



## Research Article

# The utility of digital evaluation based on automatic item generation in mathematics: Focusing on the CAFA system

Kim, Sungeun\*

Associate Professor, Incheon National University

\*Corresponding Author: Sungeun Kim (syk@inu.ac.kr)

## ABSTRACT

The purpose of this study is to specify the procedure for making item models based on ontology models using automatic item generation in the mathematics subject through the CAFA system, and to explore the generated item instances. As an illustration for this, an item model was designed as a part of formative assessment based on the content characteristics, including concepts and calculations, and process characteristics, including application, using the representative values and the measures of dispersion in Mathematics of the 9th grade based on the evaluation criteria achievement standards. The item types generated in one item model were a best answer type, a correct answer type, a combined-response type, an incomplete statement type, a negative type, a true-false type, and a matching type. It was found that HTML, Google Charts, TTS, figures, videos and so on can be used as media. The implications of the use of digital evaluation based on automatic item generation were suggested in the aspects of students, pre-service teachers, general teachers, and special education, and the limitations of this study and future research directions were presented.

**Key words:** Digital Assessment, Item Model, Mathematics, Automatic Item Generation, CAFA

## 수학교과에서 자동문항생성 기반의 디지털 평가 활용 방안: CAFA 시스템을 중심으로

김성연\*

인천대학교 부교수

\*교신저자: 김성연 (syk@inu.ac.kr)

## 초록

본 연구의 목적은 수학 교과에서 자동문항생성을 활용하여 지식의 핵심 구조인 온톨로지모형 기반의 문항모형을 CAFA 시스템을 통해 제작하는 절차를 명세하고, 생성된 문항 사례들을 탐색하는 데 있다. 이를 위한 사례로 수학 3의 대푯값과 산포도 단원의 평가준거 성취기준을 바탕으로 개념과 계산을 포함한 내용적 특성과 적용을 포함한 과정적 특성을 바탕으로 형성평가에 적합한 문항모형을 제작하였다. 하나의 문항모형에서 생성된 문항 유형은 최선답형, 정답형, 합답형, 미완성문장형, 부정형, 진위형, 배합형 등이었으며, 매체로는 Google Chart, HTML, TTS, 그림, 비디오 등을 활용할 수 있는 것으로 나타났다. 이를 바탕으로 자동문항생성 기반의 디지털 평가 활용방안에 대한 시사점을 학생, 예비교사, 일반교사 그리고 특수교육 측면에서 논의하고, 본 연구의 한계점과 향후 연구방향을 제시하였다.

**주요어:** 디지털 평가, 문항모형, 수학, 자동문항생성, CAFA

Received October 05, 2022

Revised October 16, 2022

Accepted October 21, 2022

2000 Mathematics Subject Classification : 97C40, 97U70

Copyright © 2022 The Korean Society of Mathematical Education.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 서론

2020년 코로나19가 전 세계적으로 확산되면서 우리의 삶은 언택트(untact) 방식으로 많은 변화가 일어났지만 현재 전 세계에서 코로나19 백신을 접종 중에 있으므로 머지않아 우리는 예전의 일상으로 돌아갈 수 있을지도 모른다. 그러나 사회 전반적으로 포스트 코로나시대가 코로나19 이전과는 같을 수 없으며 교육에 있어서도 상당한 변화가 예상된다(Kim, 2021). 이러한 환경의 변화와 함께 4차 산업혁명의 가속화 속에서 컴퓨터, 인터넷, 모바일 기술들의 성장을 통해 전 세계를 하나로 연결하는 초연결 사회가 도래하고 있다. 또한 각종 분석 기술들은 이러한 연결에서 생성되는 거대한 디지털 정보인 빅데이터를 실시간으로 처리 및 분석하는 수준으로 급격히 발달했다. 이러한 새로운 기술의 발전은 정치, 경제, 사회, 문화 전반에 걸쳐 디지털 전환(Digital Transformation)을 통해 광범위한 변화를 야기하고 있으며, 이는 교육 분야에서도 예외는 아니다(Choi, 2020). 교육 분야에서의 디지털 전환이란 인간의 지식을 아날로그 형태에서 디지털 콘텐츠와 서비스로 전환하고, 더 나아가 창의적인 아이디어, 표현, 주장, 공유, 학습을 통칭하는 디지털 자산으로 전환하는 것을 의미한다. 이 전환은 사람에 의해서 아날로그 방식으로 제작된 지식의 표상(Knowledge Representation)을 2차적으로 디지털화한 것에서부터 이러한 지식의 표상들을 기계가 생성(Machine Generation)하는 수준으로까지 진화하고 있다(Choi et al., 2022).

이미 세계 각국은 교육의 디지털 전환을 통해 직면한 여러 문제점들을 효율적으로 그리고 효과적으로 해결하고, 나아가 국가경쟁력을 갖추기 위해 전방위적으로 노력하고 있다. 이러한 노력 중에서 주목할 전략은 평가공학(Assessment Engineering)적 접근법이다. 평가공학은 아날로그 측정 및 평가를 위해 축적된 지식의 표상과 지식 처리 역량들을 인지과학 및 디지털 기술들과 융합하여 방대한 디지털 지식 기반 콘텐츠와 서비스의 기계 생성을 가능하게 하는 디지털 기술들과 사회적 절차를 다룬다. 즉, 평가공학은 과도하게 넘쳐나는 평가 및 학습분야의 디지털 데이터를 의미있고 유용한 디지털 정보로 바꾸고, 이를 다시 기계 생성과 효율적 관리가 가능한 가치를 창출하는 디지털 지식 표상들로 전환하는 학문이다. 또한 평가공학은 자동화(Automation)와 연결성(Connectivity)이 극대화된 평가 시스템을 목표로 기계 생성, 학습을 위한 평가, 학습으로서의 평가, 그리고 디지털 평가 플랫폼을 추구한다는 점에서 전통적인 평가 방법들과는 차이가 있다. 요컨대 현시점에서의 평가공학의 목표는 아날로그 평가의 디지털 평가로의 전환이라고 할 수 있다(Choi, 2019; Oh, 2022; Luecht, 2012; Mislevy et al., 2010).

