



AI 디지털교과서 학습맵에 기반한 정보 교과 진단평가 문항 개발

Development of Informatics Subject Diagnostic Assessment Questions Based on AI Digital Textbook Learning Map

이정숙[†] · 최현종^{††}Jeongsook Lee[†] · Hyunjong Choe^{††}

요약

본 연구는 중학교 정보 교과에서 학습 부진을 예방하기 위해 AI 디지털교과서 학습맵을 기반으로 진단평가 문항을 개발하는 것을 목표로 한다. 2022년 개정 교육과정에 따르면, 최소 성취 수준 미 도달, 학습 부진, 학습 속도 차이, 일시적 사고 등 다양한 학습자 상황에서 발생하는 학습 격차 문제를 예방하기 위해 성취 수준을 고려한 다양한 평가, 학습 과정에서의 평가, 상시 평가 등의 방안을 적극적으로 활용하도록 하고 있다. 본 연구에서는 이에 따라 개별 학습자의 학습 상태를 정확히 평가하고 학습 격차를 신속히 해소하기 위해 객관식 진단평가 문항을 개발하였다. 특히, 알고리즘과 프로그래밍 영역에서 누적된 학습 결손을 파악하기 위해 AI 디지털교과서 학습맵에서 초, 중, 고의 연계성을 가지는 내용에 중점을 두었다. 문헌 고찰, 전문가 검토, 파일럿 테스트를 통해 문항의 타당성과 신뢰성을 확보하였다. 연구 결과, 개발된 진단평가 문항은 학습자의 기본 지식과 기술을 효과적으로 평가하여 성취도가 낮은 학생을 조기에 식별하는 데 유용한 도구임을 확인하였다. 본 연구에서 개발한 진단 평가 문항의 신뢰도는 크론바흐 알파 계수 0.865로 높은 수준을 보였으며, 2차 전문가 검토 후 모든 문항의 내용타당도 비율(CVR)이 0.5 이상으로 나타나 내용타당도가 적절한 것으로 확인되었다. 이 진단평가를 활용하여 초등학교 교육과정에서의 알고리즘과 프로그래밍 영역의 학습 결손을 교사가 미리 파악하고 적절한 처방을 할 수 있게 됨으로써, 학습자들의 학업 성취도 향상에 기여할 것으로 기대된다.

주제어 AI 디지털교과서 학습맵, 학습 결손, 학습 부진 예방, 진단평가, 알고리즘과 프로그래밍 영역

ABSTRACT

This study aims to develop diagnostic assessment items based on the AI digital textbook learning map to prevent learning difficulties in middle school informatics subjects. According to the 2022 revised curriculum, various assessment methods, including achievement level-based evaluations, process evaluations, and regular assessments, should be actively utilized to prevent learning gaps arising from diverse learner situations such as underachievement, learning difficulties, differences in learning pace, and temporary cognitive issues. In this study, objective diagnostic assessment items were developed to accurately evaluate individual learners' learning status and quickly address learning gaps. Particularly, to identify cumulative learning deficits in the algorithms and programming area, the items focused on content that maintains continuity across elementary, middle, and high school levels in the AI digital textbook learning map. The validity and reliability of the items were ensured through literature review, expert evaluation, and pilot testing. The results confirmed that the developed diagnostic assessment items are useful tools for effectively evaluating learners' basic knowledge and skills, enabling early identification of low-achieving students. The reliability of the diagnostic assessment items developed in this study showed a high level with a Cronbach's alpha coefficient of 0.865, and after the second expert review, all items had a Content Validity Ratio (CVR) of 0.5 or higher, confirming appropriate content validity. It is expected that by using this diagnostic assessment, teachers can preemptively identify learning deficits in the algorithms and programming area from the elementary school curriculum and provide appropriate interventions, thereby contributing to the improvement of learners' academic achievement.

Keywords AI Digital Textbook Learning Map, Learning Deficit, Prevention of Learning Difficulties, Diagnostic Assessment, Algorithms and Programming Area

†정회원 한국교원대학교 컴퓨터교육과 박사과정
††정회원 한국교원대학교 컴퓨터교육과 교수
(교신저자)

논문투고 2024년 06월 15일
심사완료 2024년 08월 05일
게재확정 2024년 08월 07일
발행일자 2024년 08월 30일

1. 서론

1.1 연구의 필요성

교육부에서는 2022 개정 교육과정을 통해 초중등학교의 정보 교과 시수를 기존의 2배로 확대하였으며, 학생 별 디지털 교육 기회 차이를 최소화하기 위해 방학 중이나 방과 후에도 SW·AI(Artificial Intelligence, 인공지능, 이하 AI) 캠프 운영을 추진하고 있다. 또한 여러 교과를 학습하는 데 기반이 되는 기초 소양을 언어, 수리, 디지털 소양으로 강조하고 총론과 교과에 반영하였다[1, 2].

2022 개정 교육과정의 주요 변화 중 하나는 AI 디지털 교과서의 도입이다. 정보 교과는 수학, 영어와 함께 2025년부터 AI 디지털교과서가 우선 도입된다. 정보 교육이 이전 교육과정에 비해 강화되면서 정보 교과의 영역은 기존 4개 영역에서 ‘인공지능’을 추가하여 5개의 영역으로 확장되었다. 정보 교과는 초등학교 실과, 중학교 정보, 고등학교 정보 과목으로의 연계성을 가진다. 초등학교 실과의 ‘디지털 사회와 인공지능’ 영역에서 배우는 교과 내용에 대한 학습 결손이 있다면 중학교 정보 교과 학습에 영향을 미친다.

Table 1은 초등학교 실과의 2015 개정 교육과정과 2022 개정 교육과정을 비교하여 재구성한 내용이다. 2022 개정 교육과정에서는 ‘디지털 사회와 인공지능’ 영역이 새롭게 추가되었으며, 이 영역은 중학교 정보 교육을 위한 17차시에 해당한다. 이 영역의 핵심 목표는 더 명확한 디지털 정보를 찾아내고, 조합하고 평가하는 능력인 디지털 리터러시와 실생활에서 문제를 인식하고 해결하기 위해 활용되는 능력인 컴퓨팅 사고력을 강화하는 것이다[3, 4].

