 같이 컴퓨터 기반 평가 도입을 준비하고 있다. 예를 들어 기초학력 도달을 진단하고 결과를 토대로 교육적 환류를 실행할 수 있도록 마련한 Ban et al.(2021, 2022)에서는 학생 맞춤형 정보 제공을 위해 컴퓨터 기반 평가의 도입이 효과적임을 강조하였다. 이와 같은 접근은 학생들의 개별적인 평가 결과가 데이터베이스로 구축 가능하다는 장점에 기반하고 있으며, 역량 평가의 가능성과 더불어 컴퓨터 기반 평가의 특징적인 강점이라고 볼 수 있다. 한편, 과학과에서는 상대적으로 학교 현장에서 실행 가능하도록 소규모 집단을 대상으로 하는 컴퓨터 기반 평가에 대한 연구는 오래전부터 이루어졌다. 예를 들어, 전기 회로 연결에 대한 수행 평가 도구를 제안하였거나(Choi, 2013a), 문제해결력 측정을 위한 평가 도구를 개발하기도 하였다(Choi, 2013b).

이러한 흐름 속에서 한국교육과정평가원에서도 2013년 이후로 컴퓨터 기반 평가의 도입을 준비해왔으며(Kim et al., 2013), 최근에는 국가수준 학업성취도 평가(이하 학업성취도 평가)의 컴퓨터 기반 평

가 전환이 이루어지고 있다(Lee et al., 2020; 2021). 학업성취도 평가는 교육과정의 질 관리를 위한 점검의 목적에서 이루어지는 국가 단위의 대규모 평가로, 과학을 포함하여 5개 교과 평가가 이루어진다. 앞서 언급한 대로 학업성취도 평가의 컴퓨터 기반 평가 체제로의 전환을 위한 절차를 마련하여 2021년에는 체제 전환을 위한 예비시행을 실시하였고, 이를 기반으로 편의성을 증대시킨 체제를 마련하였다(Lee et al., 2021). 이어 2021년 본검사에서 지필평가와 컴퓨터 기반 평가를 병행시행하여 매체의 전환에 대한 영향을 점검하였고, 2022년에는 컴퓨터 기반 평가를 전면 도입하였다. 이러한 전환은 2015 개정 교육과정에서 역량이 도입됨에 따라 교육과정의 질 점검이라는 학업성취도 평가 본연의 목표를 달성하기 위한 방안으로 이루어졌다(Lee et al., 2017).

본 연구는 이러한 연구 선상에서 학업성취도 평가를 지필평가에서 컴퓨터 기반 평가로 전환하는 과정에서 점검되었던 이슈, 특히 새로운 평가 문항 유형 도입에 따른 현황과 학생들의 반응을 살펴보고자 한다. 컴퓨터 기반 평가 체제를 도입한 학업성취도 평가 결과를 토대로 학생들이 변화한 매체에서 새로운 방식으로 평가에 참여함에 따라 나타나는 차이를 탐색하였다. 이를 위해 2021년도에 시행된 학업성취도 평가의 지필평가와 컴퓨터 기반 평가의 병행시행 결과를 분석하였다. 구체적으로는 평가 매체가 변화함에 따라 발생하는 차이, 컴퓨터 기반 평가이기 때문에 사용할 수 있는 새로운 기능에 의한 차이를 살펴보는 것에 초점을 두었다. 분석한 결과를 토대로 컴퓨터를 활용한 평가를 위해 검사 도구를 개발하는 과정에서 유의할 사항을 논의하는 것에 연구의 목적이 있다.

II. 연구 방법

1. 컴퓨터 기반 평가의 문항 유형

컴퓨터 기반 평가에서는 컴퓨터의 다양한 기능을 활용하여 지필평가에서보다 다양한 문항 유형을 도입할 수 있게 되었다. 전통적인 지필평가에서 활용하는 문항 유형은 크게 선택형과 구성형(서답형)으로 구분될 수 있는데, 이를 세분화하면 선택형은 정답형, 최선답형, 합답형, 부정형 등으로, 서답형은 단답형과 서술형으로 구분할 수 있다. 현재 대부분의 학교 현장이나 학업성취도 평가에서도 이렇게 문항 유형을 구분하고 있다. 컴퓨터 기반 평가는 단순히 시행 매체가 종이에서 컴퓨터로 바뀌는 변화가 아니다. 컴퓨터가 제공하는 기술공학 기능을 활용하면 평가에서 다양한 문항 유형을 활용할 수 있게 된다. 컴퓨터 기반 평가를 도입하는 많은 논의에서 컴퓨터의 기술공학 기능을 활용한 문항이 몇 가지 용어로 개념화되어 왔는데, 그 예로 ‘혁신적 문항(innovative items)’(Parshall et al., 2010; Strain-Seymour, Way, & Dolan, 2009; Zenisky & Sireci, 2002), ‘기술공학 강화 문항(technology enhanced items)’(Bryant, 2017) 등의 표현을 사용하여 새로운 문항 유형을 설명하고 있다.

Parshall et al.(2010)에 따르면, 기술공학적 기능 활용 문항은 평가 구조(assessment structure), 응답 행위(response action), 미디어 포함 여부(media inclusion), 상호작용 수준(interactivity), 복잡성(complexity), 실제성(fidelity), 채점 방법(scoring method)의 7가지 측면에서 컴퓨터 기반 평가의 문항 유형을 분류할 수 있다. Bennett

et al.(1990, pp. 5-6)는 답안 작성에서의 학생의 자유도를 기준으로 자유도가 작은 것부터 높은 순으로 선다형(multiple choice), 선택형(selection/identification), 재배열형(reordering/rearrangement), 대체/수정형(substitution/correction), 단답(완성)형(completion), 구성형(construction), 발표/수행(presentation/performance)까지 7개의 문항 유형을 제시하였는데, Sireci & Zenisky(2006)는 Bennett et al.(1990)의 문항 유형에 대한 논의와 Parshall et al.(2000)의 응답 행위(라이트 펜, 마우스, 키보드 등)에 관한 논의를 종합적으로 고려하여 컴퓨터 기반 평가에서 활용할 수 있는 14개의 문항 유형을 제시하였다. Scalise & Gifford(2006)에서는 기술공학적 기능이 활용된 문항의 유형을 분류하기 위해 응답 제약 정도와 문항의 복잡성을 두 축으로 삼고 각 특성이 교차하는 지점에 해당하는 문항의 사례를 제시하였다. 가로축은 Bennett et al.(1990)에서 제시한 바와 같이 학생 응답의 제약 정도로 설정하였고, 세로축은 문항의 복잡성을 기준으로 한다. 두 개의 축을 복합적으로 고려하여 문항 유형화를 정교화하였다는 점에서 의미 있는 시도라고 볼 수 있다.

Lee et al.(2020; 2021)에서는 앞서 여러 선행연구를 살펴본 후, 학생 응답 측면과 자료 탐색 측면에서 컴퓨터 기반 학업성취도 평가에 활용될 만한 기술공학적 기능 목록을 추출하고, 이를 바탕으로 평가 도구 개발의 효율성 및 타당성을 고려하여 학생 응답 측면과 자료 탐색 측면에서 컴퓨터 기반 학업성취도 평가에 자주 활용될 만한 문항의 대표 유형을 선정하였다. 학생의 응답에 활용될 수 있는 기술공학적 기능은 라디오 버튼, 체크 박스, 짝 연결하기, 아래로 펼치기, 핫스팟, 끌어놓기, 그래프/선 그리기, 단답 입력, 서술 입력, 수식 입력, 표 채우기, 시뮬레이션, 도구 조작, 하이퍼텍스트, 탭, 멀티미디어 등이다. Table 1은 학업성취도 평가에서 활용하는 문항 유형을 보여준다.

학생 응답 측면에서 컴퓨터 기반 평가의 문항 유형은 크게 선택형과 구성형으로 범주를 구분할 수 있다. 선택형은 단순선택형과 확장선택형으로 세분할 수 있는데, 단순선택형은 기존 지필평가의 5지 선다형과 유사한 구조이지만, 경우에 따라서 선택지의 수를 조절할 수 있다. 확장선택형은 체크박스, 핫스팟, 아래로 펼치기를 활용하여 주어진 자료나 선택지에서 답안을 선택하는 유형이다. 선택지의 수를 자유롭게 조정할 수 있으며, 답안을 복수 선택하도록 구성할 수 있다. 경우에 따라서는 자료에서 직접 선택하여 응답하는 방식으로 구성하