디지털 평가에서는 평가 도구 및 평가 시스템의 제작, 검증, 사용, 분석 및 관리와 같은 평가의 전반적인 과정이 디지털 환경에서 실시간으로 분석 및 최적화되는 인터넷 상의 평가(Assessment on Internet)를 가정한다(Choi, 2019). 이는 기존의 개별적인 문항 제작과 지필평가를 기본으로 발전해오던 아날로그 평가 및 평가 시스템의 한계점과 부작용을 디지털 기술을 통해 해결하고자 하는 시도라고 볼 수 있다. 디지털 평가 전환을 위해서는 무엇보다 개별적인 문항이 아닌 하나의 모형(Model)을 통해서 다양한 유형의 문항들을 체계적으로 생성할 수 있는 컴퓨터 기반 문항모형(Item Model)으로의 전환이 필요하다. 그리고 이러한 문항모형으로의 전환에 대표적인 방법이 자동문항생성(Automatic Item Generation)이다.

자동문항생성이란 지식의 핵심 구조에 기반한 온톨로지모형과 기계 변형 알고리즘으로 구성된 문항모형을 바탕으로 컴퓨터가 임의로 규칙에 따라 서로 다른 문항 사례들을 자동적으로 만들어내는 시스템이다(Gierl et al., 2013). 자동문항생성은 타당한 문항 모형들이 충분히 제작된다면, 단시간에 많은 문제를 만들어 문제은행을 풍부하게 하고, 보안과 문제 관리에서 편리성을 가진다는 장점으로 인해(Oh et al., 2006), 단계별 학습 및 문제은행 확보가 중요한 수학 교육 분야에서 높은 관심을 받고 있다(Choi et al., 2012-2022; Embreston & Kingston, 2018; Fridenfalk, 2013; Gierl & Lai, 2013; Kang & Choi, 2020; Lim, 2017; Park, 2017; Schwinning et al., 2014; Xinxin, 2019).

수학분야에서 자동문항생성을 구현하기 위해 개발되어 있는 시스템으로는 CAFA(Computer Adaptive Formative Assessment, Choi et al., 2012-2022), HIM\_AIG(Xinxin, 2019), IGOR(Gierl & Lai, 2013), JACK(Schwinning et al., 2014), MathGen(Embreston & Kingston, 2018), 그리고 VMA(Fridenfalk, 2013) 등 다양하다. 이 중 CAFA 시스템은 문항모형을 기반으로 서버 프레임워크와 응용프로그램을 개발하기 위한 API(Application Programming Interface), 그리고 워드프레스 플러그인 기반의 기본 응용 프로그램을 제공하는 학습으로서의 디지털 평가 제작을 위해 특화된 시스템이다. 문항모형 제작을 위한 워드프레스 플러그인은 CAFA 시스템의 서버와 통신하는 API를 이용하여 구현되어 있으며, 평가자는 이를 이용하여 문항모형을 제작, 관리, 검증할 수 있다. CAFA 시스템의 문항모형

은 일반적인 문항과 마찬가지로 문두(Stem)와 선택지(Options)로 구성되어 있다. 그러나 일반 문항과 달리 하나의 문항모형으로부터 선수학습에 해당하는 전(Backward)과 후속학습에 해당하는 후(Forward)를 설계할 수 있는 연결모형(Mapped Models)과 선택지와 연동할 수 있는 피드백 슬롯(Feedback Slots)이 제시되어 있으며, 문항모형을 이루는 모든 부분을 지식의 핵심구조인 온톨로지모형에서 정의한 규칙과 매개변수에 따라 설정할 수 있다(Choi et al., 2012-2022). 구체적으로 온톨로지모형은 내용 전문가들이 측정하려는 지식을 이루는 핵심 요소 및 그 관계에 대해 합의한 바를 컴퓨터에서 조작할 수 있도록 전산적으로 도식화한 모형을 말한다(Choi, 2019). 또한 CAFA 시스템에서는 문항모형들을 결합하여 교육현장에서 학습이나 시험으로 활용할 수 있도록 평가모형을 구성할 수 있고, 이 평가모형에서 디지털 평가를 생성함으로써 자동화와 지능화를 통한 즉각적인 채점, 분석, 그리고 피드백을 제공하는 디지털 형태의 평가 서비스를 통한 교실 내 맞춤형 학습에 활용할 수 있다(Choi, 2022; Oh, 2022). 따라서 본 연구에서는 자동문항생성을 활용하여 온톨로지모형 기반의 문항모형을 CAFA 시스템을 통해 제작하는 절차를 명세하고, 생성된 문항 사례들을 제시함으로써 수학 교과에서 디지털 평가의 활용방안에 대한 시사점을 제시하고자 한다. 이를 위한 일례로 수학 3의 통계 단원 중 대푯값과 산포도 중단원의 내용을 활용하였다. 이 내용을 선정한 이유는 첫째, 우리나라 교육과정에서 제3차 교육과정 이래 취급되어 온 오래된 주제로 거의 모든 국가에서 통계학습 지도 내용으로 채택하고 있으며, 특별한 전문성 없이도 반드시 알아야 하는 개념이라고 인정되고 있기 때문이다(Lee, 2010). 또한 2015개정 교육과정에서도 수학 교과 내용 중에서 통계 영역만 실생활 중심의 통계 내용으로 재구성한다고 언급할 만큼 통계교육에 대한 중요성이 강조되고 있는 실정이다. 둘째, TIMSS와 PISA와 같은 국제 성취도 평가에서 우리나라 학생들의 성취도가 다른 수학 영역에 비해 통계 영역에서만 떨어졌기 때문이다(Kim & Kang, 2020; Park et al., 2015). 셋째, 통계용어에 대한 개념은 텍스트로, 계산은 수식으로, 그리고 데이터는 멀티미디어를 포함한 표, 그림, 그래프 등 다양한 문항 유형들을 하나의 온톨로지모형으로부터 구현할 수 있기 때문이다. 본 연구에서는 대푯값과 산포도 단원에 제시되어 있는 통계용어와 정의를 온톨로지모형으로 구성함으로써 질문하는 문항유형 자체는 다양하게 변화하지만 그 안에 내포되어 있는 지식의 핵심구조는 변화하지 않음을 확인할 수 있다. 구체적인 연구문제는 다음과 같다.