Table 1. Comparison of Elementary Technology / Home Economics curriculum

	2015 revised curriculum	2022 revised curriculum
Area (Home field)	·Human development and family ·Family life and safety	·Human development and leading life ·Living environment and sustainable choices
Area (Technology field)	·Resource management and self-reliance ·Technology system ·Take advantage of technology	·Technical problem solving and innovation ·Convergence with sustainable technology
Newly open	-	Digital society and artificial intelligence
Number of hours	17 hours of information education	More than 34 hours using classroom work and school free time Organizing / Operation
Number of achievement criteria	40	39

교육과정에 제시된 중학교 정보 교과의 목표는 컴퓨팅 사고력을 기반으로 인공지능을 포함하는 컴퓨팅 기술을 활용하여 미래 사회에서 다양한 문제를 발견하고 해결할 수

있는 기초적인 능력을 함양하는 것이다[3]. 본 연구는 정보 교과의 목표를 달성하기 위해 AI 디지털교과서 학습맵을 기반으로 교수 활동 시작 전에 학습자의 학습 결손을 파악하기 위한 진단평가 문항을 개발하였다.

진단평가를 통해 학습자의 선수학습 능력 진단과 사전 학습 능력을 진단할 수 있으며 이 정보를 바탕으로 교사는 적절한 교수 처방을 내릴 수 있다[5].

본 연구에서는 AI 디지털교과서 학습맵을 기반으로 한 진단평가 문항을 개발함으로써, 학습 내용에 대한 학생들의 이해도를 정확히 파악하고, 잠재적인 학습 결손을 조기에 식별하고자 한다. 이를 통해 학생들의 디지털 소양과 컴퓨팅 사고력의 기초를 더욱 탄탄히 다질 수 있을 것으로 기대된다. 특히, 알고리즘과 프로그래밍 영역에서의 학습 결손을 조기에 파악하고 개선할 수 있는 도구를 제공함으로써, 학생들의 정보 교과 학습 효과를 높이고 디지털 리터러시 격차를 줄이는 데 기여할 수 있을 것이다.

2. 이론적 배경

2.1 AI 디지털교과서 교육과정 표준체계(학습맵)

모두를 위한 맞춤형교육 실현이라는 비전을 가진 AI 디지털교과서는 학생 개인의 능력과 수준에 맞는 다양한 맞춤형 학습 기회를 지원하고자 인공지능을 포함한 지능 정보 기술을 활용하여 다양한 학습 자료 및 학습 지원 기능 등을 탑재한 소프트웨어이다. AI 디지털교과서 개발자들은 교육과정 표준체계(학습맵)를 활용하여 국가 교육과정과 연계되도록 교육 콘텐츠를 개발하고 있다. 학습맵이란 교육과정의 내용을 연관 관계 및 위계, 선후 관계로 연결한 체계로 학습자의 현재 학습 위치 정보와 학습경로 정보를 제공하는 데 활용되는 일종의 학습 설계도이다. 학습맵은 대주제, 중주제, 소주제로 나뉘어 각 주제 간의 연계성을 강조하며, 이를 통해 학생들은 체계적이고 일관된 학습 경험을 할 수 있게 된다[6, 7].

Table 2. Components of learning maps for AI digital textbooks

Classification	Explanation
A00	Major topics by grade (group)
B00	Mid-topics that subdivide major topics by grade (group)
C00	Sub-topics that subdivide major topics by grade (group) * Depending on the main topic, there may be no subtopics.
K	Keywords considered important in subtopic units

Table 2에서 볼 수 있듯이, AI 디지털교과서 학습맵은 대주제(A00)부터 시작하여 중주제(B00), 소주제(C00)로 세분화되는 구조로 되어 있다. 이러한 구조는 학습 내용의 위계와 연계성을 명확히 보여주어, 학습자가 전체적인 학습 구

조를 파악하는 데 도움을 준다. 특히, 키워드(K)는 각 소주제에서 중요하게 다뤄지는 개념을 나타내어 학습의 핵심을 쉽게 파악할 수 있게 한다. 이러한 체계적인 구조는 학습자의 체계적이고 효율적인 학습을 지원하며, 교사에게는 교육 내용의 구조화와 수업 설계에 유용한 틀을 제공한다[6].

2.2 알고리즘과 프로그래밍 영역

2015 개정 교육과정에서는 정보 교과의 역량은 컴퓨팅 사고력, 정보문화소양, 협력적 문제해결력은 2022 개정 교육과정에서 Figure 1에 나타난 것처럼 컴퓨팅 사고력, 디지털 문화 소양, 인공지능 소양으로 변화하였다.



Figure 1. 2022 Revised Curriculum Informatics subject competency

컴퓨팅 사고력은 두 개의 교육과정에서 공통으로 중요하게 다루는 정보 교과의 역량이며, 컴퓨팅 사고력의 하위 요소인 추상화 능력, 자동화 능력은 알고리즘과 프로그래밍 영역의 내용 체계 및 성취 기준에 포함된다.

김슬기(2024)는 2015 개정 교육과정과 2022 개정 교육과정의 분석을 통해 알고리즘과 프로그래밍 영역 단원이 정보 교과에서 강조하는 역량 중 하나인 컴퓨팅 사고력을 기르기 위한 핵심적인 부분임을 시사하였다[8]. 또한 김자미(2024)의 연구에서는 2024년 1월을 기준으로 최근 10여 년간 발표된 컴퓨팅 사고력 관련 연구는 500편 이상이었으며, 이중 프로그래밍 관련 연구가 200여 편에 이른다고 하였다[9]. 이러한 선행연구를 바탕으로 본 연구에서는 컴퓨팅 사고력을 대표하는 알고리즘과 프로그래밍 영역을 진단평가 개발 영역으로 설정하였다.

2.3 진단평가

앞서 살펴본 AI 디지털교과서 학습맵과 알고리즘 및 프로그래밍 영역의 중요성을 고려할 때, 이를 효과적으로 평가할 수 있는 진단평가의 개발이 필요하다.