여 선택지 구성을 위한 불필요한 문자화나 기호화 과정을 생략할 수 있는 장점이 있다. 구성형은 6가지 유형으로 세분되는데, 자료연결형은 서로 관련 있는 자료를 찾아 같은 영역에 끌어 놓거나 연결하는 유형이며, 순서배열형은 해당 단어 및 구, 숫자, 문자, 이미지 등을 끌어 놓아 적절한 순서로 재정렬하는 유형이다. 선택형 중 ‘확장선택형’과 구성형 중 ‘자료연결형’, ‘순서배열형’은 OMR 카드 마킹 방식을 탈피하면서도 기술공학적 기능을 활용하여 학생의 응답을 수기 채점이 아니라 기계채점이 되도록 설계된 형태이다. 특히 ‘확장선택형’과 ‘자료연결형’은 기존 지필평가의 선다형을 상당수 대체할 수 있으며 추측에 의한 응답 확률을 낮춰, 평가의 타당도를 제고할 것으로 기대된다. 구성형에서 단순히 키보드 입력만으로 작성하는 경우에는 ‘단답형’과 ‘서술형’이 기존 지필평가와 유사한 형태이나, 수식 입력기를 활용하여 작성하거나 각종 편집 기능이 제공되는 점에서 컴퓨터 기반 평가 문항의 특성이 잘 드러난다. ‘수정형’은 제시문이나 제시 자료에서 틀린 문장이나 내용을 찾아 선택한 후, 고쳐 쓰게 하는 형태이다. 기존 지필평가에서 이와 유사한 문항을 출제할 경우, 기호화 과정을 통해 틀린 문장을 먼저 찾는 선다형(또는 단답형)과 이를 고쳐 쓰게 하는 단답형(또는 서술형)의 2개 문항으로 구성하게 되어 둘 중 한 문항만 출제되거나 출제되지 못하였다. 컴퓨터 기반 평가에서는 제시 자료의 맥락 속에서 찾아 고치게 함으로써 불필요한 단계를 생략할 수 있고, 학생들이 무엇이 틀렸는지를 온전히 알고 있는지를 평가한다는 점에서 평가의 타당도를 높일 수 있다. ‘그래프/그림 완성형’은 컴퓨터 기반 평가 문항의 특색을 가장 두드러지게 보여주는 유형으로서, 그림판을 활용하여 답안을 완성하거나 격자로 제시된 화면에서 그래프 생성 기능을 활용하여 그래프를 작성할 수 있다. 이는 기존 지필평가에서 출제 및 채점을 수행하기에 불리했던 유형으로, 격자에서 특정 지점 선택을 통한 간단한 그래프 작성의 경우에는 자동채점까지 가능할 것으로 기대된다.

자료 탐색 측면의 문항 유형은 기존 지필평가의 자료 제시 형태와 가장 유사한 ‘단순제시형’부터 컴퓨터 기반의 시험 환경이기 때문에 가능한 ‘정보활용형’, ‘미디어활용형’, ‘도구 조작 및 시뮬레이션형’, ‘대화형’을 포함한다. 기존 지필평가와 가장 유사한 ‘단순제시형’도 흑백이 아닌 컬러 그래픽이나 사진을 제공할 수 있다는 점에서 문항의 타당도를 높이는 데 기여할 수 있다. ‘정보활용형’은 기존 지필평가에서 공간 배치의 한계를 극복하면서 필요한 정보 및 자료를 적절

Table 1. Item types using in the computer-based National Assessment of Educational Achievement

학생 응답 측면			자료 탐색 측면	
범주	문항 유형	기술공학적 기능	문항 유형	기술공학적 기능
선택형	1	단순선택형	1	단순제시형
	2	확장선택형	2	정보활용형
	3	자료연결형	3	미디어활용형
	4	순서배열형	4	도구조작 및 시뮬레이션형
구성형	5	그래프/그림 완성형	5	대화형
	6	단답형		가상 대화
	7	서술형		
	8	수정형		

※ 출처: Lee et al.(2021, pp. 252, 269)

Table 2. Subjects of the research

구분	지필평가 (PBT)	단순 모드 전환형 검사 (CBT-M)	컴퓨터 기술공학 기능 활용형 검사 (CBT-P)*		
문항 수 (개)	선다형	18	단순선택형 (라디오버튼)	12	
			자료연결형 (끌어놓기)	2	
	서답형	10	확장선택형 (체크박스 외)	3	
			구성형 (단답형, 서답형)	8	
			확장선택형 (핫스팟)	2	
		그래프/그림 완성형 (그래프/선 그리기)	1	그래프/그림 완성형 (그래프/선 그리기, 끌어놓기)	3
문항 수 소계 (개)	28	28	30		
참여 학생 수 (명)	6,413	477	247		

* CBT-P형 문항의 일부는 CBT-M형에 포함할 수 없는 기능을 활용하였기 때문에 대응하는 CBT-M 문항이 없으며, 해당 문항은 본 연구의 분석 대상에서 제외함.

히 제공할 수 있다. 이 유형에는 간단한 추가 정보를 제시하는 데 유용한 정보 팝업, 성격이 서로 다른 정보를 제시하는 데 유용한 탭, 연관된 정보나 필요한 정보로 빠르게 찾아갈 수 있도록 돕는 하이퍼 텍스트/하이퍼링크, 문제 해결에 필요한 정보를 주도적으로 직접 찾아볼 수 있게 하는 정보 검색 기능이 다양하게 활용될 수 있다. ‘미디어활용형’은 각종 형태의 오디오 및 동영상 자료를 활용하여 자료를 제시하는 것으로, 문항이 설정한 상황과 맥락을 학생들에게 친숙한 방식으로 더 손쉽게 전달할 수 있다. ‘도구 조작 및 시뮬레이션형’과 ‘대화형’은 기존 지필평가에서 매체의 한계로 인해 평가하기 어려웠던 성취기준을 적절히 평가하는 데 도움이 될 수 있는 유형이다. 이들 유형은 앞에서 언급한 바와 같이, 자료 탐색의 과정에서 학생과 문항 간 상호작용 수준이 높은 유형으로, 단순히 응답 결과뿐만 아니라 응답을 도출하는 과정에서 역량까지 평가할 수 있을 것으로 기대된다.

2. 연구 대상

본 연구에서는 학업성취도 평가를 컴퓨터 기반 평가로 전환하는 과정에서 새로운 문항 유형을 도입하는 것이 학생들의 답지 반응에 어떠한 영향을 미치는지 살펴보기 위해 지필평가와 컴퓨터 기반 평가를 병행시행한 2021년 학업성취도 평가 자료를 활용하여 분석하였다. 2021년 학업성취도 평가에서는 컴퓨터 기반 평가의 도입을 준비하고자 지필평가와 컴퓨터 기반 평가의 매체 차이가 학생 수행에 미치는 영향을 살펴보기 위해 두 종류의 컴퓨터 기반 평가 방식의 검사 (CBT-M, CBT-P)를 개발하여 병행시행하였다. 단순 모드 전환형 (CBT-M) 검사는 지필평가 방식의 학업성취도 평가와 동일한 문항으로 구성된 검사로 문항별 응답 결과를 통해 매체의 변화라는 측면에서 동일한 문항에 대해 지필평가와 컴퓨터 기반 평가 간의 차이를 살펴볼 수 있게 한다. 한편 컴퓨터 기술공학 기능 활용형(CBT-P) 검사는 기술공학 기능을 활용한 유형의 문항을 포함한 검사로, 새롭게 적용되는 기술공학적 기능에 대한 학생들의 수행 특성을 살펴볼 수 있게 해 준다.¹⁾ 특히 컴퓨터로 동일한 내용을 평가하지만 지필평가의 응답 방식과 자료 구성을 그대로 따르는 CBT-M형의 문항과 비교해

봄으로써, 새로운 응답 방식과 자료 구성 방식의 차이가 학생들의 응답에 미치는 영향을 살펴볼 수 있게 하여 문항 개발에 대한 시사점을 제공할 수 있다.

본 연구에서 초점을 둔 병행시행에서 학생들은 앞서 언급한 3종의 검사 중 사전에 배정된 1종의 검사에 참여하였다. 이에 따라 연구에 활용된 검사지별 문항 수와 각 평가에 응시한 학생 수는 Table 2와 같다. 표에서 보듯이 단순 모드 전환형은 지필 평가와 문항 수와 문항 유형이 동일하고, 컴퓨터 기술 공학 활용형은 일부 문항에 대해서 기술공학 기능을 활용한 것을 볼 수 있다.

연구에서 평가 매체별 문항의 정답률, 변별도를 산출하였으며, 동일한 소재를 활용하여 변형한 문항에 대해서 정답률, 변별도를 상호 비교하였다. 또한, 각 문항별 응답 특성을 분석하기 위해 학업성취도 평가 출제에 참여한 경험이 있는 과학 교사 8명을 대상으로 전문가 협의회를 개최하여 응답 특성의 원인에 대한 전문가 의견을 수렴하였다.

III. 연구 결과 및 논의

2021년 시행된 학업성취도 평가의 병행시행 결과는 대체로 평가의 매체 변화에 따른 영향은 크지 않은 것으로 나타났다. 다만 응답 특성을 분석한 결과에 따르면 응답의 편이성, 화면의 배치 등의 요인들이 학생들에게 일부 영향을 주었을 것으로 예상할 수 있었다. 병행시행이 지필평가와 함께 두 가지 종류로 구분되는 컴퓨터 기반 평가의 검사도구를 활용했다는 점에 기초하여 단순 매체 전환에 따른 응답 특성, 문항 유형의 다양화에 따른 응답 특성으로 구분하여 연구 결과를 제시하였다.

1) 본 연구에서는 컴퓨터 기반 평가와 CBT, 지필평가와 PBT를 표기할 때 기급적 국문 용어로 작성하고자 하였으나, 컴퓨터 기반 평가의 검사 결과를 CBT-M형과 CBT-P형으로 구분할 필요가 있는 경우, 영문 용어를 혼용하여 작성하였음.

1. 매체의 변화에 따른 응답 특성

매체의 변화에 따른 응답 특성의 변화를 살펴보기 위해서는 지필평가의 내용과 문항 구조를 그대로 가지고 있으면서 종이 대신 컴퓨터 화면으로 매체만 바뀐 형태의 컴퓨터 기반 평가 결과를 지필평가의 결과와 비교하는 것이 필요하다. 즉 지필평가와 단순 모드 전환형(CBT-M형) 평가를 비교해야 한다. Figure 1은 동일한 문항으로 구성된 CBT-M형과 지필평가 문항의 정답률을 비교한 결과이다. 결과에 따르면, 두 유형 간의 정답률은 대체로 유사하며, 전반적인 그래프의 개형도 유사한 것으로 나타났다. 이와 같은 결과는 컴퓨터 기반 평가로 전환한 PISA나 TIMSS, 호주의 대규모 성취도 평가인 NAP(National Assessment Program)의 과학적 소양에 대한 평가에서 두 방식이 서로 동등하게 활용될 수 있다고 보고한 것과 유사한 결과이다(ACARA, 2017; Kim, 2021; Song *et al.*, 2014). 즉, 학업성취도 평가의 병행시행 결과, 각 문항별로 다소의 정답률 편차는 발생할 수 있으나 그 편차가 크게 유의미하지 않았고, 그 분석 결과도 선행연구들과 비슷하게 나타났다.