첫째, 대푯값과 산포도와 관련하여 다양한 유형의 문항을 생성하는 문항모형은 어떻게 제작하는가?

둘째, 대푯값과 산포도와 관련하여 문항 유형에 따라 생성된 문항 사례에는 어떤 것들이 있는가?

셋째, 대푯값과 산포도와 관련하여 매체 유형에 따라 생성된 문항 사례에는 어떤 것들이 있는가?

## 이론적 배경

### 디지털 평가

디지털 평가란 다양한 디지털 기술이 측정공학 기술들과 유기적으로 통합되어 평가의 본질적 목표인 학습 및 교육의 증진과 일체화되고, 그를 위한 효율성과 효과성이 극대화된 패러다임을 일컫는다(Choi, 2020). 그동안 개별적인 문항 제작과 지필평가 중심의 아날로그 평가가 갖는 한계점으로는 제작과정이 비과학적이고, 고비용이며, 느리며, 시행면에서 있어서는 시험 보안에 취약하고, 연속으로 시행하는 것이 어려우며, 전반적인 평가 시스템을 평가의 상위목적인 학습 및 교육에 최적화하기 힘들고, 지능화하기 어렵다는 지적이 제기되어 왔다(Boo & Vispoel, 2012; Carroll, 2021; Choi et al., 2022; Dardick & Choi, 2016). 이러한 아날로그 평가가 당면한 어려움을 직시하고, 디지털 기술을 통해 해결하기 위한 노력의 학문적 체계와 집합을 평가공학이라고 한다. 평가공학이 추구하는 목표는 기존의 지필평가, 서열척도 중심의 시험, 개별적 및 분산적 평가를 디지털 시대의 특징이라 할 수 있는 전산화, 자동화, 대응적 또는 개별적으로, 학습 및 학습자 중심으로 그리고 통합화 및 지능화된 디지털 평가로 전환하는데 있다(Choi, 2017, 2020).

디지털 평가 전환을 위해서는 첫째 문항에서 문항모형으로의 전환이 필요하다(Choi, 2017). 문항모형(Bejar, 2002; Bejar et al., 2002; LaDuca et al., 1986)은 문두, 선택지, 그리고 다른 보조 정보들을 포함하는 과제에서 변수를 명시적으로 표현하는 역할을 함으로써, 하나의 모형을 통해 많은 문항들을 체계적으로 생성할 수 있게 한다(Gierl et al., 2008). 이러한 문항모형으로의 전환에 대표적인 방법이 자동문항생성이다. 구체적으로 자동문항생성은 해당 영역의 전문가가 평가 영역을 설정하여 문항모형을 개발하고, 개발된 문항모형을 코딩하여 자동문항생성 알고리즘에 따라 문항을 생성한 후, 평가 시행을 통해 학습자의 반응에 따른 문항 정보를 도출 또

는 갱신하여 데이터베이스에 저장하게 된다(Choi, 2020; Oh, 2022). 문항이 자동문항생성과 같은 디지털 기술로 혁신되면 다른 평가 및 학습 프로세스들을 디지털 환경에서 기술적으로 통합하고, 효율성과 효과성을 높이며, 자동화 및 지능화하는데 도움을 주어 평가 과정 전반에서 디지털 혁신이 연쇄적으로 촉발되는 효과를 볼 수 있게 된다(Choi, 2017, 2020). 둘째, 검사에서 학습으로의 전환이 필요하다. 이는 평가가 평가자에서 학습자 중심으로의 전환을 의미한다. 즉, 평가자 중심의 총괄평가 기법들을 진단 및 형성평가 같은 학습분석 모형으로 통합함으로써 개인차를 반영한 다양한 필요를 충족하며, 개인의 유형 및 특성을 반영한 학습자 맞춤형 평가로의 전환을 의미한다. 예를 들어, 맞춤형 형성평가 및 교사권능평가(Teacher Empowered Assessment, Dardick & Choi, 2016)와 같은 첨단 기술을 이용한 교실 내 평가의 발전을 추구할 수 있다. 셋째, 서비스에서 플랫폼으로의 전환이 필요하다. 이는 개별적인 평가 서비스에서 통합적 평가 또는 학습분석 플랫폼으로의 전환을 의미하며 평가 및 학습 서비스 전체를 지능화하는 것을 목표로 한다. 이 플랫폼은 다양한 학습 서비스들이 서로 유기적으로 연결되어 평가 또는 학습분석 자원과 정보를 공유하게 하여 이들에 대한 중복투자 방지와 효율적인 서비스를 개발 및 관리할 수 있게 한다(Choi, 2017, 2020).

따라서 본 연구에서는 디지털 평가 전환의 성공여부가 달린 문항에서 문항모형으로의 전환과 서비스에서 플랫폼으로의 전환에서 제시된 중복 투자를 방지하기 위해 기존에 개발되어 있는 CAFA 시스템을 바탕으로 지식의 핵심구조인 온톨로지모형 기반의 문항모형으로부터 문항을 생성하는 구체적인 방법을 제시하고자 한다.