박도순, 홍후조(2010)에 따르면 진단평가의 목적은 두 가지로 구분된다. 하나는 수업이 시작되기 전에 실시되는 예진적 활동이고, 다른 하나는 학습 실패의 교육 외적 원인을 알아보는 활동이다. 첫째, 학습의 예진적 활동에서는 가장 중요한 것은 학생들의 학습 결손 유무에 대한 정보이다. 학습 결손은 후속되는 학습을 저해하는, 이른바 누적적인 학습 실패의 악순환을 초래할 가능성이 높다. 둘째 학습장애의 요인을 파악하고, 적절한 교수법이나 교육적 대안을 제공하기 위한 것이다[10].

본 연구에서는 이러한 진단평가의 목적에 부합하도록, 초등학교 실과와 중학교 정보의 알고리즘과 프로그래밍 영

역을 연계한 학습 주제들에 대한 평가 문항을 개발하고자 한다. 개발된 문항을 학생들에게 적용한 후, 문항 분석을 통해 그 타당성을 검증할 것이다.

3. 연구 방법

3.1 연구 대상

한국교육학술정보원(Korea Education and Research Information Service. 이하, KERIS)의 주도로 2007년부터 초·중학생의 디지털 리터러시 수준을 측정하고 분석하여 교육적·정책적 시사점을 제시해 왔다. 최근 발표된 2022년 초·중학생 디지털 리터러시 수준 측정 연구의 주요 결과를 요약한 결과 보고서에 따르면 컴퓨팅 사고력의 자동화 점수가 4점 만점에 초등학생은 1.92점, 중학생은 1.82점으로 7개의 검사영역 중 가장 낮은 것으로 조사됐다[11].

2022년 국가 수준 초·중학생 디지털 리터러시 수준 측정 연구를 통해 학년이 올라갈수록 오히려 코딩에 대한 효능감이 낮아지는 것으로 나타났으며 중학교 1학년의 학생들이 중학교 3학년 학생들보다 정보 교과 수업을 들으면서 코딩에 대한 효능감을 더 높일 수 있었을 것으로 유추하였다[12]. 이에 따라 본 연구에서는 중학교 정보 교육과정 이전의 중학교 1학년 학생을 대상으로 개발된 진단평가 문항을 투입하여 타당도를 확보하고자 하였다.

본 연구의 표본 선정은 편의표본추출법을 사용하였다. 연구자의 접근성을 고려하여 연구 협조가 가능한 교사들이 근무하는 학교의 학생들을 대상으로 표본을 선정하였다. 이러한 방법은 시간과 비용의 제약 하에서 연구를 효율적으로 수행할 수 있게 해주지만, 표본의 대표성에 한계가 있을 수 있다. 파일럿 테스트에서는 충청남도 소재의 중학교 1개교, 전북특별자치도 소재의 중학교 2개교, 광주광역시에 있는 중학교 1개교의 총 4개교의 중학교 1학년 학생을 대상으로 연구를 진행하였으며 연구 대상의 특성은 Table 3과 같다.

Table 3. Characteristics of Research Subjects

Division		N	%
Gender	Male Student	64	50.8
	Female Student	62	49.2
Region	Jeonbuk-do	62	49.2
	Chungcheongnam-do	39	30.9
	Gwangju	25	19.9
Sum		126	100

3.2 연구 절차

본 연구는 Figure 2과 같은 절차로 진행되었다.

문헌과 교육과정 분석, AI 디지털교과서 학습맵을 바탕으로 평가 요소를 추출하여 1차 진단평가 문항을 개발하였다. 진단평가 문항은 교육과정의 성취 기준과 학습맵의 핵심 키워드를 바탕으로 개발하였으며, 본 연구에서는 채점의 객

관성과 신뢰성이 높고 채점과 통계에 유리한 선다형 형식으로 진단평가를 개발하였다. 다음은 성태제(2014a, 2014b)에서 다루고 있는 선다형 문항 제작의 원리이다[13, 14].

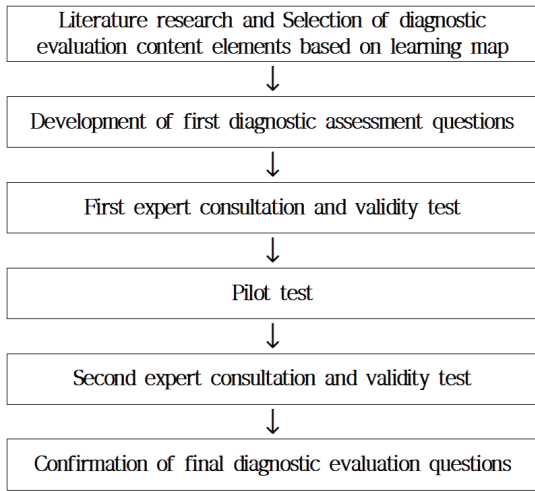


Figure 2. Diagnostic assessment item development procedure

진단평가 문항은 Table 4의 원칙에 따라 각 문항은 핵심 학습 내용을 포함하며, 간결하게 서술하였다. 긍정문을 우선하여 사용하였으며, 필요에 따라 부정문을 사용했을 때는 굵은 글씨와 밑줄을 사용하여 이를 명확하게 표시하였다. 학습자가 답지를 보고 정답을 단순히 추측하지 못하도록 매력적으로 오답지를 구성하였다. 답지의 길이는 일관성 있게 유지하였고, 유사한 내용의 답지는 인접하게 배열하였다. 또한, ‘모든 것이 정답’이나 ‘정답 없음’과 같은 답지는 사용하지 않았다.

Table 4. Principles of Developing Diagnostic Assessments

Principle	Description
Include Important Learning Content	Questions should include important learning content.
One Fact per Question	Each question should ask about one fact.
Definitions and Concepts	For questions about definitions or concepts, ask the definition or concept and list explanations in the answer choices.
Positive Statements	Questions should preferably be in positive form; if a negative form is necessary, underline or highlight it.
Plausible Interfering Factor	It creates a plausible and engaging distraction.
Best Answer	If there are multiple correct answers, prompt the selection of the best answer.
No Clues in Answer Choices	Do not provide clues in the answer choices that would allow students to select the correct answer or eliminate incorrect ones.
No Answer from Choices Alone	Students should not be able to derive the answer solely from analyzing the answer choices.