다만, 문항 수준에서 정답률의 편차가 발생한 이유에 대해 살펴보는 것은 컴퓨터 기반 평가의 문항 구성에 대한 시사점을 제공할 수 있다. 대체로 정답률의 편차는 선다형의 경우, 최대 6.39%p 이내인 것으로 나타났고, 서답형의 경우에는 최대 11.30%p인 것으로 나타났

다. 위와 같은 결과로부터 소폭이나마 서답형과 선다형 문항 간의 정답률 편차가 발생한 점, 일부 문항에서 상대적으로 높은 정답률 편차를 보이는 사례들을 살펴볼 필요가 있다. 이에 대해 각각 응답의 편이성, 화면의 배치와 연관되는 것으로 구분하여 살펴보고자 한다.

가. 텍스트 입력의 편이성

학업성취도 평가의 과학과 검사 도구는 선다형 문항을 앞에, 서답형 문항을 뒤에 배치한다. 이는 학생들에게 다소 높은 수준의 인지적 능력을 요구하는 문항을 검사지 후반에 제시하기 위함이다. 앞서 제시한 Figure 1에서 두드러진 특징은 검사 후반에 배치한 서답형 문항의 정답률이 CBT-M형에서 다소 높았던 점과 함께, 검사 후반으로 갈수록 두 방식 간의 편차가 커진다는 점이었다. Table 3은 CBT-M과 지필평가에서 서답형 문항의 정답률과 변별도를 제시한 것으로, 전반적으로 CBT-M의 정답률이 높고, 변별도는 유사하거나 다소 CBT-M에서 높은 것을 확인할 수 있다. 앞서 언급한 바와 같이 문항 번호를 기준으로 보았을 때, 검사 후반에 배치한 문항의 정답률 차가 더 커지는 결과가 나타났다. 예를 들어 서술형으로 한정하여 Table 3의 결과를 살펴보면, 상대적으로 전반부의 문항의 정답률은 2%p 이내에서 편차가 발생하는 반면, 24-1번 문항의 경우 10%p 이상 편차가 발생하는 것으로 나타났다.

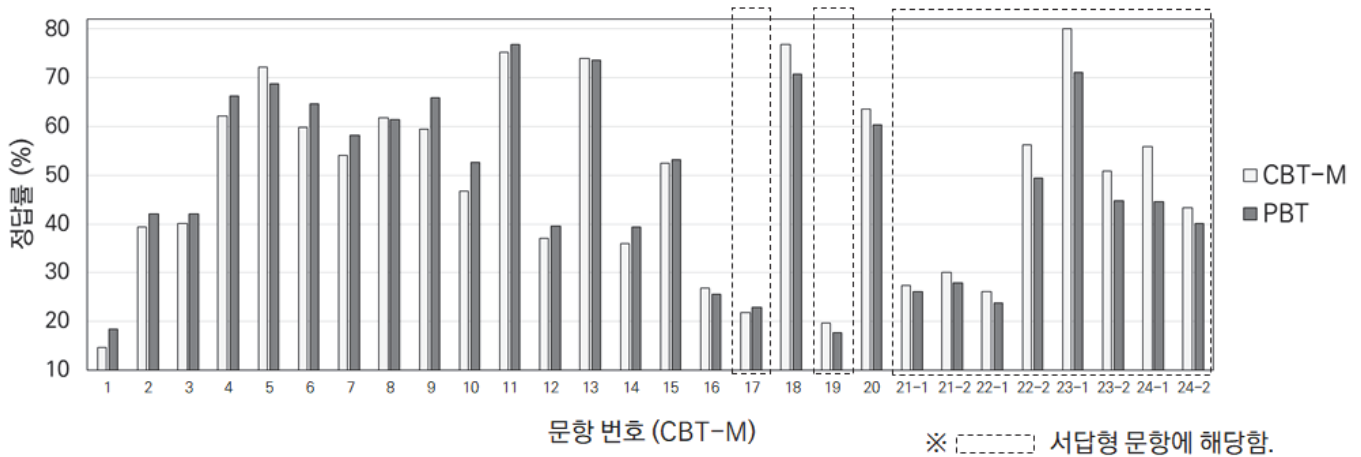


Figure 1. Percentage for correct answers in CBT-M and PBT

Table 3. Percentage for correct answers and item discrimination of constructed response items in CBT-M and PBT

구분	문항 번호 (CBT-M)	정답률(%)			변별도		
		CBT-M (A)	PBT (B)	A-B (%p)	CBT-M (C)	PBT (D)	C-D
단답형	21-1	27.31	26.13	1.18	0.64	0.61	0.03
	21-2	30.04	27.86	2.18	0.69	0.66	0.03
	22-2	56.30	49.33	6.97	0.43	0.51	-0.08
	23-1	80.04	71.05	8.99	0.43	0.54	-0.11
서술형	17	21.74	22.82	-1.08	0.44	0.52	-0.08
	19	19.64	17.72	1.92	0.54	0.54	0.00
	22-1	26.05	23.72	2.33	0.65	0.65	0.00
	23-2	50.84	44.72	6.12	0.62	0.61	0.01
	24-1	55.88	44.58	11.30	0.65	0.67	-0.02

Table 4. Results and characteristics of items which occurred scrolling on display

문항 번호 (CBT-M)	정답률 (%)			변별도			화면 구성의 특징
	CBT-M (A)	PBT (B)	A-B (%p)	CBT-M (C)	PBT (D)	C-D	
6	59.87	64.58	-4.71	0.39	0.45	-0.06	스크롤 발생
7	53.99	58.18	-4.19	0.50	0.50	0.00	스크롤 발생
9	59.45	65.84	-6.39	0.49	0.50	-0.01	스크롤 발생
11	75.21	76.77	-1.56	0.46	0.50	-0.04	주요 정보와 선택지가 한 화면에 포함되는 형태로 스크롤 발생
12	36.97	39.53	-2.56	0.44	0.39	0.05	스크롤 발생
13	73.95	73.60	0.35	0.55	0.56	-0.01	주요 정보와 선택지가 한 화면에 포함되는 형태로 스크롤 발생
14	35.92	39.30	-3.38	0.13	0.20	-0.07	스크롤 발생
16	26.89	25.62	1.27	0.07	0.15	-0.08	스크롤 발생
18	76.89	70.77	6.12	0.42	0.56	-0.14	자료와 선택지 기준으로 화면 좌우 분할
20	63.66	60.36	3.30	0.51	0.52	-0.01	자료와 선택지 기준으로 화면 좌우 분할

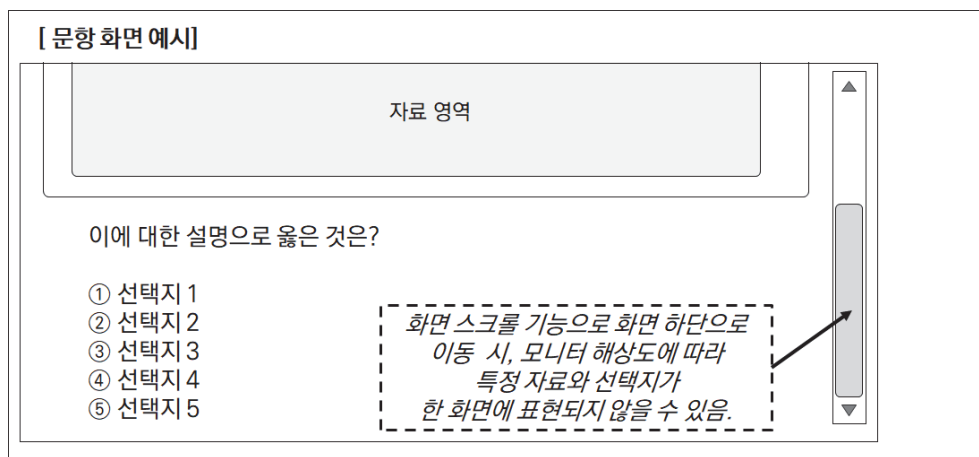


Figure 2. An example of item which occurred scrolling on display

위와 같은 결과가 나타나는 것은 학생들의 정보 입력 방식에 대한 익숙함과 관련성이 있는 것으로 보인다. 즉, 지필평가에서 필기도구로 작성하던 것에 비해 키보드를 이용하여 텍스트를 입력하는 것이 학생들에게 더 익숙한 것과 연관지을 수 있다. 이와 관련하여 교사들은 모니터를 통해 정보를 습득하는 것을 학생들이 편하게 생각함과 동시에 키보드를 이용한 정보 입력을 편하게 생각한다고 언급하면서, 높은 집중력을 가지고 성실하게 평가에 참여하는 데 영향을 끼쳤을 것이라고 말하였다. 컴퓨터 기반 평가로 전환한 대규모 평가들과 함께 다양한 연구들에서도 이와 같이 컴퓨터 활용 능력이나 디지털 매체에 대한 자신감을 주목하였는데(Fishbein *et al.*, 2018; Hardcastle, Herrmann-Abell, & DeBoer, 2017; Lee *et al.*, 2018), 이미 태블릿 등의 디지털 매체를 충분히 활용한 경험이 있는 중학생들에게는 컴퓨터 기반 평가의 방식이 더 친숙하고 편리하게 여겨지는 것으로 보인다.