### 수학교과에서 자동문항생성을 활용한 연구

수학교과에서 자동문항생성을 적용한 연구들을 국내에서 바로 적용할 수 있는지 살펴보기 위해 시스템에서 한국어를 지원하고 있는지의 유무를 반영하여 국내와 국외로 구분하고, 적용된 대상의 범위와 활용된 수학 영역을 살펴보면 다음과 같다. 국내에서 Jeong 외 (2008)는 고등학교의 삼각함수 단원을 중심으로 SCORM 기반 반복학습 LMS (Learning Management System)에서 문항을 자동생성하여 문제은행을 쉽게 구현함으로써 형성평가에 활용할 수 있는 반복학습 콘텐츠를 개발하였다. 구체적으로 문항제작을 위한 반복학습 콘텐츠의 설계에서 주제 선정은 삼각함수 원리를 학습시키는 것을 목적으로 삼각함수의 정의역을 축소하여 제작하였다. Lim (2017)은 강한 인지이론 기반의 문항모형 설계를 사용하여 문자와 식의 인수분해 단원에 대한 인지모형을 바탕으로 자동문항생성의 문항모형을 개발하였다. 또한 학생들로부터 얻은 문항반응 결과를 바탕으로 문항특성함수, 난이도, 변별도 모수를 추정함으로써 수학과 평가에서 자동문항생성이 유용하게 쓰일 수 있음을 밝혔다. Choi 외 (2018)는 실제 고등학교 학생들의 평가시스템에서 자동문항생성의 유용성을 탐색하기 위해 CAFA와 IMS 시스템을 활용하여 교사들이 자신들에게 적합한 평가 도구를 만들 수 있도록 온라인 문항 전집에서 평가 문항들을 찾고 선택할 수 있도록 하였으며, 즉각적인 결과와 피드백을 학생들이 이해하기 쉬운 그래프나 차트로 제공할 수 있도록 하였다. Kang과 Choi (2020)는 중학교 1학년 1학기에 해당하는 수와 연산, 문자와 식, 그래프와 비례 관계 단원에 대해 수학 교과서 및 문제집에 수록된 문항들을 대상으로 문제, 정답, 해설에 대한 모델링을 수행하여 동적 지식베이스 기반의 자동문항생성시스템을 개발하였다. 구체적으로 쌍둥이 문항을 생성하는 ID 기반과 단원, 난이도, 문항종류, 문항유형, 보기, 부가계산의 6개 조건을 부여하여 문항을 생성하는 조건 기반으로 자동생성된 문항들을 제시하였다. Choi 외 (2022)는 CAFA 시스템을 활용한 모형 예제집에서 초등학교 수학부터 대학교의 통계에 이르기까지 다양한 문항모형들과 평가모형들을 소개하였다. 구체적으로 초등학교 4학년의 규칙찾기 단원과 5학년의 규칙과 대응 단원, 중학교1학년의 평면도형 단원, 중학교 3학년의 대푯값과 산포도 단원, 고등학교의 다항함수의 미분 단원, 그리고 대학교의 기초통계 개념을 이해하는 문항들을 제시하였다. 그러나 부정형과 합답형 문항 유형은 생성되지 않았으며, 중학교 3학년의 대푯값과 산포도 단원의 경우에도 대푯값 소단원에 한정된 문항들만 제시하였다.

국외에서는 수학 성취도 평가에서 문장제 문제를 포함한 몇 가지 문항 유형을 생성하는 연구들(Deane et al., 2006; Embreston & Kingston, 2018; Kellogg et al., 2015; Singley & Bennett, 2002)이 수행되었다. Gierl과 Lai (2013)는 IGOR을 통해 문항을 생성하는 과정을 자세히 설명하고 질적 검토는 생성된 수학 문항 단위가 아니라 문항모형을 제작하는 데 있어 논리와 내용에 초점을 맞추어야 한다고 제안하였다. 이에 Embreston과 Kingston (2018)은 자동문항생성으로 MathGen (Wilson et al., 2014)을 활용하여 수학 성취도를 측정하는 데 있어 6학년부터 8학년까지의 수, 대수, 기하 그리고 데이터 영역에서 생성한 문항들의 특성을 경험적 관점에서뿐만 아니라 질적 관점에서도 초점을 맞추어 분석한 결과 두 관점에서 모두 일반적으로 성공적이었다고 밝혔다. 그러나 어떻게 성공물을 높

일 수 있는지에 대한 피드백에서는 특정 주제에서 현실에서는 나타날 수 없는 음수와 같은 가격이 관측되었으므로 자동문항생성을 발전시키기 위해서는 값들이 추출되는 데이터베이스에서 적절한 비율, 속도, 가격 등이 추출될 수 있도록 제한을 두어야 한다고 제시하였다. 또한 이렇게 제한을 두고 분석한 결과는 성공적이었으며, 전체 문항에 대해 검토할 것이 아니라 생성된 일부 문항모형에 대해 검토를 하고, 검토 결과가 성공적이었다면 생성된 문항에 대해서는 전문가 한 사람이 내용과 형식에 대해 검토해도 충분하다고 밝혔다. 대학수준에서 Holling 외 (2009)는 대학교 1~2학년들을 대상으로 통계와 확률론과 관련하여 자동문항생성을 활용하였고, Sinharay와 Johnson (2012)은 문항의 복잡성에 따라 4단계로 구분하여 수학 문항을 생성하였다. 이러한 연구들은 평가와 관련하여 수업, 연습, 그리고 시험에 자동문항생성이 유용하게 활용될 수 있다고 밝혔다. 또한 Fridenfolk (2013)은 자동문항생성을 수행할 수 있도록 VMA를 개발하여 대학의 원격수업 중 이산수학교과에 적용하였다. Xinxin (2019)은 중등이후 교육에서 HIM\_AIG를 활용하여 자동문항생성을 수행하였으며, Schwinning 외 (2014)는 대학에서의 기초수학과 통계수업에서 웹기반의 튜터링 시스템인 JACK을 활용하여 자동문항생성을 수행하였다.