Principle	Description
Independence of Answer Choices	The content of answer choices should be mutually independent.
Avoid Repetition of Words	Avoid repeating the same words in each answer choice.
Similar Form of Answer Choices	Ensure that the forms of answer choices are similar.
Adjacent Similar Answer Choices	Place similar answer choices next to each other.

Figure 3은 ‘알고리즘과 프로그래밍 영역’의 핵심 키워드 중 하나인 ‘변수’에 대한 문항으로, 학습자들이 변수의 개념을 이해하고 있는지 확인하기 장난감의 이동을 통해 변수의 값을 추적하도록 설계되었다.

Q. Read the text in <Example> and choose the number of toys remaining in the <toy box>.

<Example>

Hwahee and her younger sister Manyeong are organizing the toys they played with in their room together into a "toy box." To check how many toys are left, Hwahee and Manyeong always record the number of toys in the box by attaching a note to the outside of the "toy box" when they take out or put a toy in the box. Currently the "toy box" contains 5 toys.

The input and output process of the toy:

1. Hwahee added 3 toys to the "toy box".
2. Manyeong took out two toys from the "toy box."
3. Hwa-ui and Man-young added four more toys to the "toy box."
4. Manyeong gave 1 toy to her friend.
5. Hwa-ui added two more toys.
6. Manyeong took out three toys from the "toy box."

Lastly, how many toys are left in the "toy box"?

① 6 ② 7 ③ 8 ④ 10

Figure 3. Diagnostic Assessment Item (Variable)

개발된 1차 22개 문항에 대한 내용타당도를 검증하기 위해 전문가 검토 설문지를 Figure 4와 같이 구성하였다.

Question number		Difficulty		□ upper □ middle □ lower
Middle topic		Minor topic		
Content elements		Keywords		
Achievement criteria				
Diagnostic evaluation question	Q. ① ② ③ ④			

Very Inappropriate	Somewhat Inappropriate	Average	Somewhat Appropriate	Very Appropriate
①	②	③	④	⑤

Question review opinion

* Please explain why you chose "very inappropriate" or "somewhat inappropriate."

Figure 4. Expert Review Questionnaire Structure

설문지의 상단 부분에는 AI 디지털교과서 학습맵의 코딩 정보와 성취 기준을 함께 표기하여 전문가들이 타당성을 검증할 때 필요한 정보를 제공하였다. 전문가들은 문항에 대한 난이도와 현장 적합도를 평가하였으며 검토 의견을 함께 제시하였다.

개발된 1차 문항들은 Table 5와 같이 구성된 컴퓨터교육과 교수 1인, 컴퓨터교육을 전공한 초, 중, 고 교사 22인으로부터 검토를 받았다.

Table 5. Expert group

Degree	Years of Teaching Experience	Occupation
Ph.D. in Computer Education	Over 20 years	Professor
ABD Ph.D. in Computer Education	10-20 years	Elementary School Teacher
ABD Ph.D. in Computer Education	10-20 years	Elementary School Teacher
M.Sc. in Computer Education	Over 20 years	High School Teacher
M.Sc. in Computer Education	Over 20 years	High School Teacher
M.Sc. in Computer Education	Over 20 years	High School Teacher
M.Sc. in Computer Education	Over 20 years	High School Teacher
M.Sc. in Computer Education	Over 20 years	Middle School Teacher
M.Sc. in Computer Education	Over 20 years	Middle School Teacher
M.Sc. in Computer Education	10-20 years	High School Teacher
M.Sc. in Computer Education	10-20 years	High School Teacher
M.Sc. in Computer Education	10-20 years	High School Teacher
M.Sc. in Computer Education	10-20 years	High School Teacher
M.Sc. in Computer Education	10-20 years	High School Teacher
M.Sc. in Computer Education	10-20 years	High School Teacher
M.Sc. in Computer Education	10-20 years	Middle School Teacher
M.Sc. in Computer Education	10-20 years	Middle School Teacher
M.Sc. in Computer Education	5-9 years	High School Teacher
M.Sc. in Computer Education	5-9 years	Middle School Teacher
M.Sc. in Computer Education	5-9 years	Elementary School Teacher
M.Sc. in Computer Education	5-9 years	Elementary School Teacher
B.Sc. in Computer Education	10-20 years	Middle School Teacher
B.Sc. in Computer Education	10-20 years	Middle School Teacher

중학교 정보 교과를 학습하지 않은 1학년 학생 126명을 대상으로 Table 6의 20개의 진단평가 문항을 투입하여 파일럿 테스트를 진행하였다.

Table 6. Diagnostic evaluation development question information

Contents Element	Middle Topic	Minor Topic	Number of questions
Algorithm	B02 Algorithm expression, Algorithm experience	C03 Understanding and expressing problem solving procedures	8
		C06 Understanding program structure	7
Programming	B04 Problem solving programming, program development	C07 Writing a problem solving program	3
		C09 Edit shared programs	2
Total			20

본 연구에서는 개발된 진단평가 문항의 질을 평가하기 위해 고전검사이론(Classical Test Theory, CTT)과 문항 반응이론(Item Response Theory, IRT)을 적용하여 분석을 실시하였다. 고전검사이론에 근거하여 문항의 난이도, 변별도, 그리고 검사의 신뢰도를 산출하였다. 문항 난이도는 각 문항에 대한 정답률을 계산하여 평가하였으며, 문항 변별도는 상위집단과 하위집단 간의 정답률 차이를 통해 분석하였다. 검사의 신뢰도는 Cronbach's α 계수를 산출하여 전체 검사의 내적 일관성을 확인하였다. 문항반응이론을 적용하여 문항특성곡선(Item Characteristic Curve, ICC)과 문항정보함수(Test Information Function, TIF)를 분석하였다. 문항특성곡선을 통해 각 문항의 난이도, 변별도, 추측도 등의 특성을 시각적으로 분석하였으며, 문항정보함수를 통해 각 문항이 제공하는 정보량을 능력 수준에 따라 분석하여 문항이 가장 정확하게 측정하는 능력 범위를 확인하였다. 이러한 종합적인 분석을 통해 각 문항의 특성과 전체 검사의 품질을 평가하였으며, 필요한 경우 문항의 수정 및 보완 작업을 수행하였다. 이를 통해 개발된 진단평가 문항의 타당성과 신뢰성을 확보하고자 하였다. 이후 2차 전문가 검토를 시행하였으며 전문가 집단에 속해있던 컴퓨터교육 박사과정에 재학 중인 초등학교 교사 2인과 중등교사 1인에게 검토를 받아 최종 20개 문항을 확정하였다.