나. 화면 내 문항의 배치

학업성취도 평가의 병행시행 결과에서 특징적인 점은 일부 문항들에서 정답률의 편차가 다소 두드러지는 경우들이 있었다는 점이다. 기존의 5지 선다형과 유사한 형태인 단순선택형 문항을 중심으로 살

펴보면, 전체 18개의 단순선택형 문항에서 정답률 편차는 최대 6.39%p로 나타났으며, 약 5%p 내외의 편차가 발생하는 문항들이 일부 존재한다. 이와 같은 차이의 원인은 다양한 요인들을 고려해야 하며, Kim(2021)의 연구에서는 컴퓨터를 활용할 때 메모의 불가를 원인으로 꼽기도 하였다. 이러한 요인은 문항 해결과정에서 계산이 필요하거나 주요 정보의 기록이 요구되는 상황에서 학생들이 자신의 기억을 별도로 표기하기 어렵기 때문에 다소 정답률을 낮추는 방향으로 영향을 끼칠 수 있다. 다만 본 연구가 초점을 둔 학업성취도 평가의 문항에서는 계산을 요구하는 문항의 수가 매우 적었기 때문에, 다른 요인의 영향을 탐색할 필요가 있다.

컴퓨터로 매체가 변화할 때의 가장 큰 변화는 화면을 통해 정보를 습득한다는 점이다. 이는 문항에 제시되는 자료의 양에 따라 화면의 스크롤이 발생할 수 있는 가능성이 있으며, 이 경우 정보가 분절된 형태로 학생들에게 제공된다. 즉, 화면의 배치가 학생들이 정보를 읽고 해석할 때 영향을 주고 있는 것으로 볼 수 있다. Table 4는 18개의 단순선택형 문항 중 화면에 스크롤이 발생하거나, 화면의 배치가 다른 형태인 문항들의 정답률과 변별도를 비교한 것이다.2) 기본적인

2) 학업성취도 평가의 컴퓨터 기반 평가 시스템에서 제공하는 기본 해상도의 화면을 확대 또는 축소하지 않은 상태를 기준으로 구분하였음.

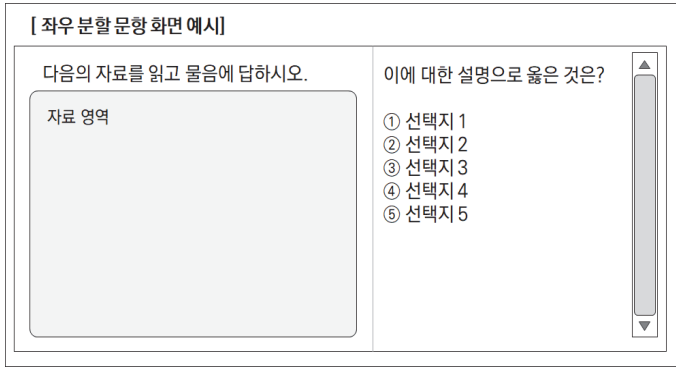


Figure 3. An example of item which is arranged information and question from side to side

문항의 배치가 위에서 아래로 나열되는 방식인 가운데, 8개의 문항은 하나의 화면에 모든 정보가 포함되지 않아 스크롤이 발생한 경우에 해당한다. 반면, 18, 20번 문항은 화면 좌측에 문항의 자료, 우측에 선택지를 배열하는 방식으로 화면이 구현된 경우이다. 이때 주요한 특징은 Figure 2와 같이 스크롤이 발생하여 문항에서 탐색해야 하는 주요 정보와 선택지 간에 정보가 분절되는 경우 정답률이 다소 하락하는 양상을 보인다는 점이다. 6, 7, 9, 12, 14, 16번의 6개 문항에서 자료와 선택지가 분절적으로 제시되었으며, 정답률의 편차가 1.27%p에서 6.39%p의 범위에서 나타났다. 반면, 화면을 좌우로 분할하여 자료와 선택지 간의 시각적 거리가 가까운 경우(18, 20번)에는 문항의 정답률이 상승하는 양상을 보였다.

이러한 응답의 특성은 문항 내에서 발생하는 자료의 분절이 학생들에게 미치는 영향을 보여준다. 자료의 분절은 학생들이 자료와 선택지를 서로 비교하며 정오를 판단함에 있어 제약을 낳게 된다. Pommerich(2004)는 지필평가에서 학생들이 경험했던 ‘위치 기억 (positional memory)’을 언급하면서, 컴퓨터 화면에서 발생하는 스크롤은 학생들이 특정 정보를 찾고 연결하는 데 불리한 영향을 끼친다고 하였다. 즉, 지필평가에서는 지면상에 정보의 위치가 고정되어 있었으나, 컴퓨터 기반 평가에서는 기존에 경험하지 못한 방식으로 화면을 옮겨가며 정보를 기억해야 하므로 정답률에 영향을 미쳤을 것으로 예상할 수 있다. Kim(2000)이 언급한 바와 같이 인간이 단기 기억으로 저장할 수 있는 정보의 양이 한정되어 웹 화면에 표시되는 정보

의 양과 복잡성을 낮추는 것이 필요함을 감안하면, 컴퓨터 기반 평가에서도 지필평가와 가급적 유사한 방식으로 자료를 탐색할 수 있게 화면을 구성하는 것이 필요하다. 예를 들어, 필수적인 자료만으로 문항을 구현하거나, 정보의 양이 상대적으로 많아지는 경우에 Figure 3과 같이 화면의 좌우 분할 기능 등을 적극적으로 활용하는 것이 필요하다.

2. 새로운 기능 도입에 따른 응답 특성

컴퓨터 기반 평가의 도입은 기존의 지필평가에서 구현하지 못했던 응답 방식이나 자료 제시 방식을 도입할 수 있다는 측면에서 혁신적이다. 앞서 Table 1에서 제시한 바와 같이 학업성취도 평가에서도 새로운 유형의 문항들이 다수 마련되었다. 이러한 변화는 다양한 이점을 가져올 수 있지만, 이와 더불어 학생들이 새로운 기능들에 어떻게 반응하는지 응답 특성을 분석하여 컴퓨터 기반 평가에서의 문항 개발에 대한 시사점을 탐색할 필요가 있다.

본 연구에서 탐색한 학업성취도 평가 병행시행의 CBT-P형에서는 컴퓨터 기반 평가에서 새롭게 도입된 기능 중 일부를 활용하여 문항을 개발하였다. Table 5는 응답 방식과 관련해 새로운 기능들이 도입된 CBT-P형과 이에 대응하는 CBT-M형 문항들을 제시한 것이다. 즉 두 문항은 문항의 내용이나 맥락은 동일하고 단지 답지 반응 유형만을 달리한 경우이다. CBT-M형과의 비교는 매체에 의한 요인을 배제하고 새로운 기능 도입에 의한 요인만을 확인하기 위한 것이다. 총 7개의 문항에 새로운 기능들이 적용되었고, 기능으로 구분하면 체크박스, 아래로 펼치기, 끌어서 놓기, 핫스팟 등 다양한 기능들의 영향을 확인할 수 있다. 결과에 따르면 4개의 문항은 정답률의 편차가 크지 않으나, 3개의 문항에서는 10%p 이상의 정답률 편차가 나타났다. 이러한 결과를 새롭게 도입된 기능의 특성들을 중심으로 살펴보고자 한다.

가. 확장선택형: 체크박스, 아래로 펼치기, 핫스팟

학업성취도 평가에서 도입한 컴퓨터 기반 평가의 새로운 문항 유형 중, 2021년 병행시행에서 활용된 확장선택형은 총 4개 문항이다. 이

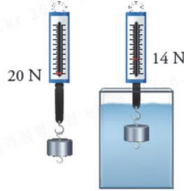
Table 5. Percentage for correct answers and item discrimination of technology-enhanced items in CBT-P and CBT-M

문항 번호 (CBT-P)	정답률 (%)			변별도			유형(기능)	
	CBT-P (A)	CBT-M (B)	A-B (%p)	CBT-P (C)	CBT-M (D)	C-D	CBT-P	CBT-M
6	52.63	53.99	-1.36	0.40	0.50	-0.10	자료연결형 (끌어놓기)	선다형 (라디오버튼)
11	17.41	36.97	-19.56	0.32	0.44	-0.12	확장선택형 (체크박스)	선다형 (라디오버튼)
12	37.25	35.92	1.33	0.14	0.13	0.01	확장선택형 (아래로 펼치기)	선다형 (라디오버튼)
18	48.18	63.66	-15.48	0.34	0.51	-0.17	자료연결형 (끌어놓기)	선다형 (라디오버튼)
21-2	51.82	56.30	-4.48	0.49	0.43	0.06	확장선택형 (핫스팟)	단답형 (단답 입력)
22-1	91.09	80.04	11.05	0.34	0.43	-0.09	확장선택형 (핫스팟)	단답형 (단답 입력)

유형은 선택지에서 복수 선택이 가능하며, 일부 기능의 경우에는 문항에 제시한 자료에서 직접 선택지를 선택하는 방식으로 응답이 가능하다는 특징을 갖는다. 병행시행의 CBT-P형에는 확장선택형 중 체크박스, 아래로 펼치기, 핫스팟 기능이 적용된 문항들이 포함되었는데, 이 중 핫스팟 문항은 2개 문항이 출제되었다. 복수 선택 가능성 및 자료에서 직접 선택지를 선택하도록 도입한 기능의 측면에서 평가 결과를 살펴보았다.