이상에서 살펴본 바와 같이 국내에서보다 국외에서 자동문항생성과 관련한 연구들이 많이 수행되고 있는 것으로 나타났으며, 학교급에 있어서도 중학교부터 대학교에 이르기까지, 수학 교과 내용에 있어서도 다양하게, 그리고 사용되는 시스템도 자체 개발하거나 기존의 것을 활용하는 등 여러 방면에서 연구가 수행되는 것으로 나타났다. 그러나 기존의 연구들은 자동문항생성을 기반으로 동일한 유형의 개별 문항들을 생성하는 연구들이 대부분이었으며, 문항 유형을 생성하는 경우에도 최선답형 및 정답형을 중심으로 한정적인 유형만 제시되었으며, CAFA를 활용한 기존 연구에서도 합답형이나 배합형 같은 문항유형은 제시되지 않은 것으로 나타났다. 또한 기하, 문자와 식, 수와 연산, 이산수학, 함수, 확률 및 통계 등 다양한 수학 분야에서 자동문항생성을 활용하고 있지만, 중학교 수준에서 생성된 문항 사례들은 대푯값에 한정되어 있는 실정이다. 따라서 본 연구에서는 지식의 핵심 구조를 반영하는 온톨로지모형을 바탕으로 형성평가에 활용할 수 있는 다양한 문항유형을 생성하는 문항모형을 제작하는데 특화되고, 한글을 지원하는 CAFA 시스템을 기반으로 산포도를 포함한 구체적인 문항모형 제작 절차와 생성된 문항들을 사례로 제시하고자 한다.

## 대푯값과 산포도 단원에서 자동문항생성 기반의 평가 개발 사례

### 문항모형 명세

CAFA 시스템에서는 수학의 전 영역에 걸쳐 문항모형을 생성할 수 있지만, 본 연구에서는 일례로 수학 3의 통계 단원 중 대푯값과 산포도 중단원의 내용을 활용하였다. 구체적인 문항모형의 명세는 Table 1과 같다. 이 문항모형은 2015개정 교육과정의 대푯값과 산포도 단원의 평가준거 성취기준을 바탕으로 개념과 계산을 포함한 내용적 특성과 적용을 포함한 과정적 특성을 바탕으로 중학교 3학년 학생을 대상으로 형성평가의 일부로 설계하였다. 즉, 여러 가지 대푯값과 산포도가 가지는 의미를 이해하는지, 주어진 자료의 특성에 따라 적절한 대푯값과 산포도를 선택하여 계산할 수 있는지, 그리고 어떤 상황에서 각 대푯값과 산포도가 유용하게 사용될 수 있는지를 평가하고자 하였다.

**Table 1.** The description of an item model

Generating Conditions	Contents
Subjects	9th grade students
Assessment Types	Formative assessment
Assessment Criteria Achievement Standard	Students can understand the meaning of <ul style="list-style-type: none"> <li>• median, mode, and mean, and be able to obtain them.</li> <li>• variance and standard deviation, and be able to obtain them.</li> </ul>
Learning Elements	Median, mode, representative values, standard deviation, variance
Assessment Goals	Students can select and obtain appropriate representative values and measures of dispersion depending on the characteristics of the data.
Assessment Domains	Concept, calculation, application
Item Varieties	<ul style="list-style-type: none"> <li>• Best answer, combined response, correct answer, incomplete statement, matching, negative, true false types</li> <li>• Google chart, HTML, Multimedia, TTS (Text To Speech)</li> </ul>

온톨로지모형 구성

대푯값과 산포도 단원의 핵심 내용인 온톨로지모형을 구성한 예시는 Figure 1과 같다. Figure 1의 온톨로지모형은 통계용어와 정의로 대푯값과 산포도 단원에서 평가하고자 하는 학습 요소와 개념을 추출 및 정리하였다. 이를 바탕으로 최선답형, 정답형, 합답형, 미완성문장형, 부정형을 포함한 선택형, 진위형, 배합형 등의 다양한 문항 유형들을 결합한 문항모형을 제작할 수 있다.

통계용어	정의
중앙값	자료를 작은 값부터 크기순으로 나열하였을 때 한 가운데 있는 값이다.
최빈값	자료의 값 중에서 가장 많이 나타나는 값이다.
평균	자료의 총합을 자료의 개수로 나눈 값이다.
분산	편차를 제곱한 값의 평균이다.
표준편차	분산의 음이 아닌 제곱근이다.
편차	각 자료의 값에서 평균을 뺀 값이다.

Figure 1. An ontology model of representative values and measures of dispersion

이러한 문항모형을 CAFA 시스템에서 구현할 수 있도록 코딩하는 예시는 Figure 2와 같다. 또한 지면 제약 상 이와 유사하게 활용 할 수 있는 매체에 따른 문항모형의 온톨로지모형에 대한 코딩 예시는 제시하지 않았으며, 일례로 매체 유형별 문항 코딩은 TTS만, 그리고 나머지 매체에 대해서는 생성된 문항 사례만을 제시하였다.