4. 연구 결과

4.1 1차 전문가 타당도 검사

본 연구에서는 전문가의 의견을 수치화하여 문항의 타당도를 평가하기 위해 개발된 각 문항에 대해 문항별 난

이도와 현장 적합성 여부를 5점 척도로 평가하도록 하였다. Lawshe(1975)가 제시한 내용타당도 비율(Content Validity Ratio, 이하 CVR)을 구하여 타당도를 검증하였다. CVR은 응답자 수에 따라 최소값이 달라지며, 본 연구에서는 전문가 집단의 인원이 20명 이상이므로 CVR의 최소값은 0.42 이상일 때 내용타당도가 확보되었다고 판단한다[15]. 1차 전문가 타당도 검사의 CVR은 Table 7과 같다.

Table 7. Primary Content Validity Ratio (CVR) Values for Each Question

Item No.	CVR	Item No.	CVR
1	0.13	12	0.48
2	0.48	13	0.39
3	0.57	14	0.13
4	0.13	15	0.48
5	0.22	16	0.48
6	0.48	17	0.30
7	0.22	18	0.13
8	0.22	19	0.22
9	0.48	20	0.48
10	0.39	21	0.13
11	0.48	22	0.13

1차 전문가 타당도 검사 결과와 전문가 검토 의견을 종합적으로 분석한 결과, 개발된 22개의 진단평가 문항 중 상당수가 개선이 필요한 것으로 나타났다. CVR 분석 결과, 0.42 미만의 값을 보인 문항이 전체의 55%인 12개 문항으로 나타나 전반적인 문항 개선의 필요성이 제기되었다.

Table 8. Expert Review Comments and Revisions

Item No.	Review comments	Revisions
4	Synthesis is not taught in the elementary school curriculum. The question is too difficult.	Change after deleting a question
5	It takes too much time to solve the problem.	Change after deleting a question
7	It is difficult because the meaning of the problem is not clear.	Change after deleting a question
8	There are no conditions on the selection structure.	Add conditions to selection structure
14	The meal guide example was good, but there are too many variables to use.	Example change
17	I think there is no need for a waiting block.	Delete wait block
18	Since the flowchart is presented only as shapes without any content, it would be better if there were examples or explanations of the shapes inside.	Present flowchart using real-life examples
19	It is difficult to solve the problem because the direction of the object is not clearly indicated.	Show object orientation

Item No.	Review comments	Revisions
21	In the fields of algorithms and programming, evaluation of definitional areas such as sharing does not seem to be appropriate.	Delete question
22	The answer seems too easy because attitude and cooperation are assessments in the emotional realm.	Delete question

전문가 검토 의견을 세부적으로 살펴보면 Table 8에 나타난 것처럼 문항의 난이도에 대한 지적이 가장 많았다. 특히 4번, 5번, 7번 문항은 초등학교 교육과정과의 연계성 부족, 문제 해결에 걸리는 시간, 문제의 명확성 등의 이유로 삭제 또는 수정이 필요한 것으로 나타났다. 8번 문항은 선택 구조에 대한 조건 추가가, 14번 문항은 사용되는 변수의 수 조정이 필요하다는 의견이 제시되었다. 이외에도 알고리즘과 프로그래밍 영역의 학습 결손을 파악하는 데 있어 정의적 영역의 평가는 적절하지 않다는 지적이 있었다. 이러한 전문가 의견을 종합적으로 반영하여 문항 수정 작업을 진행하였다. 학습자들의 인지적 부담을 줄이기 위해 복잡한 지문은 간소화하고, 선택지의 모호한 표현을 제거하여 명확성을 높였으며, 정의적 영역을 평가하는 문항은 삭제하였다. 이러한 과정을 거쳐 알고리즘과 프로그래밍 영역의 학습 결손 파악에 초점을 맞춘 20개의 문항을 완성하였으며, 이 문항은 파일럿 테스트와 2차 전문가 검토를 진행하였다.

4.2 파일럿 테스트 성취수준별 학생 분포

본 연구에서는 교육부에서 실시하는 국가수준 학업성취도 평가의 성취수준 구분 방식을 적용하여 정보 교과와 파일럿 테스트 결과를 분석하였다. 이는 김경희 외(2011)의 연구에서 제시한 방법론을 따른 것으로, 학생들의 성취수준을 우수학력, 보통학력, 기초학력, 기초학력 미달의 4단계로 구분하였다[16]. 분석 결과, 파일럿 테스트에 참여한 중학생 126명의 성취수준 분포는 Table 9와 같이 나타났다. 우수학력에 해당하는 학생은 52명(41.27%), 보통학력은 25명(19.84%), 기초학력은 24명(19.05%), 그리고 기초학력 미달은 25명(19.84%)으로 확인되었다.

Table 9. Distribution of Students by Achievement Level in the Pilot Test

Achievement Level	Score Range	Number of Students	Percentage (%)
Advanced	80 and above	52	41.27
Proficient	60 to 79	25	19.84
Basic	40 to 59	24	19.05
Below Basic	Below 40	25	19.84
Total		126	100

평균 점수는 62.54점(표준편차 25.89)으로 나타났으며, 최저점은 15점, 최고점은 100점으로 학생들 간의 성취도 차이가 큰 것으로 분석되어 정보 교과에 대한 학생들의 이해도와 능력이 양극화되어 있음을 알 수 있다.