우선, 체크박스 기능은 선다형 문항과 유사한 방식이면서 답안의 복수 선택이 가능하다는 특징을 갖는다. Figure 4에 제시한 예시 문항과 같이 각각의 선택지의 진위 여부에 대해 학생들이 판단하여 정오를 입력해야 한다.³⁾ 이는 기존의 5지 선다형 문항에서 특정 선택지의 진위를 판단하면 다른 선택지의 진위 여부는 판단하지 않아도 되었던 것과 달리, 모든 선택지의 내용을 이해하는지 평가할 수 있다는 이점을 갖는다.

1 추를 용수철저울에 매단 후, 잠시 후 물이 담긴 수조에 잠기게 하였더니 저울의 눈금이 그림과 같이 변하였다.



위의 실험 결과를 근거로 할 때, 옳게 말한 학생을 모두 고르시오.

- 추에 작용하는 중력의 크기는 20 N이다.
- 용수철의 탄성력은 추를 물에 넣게 전과 후가 같다.
- 추에 작용하는 부력은 중력과 같은 방향이다.
- 추에 작용하는 부력은 6 N이다.

Figure 4. An example of extended multiple-choice item with checkbox

2021년 병행시행에서는 11번 문항에 이 기능을 적용하였으며, 총 3개의 선택지를 제시하고 각 선택지의 진위 여부를 판단하여 답안을 선택하도록 구성하였다. 대응하는 CBT-M형에서 3개의 선택지를 활용한 합답형 문항으로 구성된 사례이다. 이 문항은 우리 몸에서 일어나는 소화 과정을 물어보는데, 각 선택지가 독립적인 성격을 갖고 있다. CBT-M형에서 정답률이 36.97%인 반면, 체크박스 기능을 적용한 경우 17.41%의 정답률을 보여 기능을 적용할 때 정답률이 19.56%p 하락했다. 이와 같은 결과는 학생들의 추측에 의한 답안 선택의 가능성과 관련하여 설명할 수 있다. Table 6은 두 검사에서 보인 학생들의 답지 반응률로, 특정 오답 유형의 답지 반응은 유사하나 체크박스 기능을 적용할 때는 이전에 없던 답지 중에서, 특정 답지인 ㄱ, ㄷ으로 응답한 반응률이 15.79%로 나타났다. 즉, 해당 선택지가 문항에 제시되어 있지 않았기 때문에 CBT-M형에서 정답을 고른 학생들이 체크박스 기능을 적용한 경우 자신의 과학적이지 않은 개념에 따라 다른 선택지를 고른 것을 볼 수 있다.

3) 본 연구에서 제시한 예시 문항들은 국가수준 학업성취도 평가의 컴퓨터 기반 평가 시스템의 문제은행에 탑재되어 있는 문항이 아니며, 본 논문에서 언급하는 문항 유형의 이해를 돕기 위해 컴퓨터 기반 평가 시스템을 이용하여 임의로 구성한 것이다.

Table 6. The response rate of options of item 11 in CBT-P

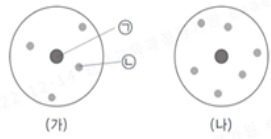
답지 반응률 (%)	정답 (ㄱ, ㄴ, ㄷ)	두 검사에 공통으로 있는 선택지 조합		CBT-P형에만 있는 선택지 조합 (ㄱ, ㄷ)
		ㄱ, ㄴ	ㄴ, ㄷ	
CBT-M	36.97	27.10	26.89	-
CBT-P	17.41	27.13	25.10	15.79

이와 같은 결과는 학생들이 평가 상황에서 일반적으로 취하던 문항의 해결 전략에서 차이가 발생함을 의미한다. 즉, 기존의 지필평가 합답형 문항에서는 선택지의 배열을 토대로 상대적으로 익숙하거나 자신이 잘 이해하는 내용을 중심으로 해석하고 응답하는 것이 가능하였으나, 체크박스 기능의 도입은 이러한 방식을 적용할 수 없게 한다. 즉, 추측에 의한 답지 선택을 배제할 수 있다는 측면에서 새로운 기능 도입의 의미를 확인하게 한다. 특히, 각 선택지가 독립적이며 서로 다른 과학 개념에 대응하는 경우, 학생들의 과학적이지 않은 개념을 세밀하게 살펴볼 수 있다는 이점이 있다.

다만, 컴퓨터 기반 평가에서 문항을 개발하는 경우, 각각의 선택지를 모두 판단해야 하는 이와 같은 문항에서 선택지의 수를 과도하게 증가시키는 것은 유의할 필요가 있다. 각 선택지가 서로 독립적인 경우, 학생들의 관점에서는 서로 다른 문항의 수가 증가하는 것과 유사한 효과를 낼 수 있다. 따라서 검사 전체의 문항 수와 학생들에게 제공되는 평가 시간을 고려하여 과도한 수의 선택지를 배치하지 않도록 유의해야 한다.

아래로 펼치기 기능 적용 문항에서도 선택지 배치와 관련된 학생들의 문항 풀이 전략을 살펴볼 수 있다. 아래로 펼치기 기능은 컴퓨터에서 자주 활용하는 기능인 드롭다운(drop-down list) 기능을 문항에 직접 구현한 것으로 Figure 5와 같은 방식으로 제시된다. Figure 4의 사례와 같이 자료의 문장이나 이미지상에 직접 단어 등을 선택하여 배치함으로써 불필요한 문자를 포함하는 것을 막고, 직관적으로 학생들이 자신의 응답을 자료와 비교할 수 있도록 구현하였다.

1 그림은 학생이 그린 원자 (가), (나)의 모형을 나타낸 것이다.



학생이 그린 모형이 적절할 때, 빈 칸에 들어갈 내용으로 가장 적절한 것을 선택하시오

모형 (가)에서 전자는 으로, 원자핵의 주위에서 움직인다. 두 원자 중 원자핵의 전하량이 큰 것은 이다.

(가)
(나)

Figure 5. An example of extended multiple-choice item with drop-down list

2021년 병행시행에서는 12번 문항에 이 기능을 적용하였으며, 2개의 아래로 펼치기 항목에 각각 동일한 선택지를 3개씩 배치하였다. 지구 자전에 의한 천체의 겉보기 운동을 묻는 이 문항은 시간에 따른

천체의 위치를 물어보았다. 이 문항에 대응하는 CBT-M형의 문항은 아래로 펼치기 항목의 선택지들을 짝지어 5지 선다형으로 개발한 것이다. 두 검사에서 문항의 정답률은 CBT-P형과 CBT-M형에서 각각 37.25%, 35.92%로 두드러지는 차이는 확인할 수 없었다. 이는 이 문항의 선택지 배열이 갖는 특징이자, 앞서 언급한 학생들의 문항 풀이 전략과 일부 연관된다고 볼 수 있다. Figure 6에 제시한 예시와 같이 두 가지 아래로 펼치기 항목의 선택지들이 중복되는 경우, ①번 항목에서 특정 선택지를 선택하면 ②번 항목에서 해당 선택지는 배제가 가능하다.

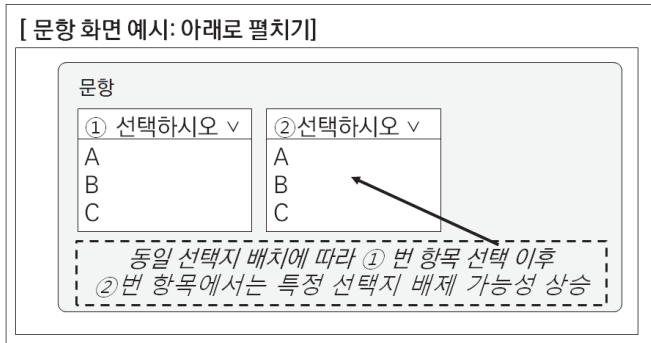


Figure 6. An example of item using drop-down list which has duplicated options

이러한 특성은 두 방식 간의 답지 반응 비교에서도 살펴볼 수 있다. Table 7은 12번 문항의 답지 반응률을 나타낸 것으로, 앞서 언급한 문항 풀이 전략을 확인할 수 있다. 이 표에서 확인할 수 있듯이, 두 검사에서 정답률은 큰 차이가 없는 가운데 두 검사에 공통으로 존재하는 선택지 조합의 답지 반응률이 다소 차이가 있는 것을 확인할 수 있다. 우선, CBT-P형에만 있는 선택지 조합 중, 'A,A', 'B,B', 'C,C'와 같이 두 선택지를 같은 것으로 고른 학생들의 비율은 4.04%로 나타나, 두 선택지를 같은 것으로 고르지 않는다는 학생들의 응답 양상을 확인할 수 있었다. 한편, 'B,C' 유형은 CBT-M형에서 없던 답안 유형인데, 이를 응답한 비율의 수가 17.00%로 나타났다. 이는 앞서 체크박스 기능 적용 문항에서 살펴볼 수 있던 추측에 의한 응답 배제를 의미함과 동시에, 학생들이 보이는 과학적이지 않은 개념을 확인할 수 있다는 점에서 의미가 있다.

핫스팟은 자료의 그림이나 텍스트를 직접 선택하여 응답하는 방식을 구현한 기능으로, Figure 7과 같이 구현된다. 이 기능은 선택지를 선택하면 Figure 6과 같이 색상이 변경되거나 표기가 이루어져 학생들이 자신의 응답에 주목할 수 있도록 돕는 효과를 가져온다. 또한 문항에서 자료의 특정 영역이나 텍스트를 지칭하기 위한 문자를 배치하지 않아도 된다는 점에서 이점이 있다.