매개 변수	값	조건
@1@	중앙값최빈값평균분산표준편차편차	SHUFFLE
@2@	자료를 작은 값부터 크기순으로 나열하였을 때 한 가운데 있는 값이다.자료의 값 중에서 가장 많이 나타나는 값이다.자료의 총합을 자료의 개수로 나눈 값이다.편차를 제곱한 값의 평균이다.분산의 음이 아닌 제곱근이다.각 자료의 값에서 평균을 뺀 값이다.	LINK(1)
@3@	자료를 작은 값부터 크기순으로 나열하였을 때 한 가운데 있는 값이다.자료의 값 중에서 나타나는 값이다.자료의 총합을 자료의 개수로 나눈 값이다.편차를 제곱한 값의 평균이다.분산의 음이 아닌 제곱근이다.각 자료의 값에서 평균을 뺀 값이다.	LINK(1)
@4@	한 가운데가장 많이총합제곱한제곱근평균	LINK(1)
@6@	@1[1]@ @2[1]@ @1[2]@ @2[2]@ @1[3]@ @2[3]@ @1[4]@ @2[4]@ @1[5]@ @2[5]@ @1[6]@ @2[6]@	SHUFFLE
@7@	@1[1]@ @2[2]@ @1[2]@ @2[4]@ @1[3]@ @2[5]@ @1[4]@ @2[6]@ @1[5]@ @2[1]@ @1[6]@ @2[3]@	SHUFFLE
@8@	@ONEOF("a)(b)(c)(d)(e)", "12345", "디디디디", "히히히히")@	SELF
@9@	참거짓	SHUFFLE
@10@	@6[1]@ @7[1]@	LINK(9)
@11@	@1[1:5]@	SHUFFLE
@12@	@8@	LINK(11)
@13@	@8[1]@ @7[2]@   @8[2]@ @7[1]@   @8[3]@ @6[3]@   @8[4]@ @6[4]@   @8[1]@ @7[3]@   @8[2]@ @6[2]@   @8[3]@ @7[1]@   @8[4]@ @6[4]@   @8[1]@ @6[1]@   @8[2]@ @6[2]@   @8[3]@ @7[4]@   @8[4]@ @7[5]@   @8[1]@ @6[1]@   @8[2]@ @7[4]@   @8[3]@ @6[3]@   @8[4]@ @7[2]@   @8[1]@ @6[1]@   @8[2]@ @7[3]@   @8[3]@ @7[2]@   @8[4]@ @6[4]@   @8[1]@ @7[4]@   @8[2]@ @6[2]@   @8[3]@ @6[3]@   @8[4]@ @7[1]@   @8[1]@ @6[1]@   @8[2]@ @6[2]@   @8[3]@ @6[3]@   @8[4]@ @6[4]@   @8[1]@ @6[1]@   @8[2]@ @7[2]@   @8[3]@ @7[4]@   @8[4]@ @7[3]@   @8[1]@ @7[2]@   @8[2]@ @7[4]@   @8[3]@ @7[1]@   @8[4]@ @6[4]@   @8[1]@ @6[1]@   @8[2]@ @6[2]@   @8[3]@ @6[3]@   @8[4]@ @7[6]@   @6[1]@ @6[1]@   @6[2]@ @6[2]@   @6[3]@ @7[5]@   @6[4]@ @6[4]@	SHUFFLE
@14@	@8[3]@, @8[4]@   @8[4]@, @8[2]@, @8[3]@   모두 정답임. @8[1]@   @8[4]@   @8[1]@, @8[2]@, @8[3]@   @8[1]@, @8[2]@, @8[4]@	LINK(13)
@15@	@8[1]@, @8[2]@   @8[1]@, @8[3]@   @8[4]@, @8[3]@   @8[2]@, @8[4]@   @8[2]@, @8[3]@   @8[1]@, @8[4]@   @8[1]@, @8[2]@, @8[3]@   @8[3]@   @8[4]@   @8[3]@	LINK(13)

Figure 2. An example of coding about an ontology model of representative values and measures of dispersion

Figure 2에서 CAFA 시스템의 매개변수 @1@과 @2@는 Figure 1의 온톨로지모형에 제시된 통계용어와 정의를 나타내며, @1@의 조건으로 SHUFFLE 함수를 사용함으로써 통계용어를 무작위로 섞은 후 처음에 나타나는 값을 반환하며, @2@의 조건으로 LINK 함수를 사용함으로써 무작위로 섞인 @1@의 순서에 따라 @2@의 정의가 연결된다. @3@과 @4@는 통계용어를 정의하는 데 있어 핵심 정보를 빈 칸으로 나타내고, 이 빈 칸에 들어갈 단어를 각각 제시함으로써 미완성문장형 문항을 생성하는 데 활용된다. @6@과 @7@은 각각 통계용어와 정의에 대한 참 명제와 거짓 명제로써 SHUFFLE 함수를 사용하여 @6@과 @7@에 제시된 목록의 순서를 무작위로 섞어 그 중 첫 번째 값을 반환하며, @9@는 참과 거짓을 섞어 그 중 먼저 나타나는 값을 반환하며, @10@은 @9@에서 참인 경우는 @6@을, 그리고 거짓인 경우는 @7@을 연결하는 진위형 문항을 생성하는 데 활용된다. @11@은 배합형 문항을 생성하는 데 있어 문제에 제시될 통계용어를 나타내며, @12@와 @8@은 설명을 나타내는 답지의 번호를 알파벳, 숫자, 한글 등 다양하게 생성하는 데 활용된다. @13@-@15@는 합답형 및 부정형을 포함한 선택형 문항을 생성하기 위해 @13@의 첫 번째 목록인 @8[1]@ @7[2]@<br> @8[2]@ @7[1]@<br> @8[3]@ @6[3]@<br> @8[4]@ @7[5]@<br> @8[5]@ @6[5]@는 거짓, 거짓, 참, 거짓, 참을, @14@의 첫 번째 목록은 @13@에서 참을 나타내는 @8[3]@과 @8[5]@를, 그리고 @15@의 첫 번째 목록은 @13@에서 거짓을 나타내는 @8[1]@, @8[2]@, @8[4]@를 제시하였다. 통계용어의 목록을 무작위로 섞기 위해 @13@의 조건은 SHUFFLE 함수를, 그리고 이에 연결이 되도록 @14@와 @15@의 조건으로 LINK 함수를 사용하였다. 또한 추가할 내용이 있는 경우에는 @5@를 수정하여 사용할 수 있다.

**문항모형 제작 및 생성 문항 사례**

*문항 유형에 따라 생성된 문항 사례*

대푯값과 산포도의 내용적 특성 중 개념을 평가하기 위한 다양한 문항 유형을 CAFA 시스템에서 구현할 수 있도록 제작한 문항모형의 질문, 선택지, 그리고 정답은 Figure 3과 같다. 미완성문장형, 정답형, 최선답형, 진위형, 배합형, 합답형, 부정형 문항 유형의 질문은 모두 문항모형의 문두에 포함시켰다. @24@는 @SEQ(23, 16, -1)@로 @16@부터 @23@에 제시되어 있는 모든 문항 유형을 의미하며, @@24@[1]@은 이들 중 임의의 한 개의 문항 유형을 제시하게 된다. 또한 선택지에는 답지를 제시하며, 본 연구에서는 답지를 5개로 선정하였고 @@24@[2]@는 정답인 A를, 그리고 B부터 E는 각각 @@24@[3]@-@@24@[6]@인 오답을 제시하였다.