4.3 문항의 난이도와 변별도

문항의 난이도와 변별도는 고전검사이론(Classical Test Theory, CTT)과 측정하여 Table 10에 제시하였다. 고전검사이론에 근거하면 난이도는 문항의 쉽고 어려운 정도를 나타내는 지수로써 0에서 1까지이며, 진단평가의 난이도는 선행 기능 및 능력의 진단을 위해 대부분 쉬운 문항으로 0.60 이상에서 1.00 미만의 난이도가 적절하다[5, 10]. 본 연구에서 개발한 평가 문항의 난이도는 평균 0.60으로 적절하다고 볼 수 있다. 그러나 일부 문항의 난이도가 0.40 이하로 낮게 나타났는데, 이는 해당 영역에서 많은 학생들에게 학습 결손이 있음을 시사한다. 변별도 역시 평균 0.59로 대부분의 문항이 학생들의 능력 차이를 잘 구분해주는 문항으로 나타났으며 이는 Ebel(1965)이 제시한 문항변별도 평가 기준에 부합하는 결과이다[17]. 반면, 변별도가 낮은 문항들은 학습 결손 파악에 부적합하여 오답지의 매력도를 높이고, 실생활 맥락이나 학습자들에게 흥미 있는 소재로 문제를 재구성하였다. 학습자들의 수준에 맞는 적절한 단어를 사용하고 문제나 보기의 내용을 더 명확하고 간결하게 수정하였다. 또한 전문가 집단에 속해있는 컴퓨터교육 박사과정 3인에게 문항 검토를 의뢰하여 검토 의견을 문항에 반영하였다.

Table 10. Difficulty and discrimination of questions

Item No.	Contents Element	DI (p)	Discrimination (d)	Std. Dev.
1	Algorithm	0.87	0.14	0.33
2	Algorithm	0.90	0.23	0.31
3	Algorithm	0.56	0.56	0.50
4	Algorithm	0.72	0.65	0.45
5	Algorithm	0.59	0.59	0.49
6	Algorithm	0.67	0.62	0.47
7	Algorithm	0.58	0.72	0.50
8	Algorithm	0.70	0.60	0.46
9	Programming	0.66	0.67	0.48
10	Programming	0.67	0.80	0.47
11	Programming	0.69	0.73	0.46
12	Programming	0.59	0.64	0.49
13	Programming	0.40	0.70	0.49
14	Programming	0.39	0.44	0.49
15	Programming	0.52	0.69	0.50
16	Programming	0.56	0.58	0.50
17	Programming	0.56	0.69	0.50
18	Programming	0.61	0.72	0.49
19	Programming	0.57	0.79	0.50
20	Programming	0.33	0.39	0.47

4.4 문항의 신뢰도

문항의 내적 일관성 검증을 위해 신뢰도를 분석하였다.

Table 11. Item reliability analysis

CT Element	Alpha	Alpha deleting specific item
Algorithm	.707	.731
		.704
		.686
		.639
		.675
		.654
		.643
		.676
Programming	.819	.803
		.791
		.796
		.808
		.804
		.827
		.799
		.810
		.801
		.803
.792		
Total	.865	

주로 신뢰도 검사는 크론바흐 알파 계수(Cronbach's alpha)를 산출하여 판단하는데, 일반적으로 0.7 이상이면 신뢰도가 양호한 것으로 판단한다.

Table 11은 크론바흐 알파 계수를 산출한 결과로 전체 신뢰도는 0.865로 높게 나타났으며 알고리즘 영역은 0.707, 프로그래밍 영역은 0.819로 두 영역 모두 높은 신뢰도를 보였다.

4.4 문항특성곡선과 문항정보함수

문항반응이론(Item Response Theory, IRT)을 통해 파일럿 테스트 결과를 문항특성곡선(Item Characteristic Curve, ICC)과 문항정보함수(Test Information Function, TIF)를 이용하여 추가로 분석하였다.

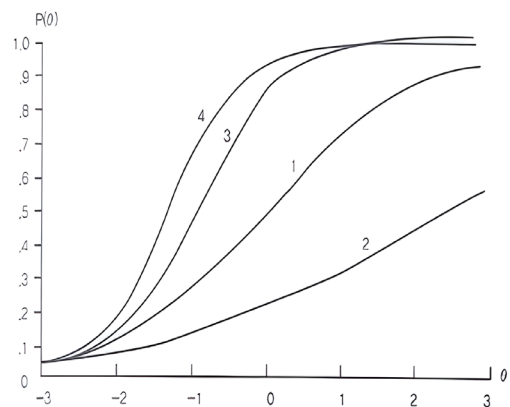


Figure 5. Item characteristic curves based on verbal expressions

문항특성곡선이란 준거 변수(criterion variable)와 한 문항에 답을 맞힐 확률과의 함수적 관계를 말한다. 문항반응이론에 의한 문항특성곡선에서 준거변수는 피험자가 가지고 있는 직접 관찰이 불가능한 잠재적 특성인 능력(ability)을 의미한다[17]. 문항특성곡선에는 문항난이도와 문항변별도의 두 가지의 기술 속성이 있다. Figure 5는 문항 난이도와 변별도의 언어적 표현에 따른 문항특성곡선이다. 1번 문항은 중간 수준의 난이도와 적절한 수준의 변별도를 지닌 문항이다. 2번 문항은 매우 어려우며 변별력이 낮은 문항이며, 3번 문항은 쉬우면서 변별력이 높은 문항이라 할 수 있다. 4번 문항은 매우 쉬우면서 변별력은 매우 높은 문항이다[18].