1 다음은 중위도 온대지구의 구조를 간략히 나타낸 것이다. 그림에서 적운형 구름이 발달하여 소나기성 비가 내릴 것으로 보이는 지역 위의 한 곳을 골라 선택하시오.

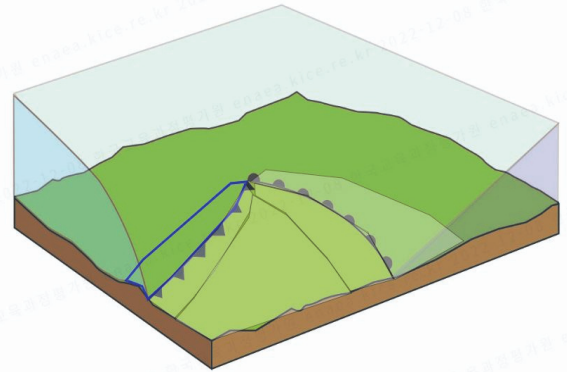


Figure 7. An example of extended multiple-choice item with hot spot

2021년 병행시험에서는 2개 문항에 핫스팟 기능을 적용하였으며, 이 문항들은 자료상의 특정 지점을 선택하도록 구성하였다. 이 문항들에 대응하는 CBT-M형의 문항은 모두 단답형으로 자료에 표기된 기호를 적는 방식으로 구성된 문항이다. 두 문항에 대한 CBT-P형과 CBT-M형의 정답률 편차는 서로 반대의 양상을 보인다. 21-2번 문항은 핫스팟 기능의 적용에 따라 정답률이 4.48%p 하락하였고, 22-1번 문항의 정답률은 11.05%p 증가하였다. 두 문항이 각각의 검사에서 구현된 방식이 동일하였다는 점을 고려하였을 때, 문항 구성의 특성을 살펴볼 필요가 있다.

문항 구성의 특성을 위해 살펴볼 것은 두 문항에서 모두 무응답 또는 무의미한 응답의 비율이 감소하였다는 점이다. 21-2번 문항에서는 이 비율이 7.05%p 감소하였고, 22-1번 문항에서는 13.89%p 감소하였다. 즉, 응답을 위해 텍스트를 작성하는 방식에서 선택의 방식으로 변화함에 따라 학생들이 보다 적극적으로 응답에 임했다고 볼 수 있다. 이 과정에 대해 교사들은 제시된 자료의 특성에 주목하였다. 21-2번의 경우 문항의 특성상 자료에 포함된 기호만이 선택 영역으로 지정된 반면, 22-1번의 경우 그림과 기호가 모두 선택 영역으로 지정되어 보다 직관적으로 응답하기 용이했다는 점이다. 즉, 학생들이 얼마나 직관적으로 답안을 선택할 수 있는지가 응답에 영향을 미친다고 볼 수 있다. 이 사례와 같이 컴퓨터 기반 평가에서 자료와 선택지 간의 경계가 모호해진다는 점은 학생들에게 명확한 응답 방식의 안내가 필요함을 의미한다. 이와 관련해 호주의 대규모 학업성취도 평가인 NAPLAN은 핫스팟과 같은 유형의 문항에서 선택 영역별로 색상을 다르게 지정하는 등의 방식을 적용하고 있으며, 이는 컴퓨터 기반 평가 문항 개발에 있어 시사점을 제공한다(ACER, 2016, p. 74). 아울러 문항의 발문 등에서 명확하게 응답 방법을 기술해주는 것이 필요하다고 할 수 있다.

Table 7. The response rate of options of item 12 in CBT-P

답지 반응률 (%)	정답 (A,C)	두 검사에 공통으로 있는 선택지 조합				CBT-P형에만 있는 선택지 조합				무응답
		A,B	B,A	C,A	C,B	A,A	B,B	C,C	B,C	
CBT-M	35.92	15.97	15.76	25.00	7.14	-	-	-	-	0.21
CBT-P	37.25	14.57	7.69	13.36	5.67	2.02	1.62	0.40	17.00	0.40

나. 자료연결형: 끌어놓기

학업성취도 평가에서 도입한 컴퓨터 기반 평가의 새로운 문항 유형 중, 2021년 병행시행에서 활용된 자료연결형은 총 2개 문항이며, 모두 끌어놓기 기능을 적용하였다. 끌어놓기 기능은 텍스트 또는 이미지를 자료의 특정 위치에 배치시킴으로써 응답하는 방식으로, 자료들을 관련 과학 개념이나 사례에 따라 분류하고 대응시키도록 문항을 구성할 수 있다. Figure 8과 같이 구현되는 끌어놓기 기능은 끌어놓을 대상의 수나 끌어놓을 영역의 수를 자유롭게 설정할 수 있다는 점이 특징이다.

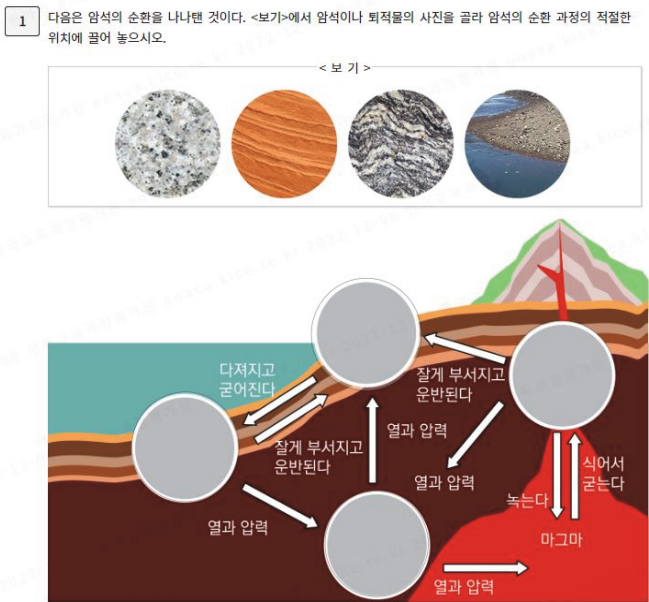


Figure 8. An example of specifying relationships item with drag-and-drop

2021년 학업성취도 평가 병행시행의 CBT-P형에서 사용한 2개의 끌어놓기 기능 문항은 선다형 문항을 변환한 것이다. Table 5의 6번과 18번 문항이 이에 해당하는 것으로, 6번 문항은 두 검사 간의 정답률의 편차가 1.36%p로 유사하게 나타났고, 18번 문항은 CBT-P형에서 정답률이 15.48%p 하락한 것으로 나타났다. 6번 문항은 혼합물 분리 과정을 자료로 제시하고 분리한 물질들을 각각 끌어놓아 응답하도록 구성하였는데, 3개의 물질을 3개의 영역에 끌어놓도록 제시하였다. 한편, 18번 문항은 특정 사례에서 일어난 물질의 상태 변화를 표현하기 위해 변화 전과 후의 입자 모형을 각각 끌어놓도록 구성한 것으로, 3가지 입자 모형 중 2개를 골라 끌어놓도록 제시하였다.

두 문항의 차이점은 끌어놓을 영역의 수로, 이것이 정답률에 일부 영향을 주었을 것으로 해석할 수 있다. 6번 문항에서는 학생들이 2개의 대상을 끌어놓게 되면, 1개의 대상은 자동적으로 위치가 지정이 되는 형태이기 때문에 특정 선택지의 진위 여부가 불분명해도 응답에 어려움이 없다. 반면, 18번 문항은 특정 선택지는 끌어놓을 수 없기 때문에 선택지의 배제가 불가능하다는 측면이 일부 영향을 미친 것으로 보인다. 18번 문항에서 정답과 반대되는 방향의 상태 변화로 응답한 학생의 비율이 13.77%인 점과, 끌어놓을 영역을 일부 채우지 않은 학생들의 비율이 18번 문항에서 소폭 높다는 점(2.42%p)은 이와 관련

이 있어 보인다.

이처럼 끌어놓기 기능과 같이 자료를 직접 연결시키도록 문항을 구성하는 것은 일종의 서답형과 비슷하게 활용할 수 있어 학생들의 정확한 이해를 파악할 수 있다는 점에서 의미가 있다. 이러한 결과는 앞서 살펴본 체크박스 기능을 활용한 11번 문항과 같이 교수학습을 위한 피드백 구성을 보다 세밀하게 할 수 있다는 이점이 있다. 다만, 체크박스 기능과 마찬가지로 과도하게 복잡한 형태로 자료 간의 연결을 요구하는 경우 검사 전반에 걸쳐 학생들에게 부담을 가중시킬 수 있다. 따라서, 이러한 유형의 비율은 검사 전반에서 일정 수를 넘지 않게 하는 등의 노력이 필요하다.

IV. 결론 및 제언

컴퓨터 기반 평가는 학생들이 평가 상황에서 접하는 평가의 매체가 변화함과 동시에, 문항 유형의 다각화를 위해 컴퓨터의 기술공학적인 기능을 도입할 수 있다는 측면에서 기존의 지필평가와 차별점을 갖는다. 이러한 변화는 기존의 지필평가보다 역량 평가의 측면에서 효과적이라는 강점과 더불어 학생들의 평가에 대한 참여도를 높일 수 있다는 장점을 가지고 있다. 반면, 매체가 변화하고 기존과 다른 방식과 형태의 문항들이 학생들에게 제시된다는 점은 평가의 타당성을 확보해야 하는 과제가 있기도 하다. 이러한 선상에서 본 연구는 컴퓨터 기반 평가 체제를 처음 도입한 학업성취도 평가의 결과를 바탕으로 컴퓨터 기반 평가 도입에 따른 학생 응답의 특성을 살펴보았다. 앞서 기술한 변화의 영향을 살펴보기 위해 2021년 학업성취도 평가의 병행시행 결과를 활용하여 기존의 지필평가와 비교하여 평가의 매체만 변환시킨 문항 유형, 컴퓨터 기반 평가의 특징을 적극적으로 반영한 문항 유형의 학생 반응 결과를 비교하고 분석하였다.