문두	선택지		정답
@24@[1]	A	@24@[2]	A
	B	@24@[3]	
	C	@24@[4]	
	D	@24@[5]	
	E	@24@[6]	

**Figure 3.** An example of coding about questions, stems, options, and answers

구체적으로 각 문항 유형의 질문들을 생성하는 문항모형의 코딩 예시와 생성된 문항 사례는 Figure 4-Figure 6과 같다. Figure 4의 @16@은 미완성문장형으로 Figure 2의 @1@에서 임의로 추출된 통계용어와 통계용어의 의미를 설명하는 데 있어 중요한 단어를 공백으로 둔 @3@을 연결함으로써 빈 칸에 들어갈 단어를 묻는 문항이다. @4[1]@은 @1@에서 임의로 추출된 통계용어와 연결이 되는 정답을 의미하며, @DISTRACTOR( "@4[2:6]@정확하게처음에마지막에", "@4[1]@" )@는 오답지로 정답인 @4[1]@을 제외한 나머지 답지들과 추가한 단어들로 후보를 정하여 이들 중 default 값에 따라 임의의 4개로 구성한다.

매개변수	값	문항 사례
@16@	다음 빈 칸에 들어갈 말은?                       @1[1]@: @3[1]@ @4[1]@  @DISTRACTOR("   @4[2:6]@  정확하게 처음에 마지막에",   "@4[1]@")@	<b>Item instance with seed 16 [ ]</b> <b>Stem:</b> 다음 빈 칸에 들어갈 말은? 최빈값: 자료의 값 중에서 _____ 나타나는 값이다. A. 가장 많이 B. 가장 적게 C. 마지막에 D. 중간에 E. 처음에 <b>Answer Key: A</b> <b>Feedback</b>
@17@	다음 중 용어와 설명이 올바른   것은?                      @1[1]@: @2[1]@  @1[2]@: @2[3]@  @1[3]@: @2[4]@  @1[4]@: @2[5]@  @1[5]@: @2[2]@	<b>Item instance with seed 2 [ ]</b> <b>Stem:</b> 다음 중 용어와 설명이 올바른 것은? A. 평균: 자료의 총합을 자료의 개수로 나눈 값이다. B. 중앙값: 편차를 제곱한 값의 평균이다. C. 분산: 자료의 값 중에서 가장 많이 나타나는 값이다. D. 최빈값: 각 자료의 값에서 평균을 낸 값이다. E. 편차: 자료를 작은 값부터 크기순으로 나열하였을 때, 한 가운데 있는 값이다. <b>Answer Key: A</b> <b>Feedback</b>
@18@	다음 설명을 나타내는 용어는?   @2[1]@  @1[1]@  @DISTRACTOR("   @1[2:6]@ 최댓값 최솟값",   "@1[1]@")@	<b>Item instance with seed 24 [ ]</b> <b>Stem:</b> 다음 설명을 나타내는 용어는? 자료를 작은 값부터 크기순으로 나열하였을 때, 한 가운데 있는 값이 A. 중앙값 B. 편차 C. 평균 D. 최댓값 E. 표준편차 <b>Answer Key: A</b> <b>Feedback</b>
@19@	다음 용어를 가장 잘 설명하는   것은?                      @1[1]@  @2[1]@  @DISTRACTOR("@2[2:6]@",   "@2[1]@")@	<b>Item instance with seed 13 [ ]</b> <b>Stem:</b> 다음 용어를 가장 잘 설명하는 것은? 표준편차 A. 분산의 음이 아닌 제곱근이다. B. 각 자료의 값에서 평균을 낸 값이다. C. 자료의 총합을 자료의 개수로 나눈 값이다. D. 자료를 작은 값부터 크기순으로 나열하였을 때, 한 가운데 있는 값이다. E. 자료의 값 중에서 가장 많이 나타나는 값이다. <b>Answer Key: A</b> <b>Feedback</b>

Figure 4. An example of item coding and instances about an incomplete statement type, a correct answer type, and a best answer type

@17@은 정답형으로 @1@에서 임의로 추출된 통계용어 @1[1]@의 설명이 올바른 것을 묻는 문항이며, 추출된 통계용어의 의미를 설명하는 정의로 구성된 @2[1]@이 연결되는 경우는 정답으로, 그리고 나머지는 오답으로 구성하게 된다. 문항 사례에서 답지 B는 오답으로 @1@에서 임의로 추출된 통계용어가 두 번째 @1[2]@와 연결이 되는 정의로 @2[2]@가 아닌 @2[3]@이 제시되게 된다. 나머지 경우도 순서를 나타내는 번호들이 다르게 연결됨으로써 오답을 나타낸다. @18@도 정답형으로 @2@에서 임의로 추출된 통계용어에 대한 정의인 @2[1]@을 나타내는 통계용어를 묻는 문항이며, 정의에 해당하는 통계용어에 해당하는 @1[1]@은 정답을 나타낸다. 오답지는 @DISTRACTOR("@1[2:6]@최댓값|최솟값", "@1[1]@")@를 활용하여 정답인 @1[1]@을 제외한 나머지 답지들과 추가한 단어들로 후보를 정하여 이들 중 default 값에 따라 임의의 4개로 구성한다. @19@는 최선답형으로 @1@에서 임의로 추출된 통계용어 @1[1]@의 설명으로 가장 올바른 것을 묻는 문항이며, 추출된 통계용어의 의미를 설명하는 정의로 구성된 @2[1]@이 연결되는 경우는 정답으로, 그리고 @DISTRACTOR("@2[2:6]@", "@2[1]@")@는 오답지로 @2@에서 정답인 첫 번째로 제시된 @2[1]@을 제외한 나머지 4개로 구성된다.