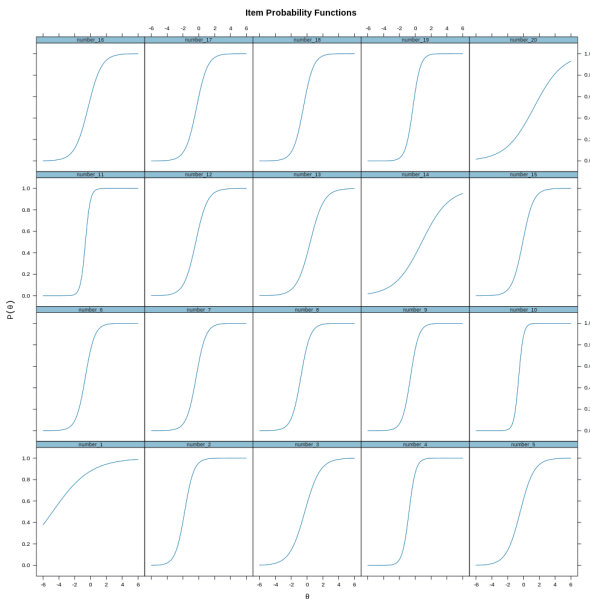


Figure 6. Item Characteristic Curve for Each Question

Figure 6은 파일럿 테스트를 통해 도출된 개별 문항에 대한 문항특성곡선이다. 진단평가는 난이도는 쉽고 변별도는 높아야 하므로 Figure 5의 4번 문항에 가까운 문항특성곡선의 형태를 취해야 한다.

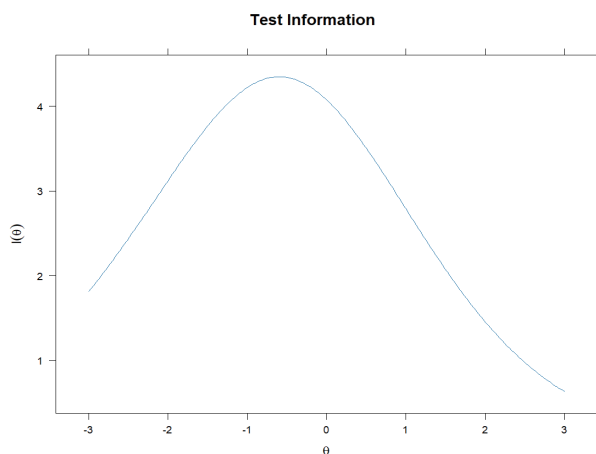


Figure 7. Test Information Function

Figure 7은 전체 문항에 대한 문항정보함수로 학습자의 능력 수준이 -1(약간 낮음)에서 0(평균) 사이일 때 약 4.5로 검사 정보가 가장 높게 나타난다. 이는 개발한 문항이 해당 범위의 학생들에 대한 능력을 가장 잘 측정한다고 해석할 수 있으며 해당 범위에서 특성 학습 결손을 신뢰성 있게 식별할 수 있음의 의미한다.

4.5 2차 전문가 타당도 검사 및 문항 확정

본 연구에서는 알고리즘과 프로그래밍 영역의 학습 결손을 진단하기 위한 평가 문항 개발을 위해 체계적인 과정을 거쳤다. 초기 개발된 22개 문항에 대해 1차 전문가 검토를 시행한 결과, CVR이 0.42 미만인 문항이 55%를 차지하여 대폭적인 수정이 필요함을 확인하였다. 이후 파일럿 테스트를 통해 문항의 난이도와 변별도를 분석하고, 그 결과를 바탕으로 문항을 수정 및 보완하였다. 2차 전문가 검토에서는 Table 12에 나타난 것처럼 모든 문항의 CVR이 0.5 이상으로 나타나, 1차 검토 결과와 비교하여 문항의 타당도가 크게 향상되었음을 확인하였다.

Table 12. Content Validity Ratio (CVR) Values for Each Question

Item No.	CVR	Item No.	CVR
1	0.6	11	0.6
2	0.6	12	0.6
3	0.8	13	0.8
4	0.6	14	0.6
5	0.6	15	0.6
6	0.6	16	0.6
7	0.6	17	0.5
8	0.8	18	0.7
9	0.6	19	0.7
10	0.6	20	0.6

최종적으로, 컴퓨터교육 박사과정에 재학 중인 초등학교 교사 2인과 중학교 교사 1인에게 추가 검토를 받아 문항의 내용, 난이도, 학습 목표와의 연계성을 재확인하였다. 이 과정에 중학교 1학년 학생들의 수준에 적합한지를 중점적으로 검토하고 20개의 최종 문항을 확정하였다. 이러한 다단계 검토 및 수정 과정을 통해, 본 연구에서 개발한 진단평가 문항은 내용타당도와 신뢰도를 확보하였다. 최종 확정된 20개 문항은 알고리즘과 프로그래밍 영역에서의 학습 결손을 효과적으로 진단할 수 있을 것으로 기대되며, 이는 향후 학생들의 개별화된 학습 지원 및 교육과정 개선에 유용한 정보를 제공할 수 있을 것이다.

5. 결론

본 연구는 2022 개정 교육과정의 AI 디지털교과서 학습맵을 기반으로 초등학교 실과 교육과정에서의 알고리즘과 프로그래밍 영역에 대한 학습 결손을 점검하기 위한 진단평가 문항을 개발하고 타당화하였다. 이를 위해 문헌 고찰, 교육과정과 학습맵 분석을 통해 알고리즘과 프로그래밍 영역의 핵심 키워드를 도출하고, 파일럿 테스트와 전문가 타당도 검증 과정을 거쳐 중학교 1학년 학생들을 대상으로 하는 진단평가 문항을 개발하였다. 본 연구에서 개발한 진단평가 문항의 양호도와 신뢰도는 고전검사이론과 문항반응이론을 통해 분석되었다. 고전검사이론에 따른 분석 결과, 문항의 평균 난이도는 0.60으로 나타나 진단평가로서 적절한 수준을 보였으며, 평균 변별도는 0.59로 대부분의 문항이 높은 수준의 변별력을 갖춘 것으로 확인되었다. 이는 Ebel(1965)의 문항변별도 평가 기준에 부합하는 결과이다[17]. 문항반응이론에 기반한 문항특성곡선(ICC) 분석 결과, 개발된 문항들은 대체로 진단평가의 목적에 부합하는 특성을 보였다. 특히, 문항정보함수(ITF) 분석을 통해 본 평가 문항들이 평균 능력 수준보다 약간 낮은 실력을 갖춘 학생들의 능력을 효과적으로 측정할 수 있음을 확인하였다[5]. 이는 학습 결손을 파악하고자 하는 진단평가의 목적에 잘 부합하는 결과라고 할 수 있다. 내용타당도 검증을 위해 실시한 전문가 평정 결과, 2차 전문가 검토에서 모든 문항의 CVR이 0.5 이상으로 나타났다. 이는 Lawshe(1975)가 제시한 기준값인 0.42를 상회하는 결과로, 개발된 문항들이 전문가들로부터 충분한 타당성을 인정받았음을 의미한다[15].