종이와 펜을 이용한 시험에서 컴퓨터를 이용한 시험으로 매체가 변화한 측면에서 보았을 때, 2021년 학업성취도 평가의 결과에서 비교하면 두 검사 간의 큰 차이는 없었다. 즉, 구인과 무관한 변인으로 매체 효과(mode effect)가 작용하지 않는다고 볼 수 있어 컴퓨터 기반 평가 도입에 따른 평가 결과 산출의 문제는 크지 않음을 확인할 수 있었다. 다만, 일부 문항에서 컴퓨터 기반 평가 도입에 따른 주목할 점을 발견할 수 있었다. 우선 검사 후반부에 배치된 서답형 문항의 정답률을 보았을 때, 정답률이 지필평가에 비해 컴퓨터 기반 평가에서 다소 높게 나타났다. 이는 컴퓨터 기반 평가는 학생들에게 응답의 편의성을 높여 평가에 대한 적극적 참여를 강화한 것과 연관된다고 볼 수 있다. 한편, 컴퓨터 기반 평가 도입에 따라 새로운 기능들이 적용된 문항들의 결과에 따르면, 일부 문항들에서 정답률의 차이가 10%p 이상 나타남을 확인할 수 있었다. 학생들의 답지 반응을 분석한 바에 따르면, 이는 새로운 기술공학적인 기능을 통해 마련한 혁신적인 문항들이 학생들의 이해 수준을 더 세밀하게 파악한 결과로 볼 수 있었다. 즉, 학생들의 이해를 보다 구체적으로 파악할 수 있다는 컴퓨터 기반 평가 도입의 강점을 확인할 수 있었다.

다만, 컴퓨터 기반 평가를 도입함에 있어 평가 도구와 문항 구성에 대한 몇 가지 시사점을 확인할 수 있었다. 지필평가와 CBT-M형의 결과를 비교한 바에 따르면, 문항에 포함되는 정보의 양과 길이가 과도하게 늘어나지 않도록 유의할 필요가 있다. 지면상에서는 학생들이 자유롭게 자신의 시선을 옮겨가며 문항의 정보들을 조합하고 활용

할 수 있지만, 컴퓨터 화면에서는 스크롤이 발생할 여지가 있어 한번에 확인할 수 있는 정보의 양이 제한된다. 이는 학생들이 정보를 처리하는 데 분절의 우려를 낳을 수 있다. 따라서, 문항에 포함된 정보의 양을 조절하거나, 자료와 질문의 배치를 효율적으로 처리할 필요가 있다. 또한 지필평가에서는 지면상에 정보를 기재하거나 표기가 자유롭게 가능하나 컴퓨터 기반 평가에서는 같은 수준에서 자유도 높은 정보의 기재가 어렵다는 점을 감안하면, 문항의 정보가 과도하게 증가하는 것은 경계할 필요가 있다.

컴퓨터 기반 평가에서 새로운 문항 유형을 도입할 때도 유의해야 하는 점들이 있다. 끌여넣기, 아래로 펼치기, 핫스팟 등의 새로운 기술 공학적 기능들은 일반적인 컴퓨터 활용 상황에서는 익숙할 수 있으나 평가 상황에서는 새로운 것이다. 즉, 이전에 경험하지 못한 방식들에 대해서는 구체적인 응답 방법이나 자료의 배열을 상세하게 안내해주는 것이 필요하다. 본 연구에서 탐색했던 CBT-P형의 문항 중 핫스팟 기능을 적용한 두 문항이 대표적인 사례로, 학생들에게 무엇을 선택해야 하는지 구체적으로 문항에 표기하는 것이 필요하다. 즉, 문항의 응답 방식 및 응답해야 하는 항목의 위치 등에 대하여 보다 상세하게 안내하는 접근이 필요하다.

한편, 이러한 유형들의 문항과 관련해 검사지 전반에 걸쳐 고려할 사항도 있다. 앞선 결과 중, 체크박스, 아래로 펼치기 기능 적용 문항에서 살펴볼 수 있듯이, 기존의 5지 선다형 문항들이 갖는 임의의 추측에 의한 응답은 새로운 기능을 통해 어느 정도 배제할 수 있었다. 하지만, 이와 같은 유형의 문항들은 상황에 따라 학생들에게 과도하게 많은 시간을 소요하게 할 가능성이 있어, 검사 전반에 걸쳐 그 수를 적절하게 조정하는 것이 필요하다. 학교 현장에서의 수업 상황에서 수업 시작이나 중간 시점에 한 문항씩 제시하는 경우에는 이러한 점을 고려할 필요가 없다. 하지만 총괄평가와 같은 상황에서는 제공된 시간 내에 학생들이 충분히 해결할 수 있는 분량으로 검사를 구성하는 것이 필요하다. 따라서 위와 같은 새로운 유형의 문항들이 학생들에게 과도하게 제시되지 않도록 검사를 구성할 때 유의해야 한다.

본 연구의 결과는 컴퓨터 기반 평가 체제를 도입한 대규모의 평가 결과를 토대로 평가 체제의 전환 과정에서 연구자와 교사들이 유의할 사항들을 짚어냈다는 점에서 의미가 있다. 다만, 추후 추가적인 연구와 사례의 조사가 필요하다. 우선 본 연구는 학업성취도 평가에서 새롭게 도입하고자 하는 문항 유형 중 일부 유형에 대한 학생들의 반응만을 분석하였다는 한계가 있다. 앞서 기술한 것과 같이 학업성취도 평가에서는 본 연구에서 분석한 문항 유형 외에 순서배열형, 정보활용형 등 보다 다양한 유형이 활용될 것이다. 연구에서는 2021년에 지필평가와 병행시행한 컴퓨터 기반 평가를 분석 대상으로 하여 일부 유형에 대한 학생 반응을 분석할 수 없었으나 향후 컴퓨터 기반 평가가 전면적으로 시행된 후, 새롭게 도입된 문항 유형에 대한 전반적인 분석이 필요하다. 그리고 본 연구에서는 중학교 대상으로 학생 반응을 분석하였으나, 교육부의 발표에 따르면 향후 초등학교에서도 자율평가 형태로 컴퓨터 기반 학업성취도 평가를 시행할 예정이다. 따라서 본격적으로 컴퓨터 기반 학업성취도 평가가 초등학교에서 시행되고 검사의 타당성이 확보되기 위해서는 초등학교 학생들을 대상으로 컴퓨터의 기술공학적 기능이 활용된 문항 유형에 대한 답지 반응과 어려움에 대해서 분석할 필요가 있다. 또한 학교 현장에서 개인

용 태블릿이나 컴퓨터의 보급이 점차 확대되고, 온라인 학습이 일상화되면서 컴퓨터를 활용한 평가의 기회도 점차 늘어나고 있다. 따라서 컴퓨터를 활용한 평가나 이에 대한 학생들의 반응 특성, 그리고 컴퓨터 기반 평가와 교수학습을 연계하는 방안이 체계적으로 연구될 필요가 있다.

국문요약

디지털 기반 지능 정보 사회로의 진입에 발맞추어 과학 교육과정에서는 과학과의 교과역량 함양을 강조하고 있으며, 역량 평가의 측면에서 컴퓨터 기반 평가가 관심을 받고 있다. 컴퓨터 기반 평가는 높은 실제성을 갖는 형태로 문항을 구현할 수 있고, 평가 결과를 데이터베이스로 축적하여 환류 체계를 구축함에 있어서도 이점이 있다. 다만, 평가 타당도 개선, 측정 효율성 저하, 관리 요소 증가 등의 문제를 해결할 필요가 있다. 본 연구에서는 학업성취도 평가가 지필평가에서 컴퓨터 기반 평가로 전환되는 과정에서 새로운 평가의 도입에 따른 학생들의 반응을 살펴보기 위해 2021년도에 시행된 학업성취도 평가의 지필평가와 컴퓨터 기반 평가의 병행시행 결과를 분석하였다. 특히, 동일한 문항을 평가 매체만을 변화시켰을 때 학생들의 성취에 미치는 영향, 컴퓨터 기반 평가의 장점을 살린 새로운 기능을 포함하여 문항을 구성했을 때의 변화가 학생들의 성취에 미치는 영향을 살펴보았다. 중학교 3학년 학생 7,137명이 지필평가와 2종의 컴퓨터 기반 평가 중 하나에 응시한 결과를 분석하였다. 평가 시행 후 집단별로 문항의 정답률과 변별도 평균을 산출하였으며, 학업성취도 출제 경험이 있는 과학교사 8명이 참여한 전문가 협의회를 통해 응답 특성에 대한 전문가 의견을 수렴하였다. 결과에 따르면 지필평가와 단순 모드 전환형 평가에서의 학생들의 성취 결과는 큰 차이는 없어 매체 효과가 거의 나타나지 않았다. 다만, 서답형 문항의 정답률이 컴퓨터 기반 평가에서 다소 높게 나타났음을 확인하였고, 이는 응답의 편이성과 관련되는 것으로 분석하였다. 한편, 컴퓨터 기반 평가 도입에 따라 새로운 기능들이 적용된 문항들에서 유사한 문항의 정답률과 차이가 10%p 이상인 문항들이 존재하였다. 학생들의 답지 반응을 분석한 바에 따르면, 이는 새로운 기능을 통해 마련한 혁신적인 문항들이 학생들의 이해 수준을 보다 세밀하게 파악한 결과로 볼 수 있었다. 결과를 토대로 컴퓨터 기반 평가를 도입하고 개발할 때 유의할 사항을 논의하고 시사점을 제시하였다.

주제어 : 국가수준 학업성취도 평가, 중학생, 학생 응답 특성, 컴퓨터 기반 평가, 매체 효과

References

- ACARA(Australian Curriculum, Assessment and Reporting Authority) (2017). NAP sample assessment science literacy 2015: Technical report. Retrieved December 04, 2022, from https://www.nap.edu.au/docs/default-source/default-document-library/20170317-nap-sl-tech-report.pdf?sfvrsn=c0316d5e_4.
- ACARA (2018). National Assessment Program - Literacy and Numeracy (NAPLAN) 2018: Technical report. Australian Curriculum, Assessment and Reporting Authority. Retrieved May 29, 2020, from https://www.nap.edu.au/docs/default-source/default-document-library/2018_naplan_technical_report_full_v1.pdf?sfvrsn=0.
- ACER(Australian Council for Educational Research) (2016). NAPLAN online readability and layout study - literature review (p. 74).

- Ban, J., Kim, S., Kim, S. J., Lee, S., Kwak, C., Lee, J., Park, H., Jeon, M., Shin, D., Kim, M., Jeon, J., Yeom, J., & Kim, J.(2021). 2020 research service report for developing and using of contents for 2020 diagnosis-supplement system for basic academic skills. Chungnam National University Applied Measurement and Evaluation Center.
- Ban, J., Kim, S., Kim, M., Kim, S. J., Kim, J., Lee, S., Lee, Y., Lee, J., Yeom, J., Jeon, M., Lee, Y. J., & Jin, J.(2022). 2021 research service report for developing and using of contents for 2021 diagnosis-supplement system for basic academic skills. Chungnam National University Applied Measurement and Evaluation Center.
- Bennett, R. E., Ward, W. C., Rock, D. A., & LaHart, C. (1990). Toward a framework for constructed-response items. ETS Research Report Series(RR-90-7). Retrieved May, 27, 2020, from <https://onlinelibrary.wiley.com/doi/epdf/10.1002/j.2333-8504.1990.tb01348.x>.
- Bryant, W. (2017). Developing a strategy for using technology-enhanced items in large-scale standardized tests. *Practical Assessment, Research, and Evaluation*, 22(1), 1-10.
- Choi, H. (2013a). The development of science item on the computer-based performance assessment: A experiment on constructing circuits with an ammeter and a voltmeter. *Journal of Science Education*, 37(2), 348-358.
- Choi, H. (2013b). Developing science items of computer-based problem solving assessment. *Korean Journal of Teacher Education*, 29(4), 255-276.
- Cotter, K. E., Centurino, V. A. S., & Mullis, I. V. S. (2020). Developing the TIMSS 2019 mathematics and science achievement instruments. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and Procedures: TIMSS 2019 Technical Report* (pp. 1.1-1.36). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education and Human Development, Boston College and International Association for the Evaluation of Educational Achievement (IEA).
- de Klerk, S. (2012). An overview of innovative computer-based testing. In T. J. H. M. Eggen, & B. P. Veldkamp (Eds.), *Psychometrics in Practice at RCEC* (pp. 137-150). Retrieved May 27, from https://ris.utwente.nl/ws/portalfiles/portal/5579570/Chapter_12.pdf.
- Fishbein, B., Martin, M. O., Mullis, I. V. S., & Foy, P. (2018). The TIMSS 2019 item equivalence study: Examining mode effects for computer-based assessment and implications for measuring trends. *Large-scale Assessments in Education*, 6(1), 1-23.
- Jewsbury, P., Finnegan, R., Xi, N., Jia, Y., Rust, K., Burg, S., Donahue, P., Mazzeo, J., Cramer, B., Lin, A., & Weil, N. (2020). 2017 NAEP transition to digitally based assessments in mathematics and reading at grades 4 and 8: Mode evaluation study. Retrieved January 11, 2021, from https://nces.ed.gov/nationsreportcard/subject/publications/main2020/pdf/transitional_whitepaper.pdf.
- Hardcastle, J., Herrmann-Abell, C. F., & DeBoer, G. E. (2017). Comparing student performance on paper-and-pencil and computer-based tests. Paper presented at the 2017 AERA Annual Meeting, San Antonio, TX. Retrieved January 14, 2021, from <https://files.eric.ed.gov/fulltext/ED574099.pdf>.
- Kim, H. K. (2021). The study of the mode effect between computer-based and paper-based science tests in TIMSS 2019. *Journal of the Korean Chemical Society*, 65(1), 48-57.
- Kim, K., Kim, D., Kim, W., Kim, M., Kim, J., Shin, D., Choi, I., Lee, I., Park, I., Choi, I., Shin, J., Han, J., Song, M., Han, K., Kim, H., & Park, G. (2013). Introduced plans for computer based national assessment of educational achievement (CRE 2013-5). Korea Institute for Curriculum and Evaluation.
- Kim, M. (2000). Strategies on screen design of learner-centered web-based instructional systems. *Journal of Educational Technology*, 16(4), 51-76.
- Kozma, R. (2009). Transforming education: Assessing and teaching 21st century skills – Assessment call to action. In F. Scheuermann, & J. Björnsson (Eds.), *The Transition to Computer-Based Assessment* (pp. 13-23). Luxembourg: Office for Official Publications of the European Communities.
- Lee, M., Dong, H., Park, I., Kim, W., Seo, M., Jung, H., Kim, K., Kang, M., Jang, E., Sung, K., Rim, H., Kim, S., Pae, J., Kim, S. Y., Lee, J., Park, J., Yang, K., Kang, T., Shin, Y., & Park, Y. (2017). A study on the enhancement of the NAEA system incorporating the aims of the revised 2015 national curriculum (CRE 2017-8). Korea Institute for Curriculum and Evaluation.
- Lee, C., Kim, H., Sang, K., & Choi, J. (2018). Comparison of item characteristics between computer-based assessment and paper-based assessment-focusing on mathematics and science subjects (ORM 2018-39-22). Korea Institute for Curriculum and Evaluation.
- Lee, J., Kim, J., Park, J., Sung, K., Lee, K., Lee, S., Jung, H., Choi, S., Kim, K., Ahn, Y., & Ha, M. (2020). Introducing the computer-based test for the national assessment of educational achievement: A developmental study for the eNAEA (RRE 2020-5). Korea Institute for Curriculum and Evaluation.
- Lee, J., Kim, Y., Kim, J., Nam, M., Park, J., Park, J. H., Sung, K., Lee, S., Jang, E., & Jung, H., (2021). Advancing the test tools for electronic national assessment of educational achievement (eNAEA) (RRE 2021-2). Korea Institute for Curriculum and Evaluation.
- McMillan, J. H. (2014). *Classroom assessment: Principles and practice for effective standards-based instruction* (6th Ed.). Boston, MA: Pearson.
- Ministry of Education(MOE) (2015). 2015 revised science curriculum. Ministry of Education 2015-74 [issue 9].
- Parshall, C. G., Davey, T., & Pashley, P. J. (2000). Innovative item types for computerized testing. In W. J. van der Linden & C. A. W. Glas (Eds.), *Computerized Adaptive Testing: Theory and Practice* (pp. 129-148). Netherlands: Kluwer Academic Publishers.
- Parshall, C. G., Harnes, J. C., Davey, T., & Pashley, P.J. (2010). Innovative items for computerized testing. In W. J. van der Linden & C. A. W. Glas(Eds.), *Elements of Adaptive Testing*(pp. 215-230). New York: Springer.
- Pommerich, M. (2004). Developing computerized versions of paper-and-pencil tests: Mode effects for passage-based Tests. *The Journal of Technology, Learning, and Assessment*, 2(6), 1-45.
- Ripley, M. (2009). Transformational computer-based testing. In F. Scheuermann, & J. Björnsson (Eds.), *The Transition to Computer-Based Assessment* (pp. 85-91). Luxembourg: Office for Official Publications of the European Communities.
- Scalise, K., & Gifford, B. (2006). Computer-based assessment in E-Learning: A framework for constructing “intermediate constraint” questions and tasks for technology platforms. *The Journal of Technology, Learning, and Assessment*, 4(6). Retrieved March 23, 2020, from <https://ejournals.bc.edu/index.php/jtla/article/view/1653/1495>.
- Sireci, S. G., & Zenisky, A. L. (2006). Innovative item formats in computer-based testing: In pursuit of improved construct representation. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of Test Development* (pp. 329-347). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Song, M., Kim, S., Ku, J., Rim, H., Park, H., Han, J., Son, S., & Yang, S. (2014). OECD programme for international students assessment: PISA 2015 field trial report (RRE 2014-4-3). Korea Institute for Curriculum and Evaluation.
- Strain-Seymour, E., Way, W. D., & Dolan, R. P. (2009). Strategies and processes for developing innovative items in large-scale assessments. pearson education, Inc. Retrieved March 29, 2020, from <https://images.pearsonassessments.com/images/tmrs/StrategiesandProcessesforDevelopingInnovativeItems.pdf>.
- Zenisky, A. L., & Sireci, S. G. (2002). Technological innovations in large-scale assessment. *Applied Measurement in Education*, 15(4), 337-362.

저자정보

백중호(한국교육과정평가원 부연구위원)
이재봉(한국교육과정평가원 연구위원)
구자옥(한국교육과정평가원 연구위원)