본 연구를 통해 개발된 진단평가 문항은 초등학교에서 중학교로 이어지는 정보 교육과정에서 알고리즘과 프로그래밍 영역의 학습 결손을 효과적으로 진단할 수 있을 것으로 기대된다. 이는 교사들이 학생들의 개별적인 학습 상태를 정확히 파악하고, 그에 따른 맞춤형 교육 전략을 수립하는 데 유용한 도구로 활용될 수 있을 것이다.

향후 연구에서는 본 진단평가 문항을 2022 개정 교육과정이 도입된 실제 교육 현장에 적용하여 그 효과성을 검증하고, 필요에 따라 지속적인 개선과 보완을 진행할 필요가 있다. 또한, AI 기술의 발전과 교육과정의 변화에 맞춰 진단평가 문항을 주기적으로 업데이트하고 확장하는 연구도 필요할 것이다. 이를 통해 변화하는 디지털 환경에서 학생들의 컴퓨팅 사고력과 프로그래밍 능력을 지속적으로 향상하는데 기여할 수 있을 것으로 기대된다.

참고문헌

- [1] Ministry of Education. (2021). *2022 Revised Curriculum Overview*. Sejong city. Ministry of Education.
- [2] Ministry of Education. (2022). *Basic Plan for Digital Talent Development in 2022*. Sejong city. Ministry of Education.
- [3] Ministry of Education. (2022). *Practical (Technology/Home Economics)/Information Department Curriculum*. Ministry of Education Notification No. 2022-33. Separate Volume 10.
- [4] Ministry of Education. (2015). *Practical (Technology, Home Economics)/Informatics Curriculum*. Ministry of Education Notice No. 2015-74, Separate Volume 10.
- [5] Seong, T. J. (2019). *Signature responsibilities basics of educational evaluation (3rd ed.)*. Hakjisa.
- [6] Korea Education and Research Information Service. *AI Digital Textbook Guidelines*. Korea Education and Research Information Service. (2024, May 16). Retrieved from <https://www.keris.or.kr/main/ad/pblcte/selectPblcteETCInfo.do?mi=1142&pblcteSeq=13722>
- [7] KERIS. (2023). *AI digital textbook development guidelines*. (2024, May 24). Retrieved from <https://www.keris.or.kr/main/ad/pblcte/selectPblcteETCInfo.do?mi=1142&pblcteSeq=13722>.
- [8] Kim, S. K. (2024). Analysis of Learning Problems in Middle School Informatics Textbooks: Focusing on Problem Solving and Programming Chapter. *Journal of Korean Association of Computer Education*, 27(1), 113-125. <https://doi.org/10.32431/kace.2024.27.1.009>.
- [9] Kim, J. M. (2024). Extracting the Hierarchical Learning Elements for the 'Algorithms and Programming' Area of the 2022 Revised Secondary Informatics Curriculum. *Journal of the Korea Internet Information Society*, 25(2), 123-132. <https://doi.org/10.7472/jksii.2024.25.2.123>.
- [10] Hong, D. S., Hong, H. J. (2010). *Curriculum & Educational Evaluation (3rd ed.)*. Muneumsa.
- [11] KERIS. (2022). *Status of digital education infrastructure and student digital capabilities in 2022*. Daegu. KERIS.
- [12] KERIS. (2022). *Research on measuring the digital literacy level of elementary and middle school students at the national level in 2022*. Daegu. KERIS.
- [13] Seong, T. J. (2014a). *Educational Evaluation Basics (2nd ed.)*. Seoul. Hakjisa.
- [14] Seong, T. J. (2014b). *Theory and Practice of Item Writing and Analysis*. Seoul. Hakjisa.
- [15] Lawshe, C. H. (1975). A Quantitative Approach to Content Validity. *Personnel Psychology*, 28(4), 563-575.
- [16] Kim, G. H., Kim, W. S., Choi, I. B., Sang, K. A., Kim, H. K., Shin, J. A., Kim, J. Y., & Son, W. S. (2011). *Analysis of the characteristics of Korea's academic achievement improvement as shown in the national level academic achievement evaluation*. Seoul. Korea Institute for Curriculum and Evaluation.

- [17] Robert L. Ebel. (1965). *Measuring Educational Achievement*. Englewood Cliffs, N. J. Prentice-Hall.
- [18] Seong, T. J. (2000). *Understanding and Application of Item Response Theory (2nd ed.)*. Education Science Press.
- [19] Baker, F. B. (2019). *Introduction item response theory* (Seong, T. J. Trans., 2nd ed.). Hakjisa.



이정숙

- 2013년 아주대학교 소프트웨어교육전공 (교육학석사)
- 2022년 ~ 현재 한국교원대학교 대학원 컴퓨터교육과 박사과정
- ✦ 관심분야 : 교육평가, 학습 부진, CT, EPL, SW/AI 교육
- ✉ jeongbo2019@knie.ac.kr



최현종

- 2005년 한국교원대학교 컴퓨터교육전공(교육학박사)
- 2006년 ~ 2021년 서원대학교 컴퓨터교육과 교수
- 2021년 ~ 현재 한국교원대학교 컴퓨터교육과 교수
- 2021년 ~ 현재 한국교원대학교 정보교육연구소 소장
- 2023년 ~ 현재 한국교원대학교 산학협력단 단장
- ✦ 관심분야 : 컴퓨터교육학, 정보교과교육, 인공지능교육 등
- ✉ chj@knie.ac.kr