Figure 5는 진위형과 배합형 문항을 생성하는 문항모형의 코딩 예시와 생성된 문항 사례를 나타낸다. 구체적으로 @20@은 진위형으로 @6@의 통계용어와 정의가 올바르게 연결되는 참인 명제와 @7@의 통계용어와 정의가 올바르게 연결되지 않는 거짓인 명제를 묻는 문항이며, @21@는 배합형으로 문제군은 @11@로 @1@에서 임의로 추출한 4개의 통계용어를 제시하였으며, 답지군은 @12@로 @11@에 제시되어 있는 5개의 통계용어에 연결이 되는 설명을 제시하였다. 또한 @21@의 중간에 제시되어 있는 @8@은 답지군인 설명의 답지 번호들을 (a)~(e), ①~⑤, ㉠~㉡, 또는 ㉠~㉡ 중 하나로 통일하여 제시하게 된다. 정답은 문제군에 제시된 순서에 맞추어 @11@에 링크가 되어 있는 @12[1]@, @12[2]@, @12[3]@, @12[4]@로, 그리고 오답지는 정답의 순서가 아닌 경우를 제시하였다. 문항 사례에서는 정답으로 A에 ④, ③, ①, ②가 제시되었으며, 오답인 B의 경우는 A에서 각각 2번째, 1번째, 3번째, 4번째를 나타내는 @12[2]@, @12[1]@, @12[3]@, @12[4]@로 ③, ④, ①, ②가 제시되었다.



매개변수	값	문항 사례
@20@	<pre> @10[1]@&lt;br&gt; 다음은 참인가 거짓인가? @9[1]@ @9[2]@ Dummy Distractor 1  Dummy Distractor 2  Dummy Distractor 3                     </pre>	<p><b>Item instance with seed 4 [ ]</b>  <b>Stem:</b> 평균: 자료의 총합을 자료의 개수로 나눈 값이다. 다음은 참인가 거짓인가?  A. 참  B. 거짓  C. Dummy Distractor 1  D. Dummy Distractor 2  E. Dummy Distractor 3  <b>Answer Key:</b> A  <b>Feedback</b></p>
@21@	<pre> 다음 용어와 설명이 모두 올바르게 연결된 것은? &lt;table&gt;&lt;tr&gt;&lt;th align = left&gt; 용어&lt;/th&gt; &lt;th align = "left"&gt; 설명 &lt;/th&gt; &lt;/tr&gt;&lt;tr&gt;&lt;td width = "100px"&gt; @11[1]@ &lt;br/&gt; @11[2]@&lt;br/&gt; @11[3]@&lt;br/&gt; &lt;td&gt; @8[1]@ @2[1]@&lt;br/&gt; @11[4]@&lt;/td&gt; @8[2]@ @2[2]@&lt;br/&gt; @8[3]@ @2[3]@ &lt;br/&gt; @8[4]@ @2[4]@ &lt;br/&gt;@8[5]@ @2[5]@ &lt;/td&gt; &lt;/tr&gt;&lt;/table&gt;  @12[1]@, @12[2]@, @12[3]@, @12[4]@ @12[2]@, @12[1]@, @12[3]@, @12[4]@ @12[3]@, @12[2]@, @12[1]@, @12[4]@ @12[3]@, @12[1]@, @12[4]@, @12[2]@ @12[1]@, @12[4]@, @12[3]@, @12[2]@                     </pre>	<p><b>Item instance with seed 6 [ ]</b>  <b>Stem:</b> 다음 용어와 설명이 모두 올바르게 연결된 것은?  용어 설명  평균 ① 분산의 용이 아닌 제공근이다.  분산 ② 각 자료의 값에서 평균을 뺀 값이다.  표준편차 ③ 편차를 제공한 값의 평균이다.  편차 ④ 자료의 총합을 자료의 개수로 나눈 값이다.  ⑤ 자료를 작은 값부터 크기순으로 나열하였을 때, 한 가  운데 있는 값이다.  A. ①, ③, ⑤  B. ①, ④, ⑤  C. ①, ③, ④  D. ①, ②, ③  E. ③, ②, ③  <b>Answer Key:</b> A  <b>Feedback</b></p>

Figure 5. An example of item coding and instances about a true-false type and a matching type

Figure 6은 합답형과 부정형 문항을 생성하는 문항모형의 코딩 예시와 생성된 문항 사례를 나타낸다. @22@와 @23@은 각각 정답과 오답을 모두 찾는 합답형 문항으로, 본 절에서는 오답을 찾는 합답형 문항을 부정형의 예시로 제시하였다. 합답형과 부정형의 질문에 포함된 보기는 4개로 구성되어 있으며, Figure 2에 제시되어 있는 전체 참, 거짓을 나타내는 @13@에서 무작위로 섞인 후 처음에 나타나는 4개의 값이 반환된다. @22@는 Figure 2에 제시되어 있는 참을 나타내는 @14@에 해당하는 목록을 선택하는 합답형 문항을, 이와 유사하게 @23@은 Figure 2에 제시되어 있는 거짓을 나타내는 @15@에 해당하는 목록을 선택하는 부정형 문항을 생성한다.

매개변수	값	문항 사례
@22@	<pre> &lt;style&gt; fieldsetroundedCorners {border: 1px solid Black; border-radius: 10px; box-shadow: 2px 2px; padding: 5px;} &lt;/style&gt; &lt;fieldset class="roundedCorners"&gt; &lt;legend&gt;보기&lt;/legend&gt; @13[1]@ &lt;/fieldset&gt; 위의 보기 중 맞는 것을 모두 고른 것은? @14[1]@ @14[2]@ @14[3]@ @14[4]@ @14[5]@                     </pre>	<p><b>Item instance with seed 108 [ ]</b>  <b>Stem:</b> 보기  ① 표준편차: 분산의 용이 아닌 제공근이다.  ② 평균: 자료의 총합을 자료의 개수로 나눈 값이다.  ③ 분산: 편차를 제공한 값의 평균이다.  ④ 중앙값: 자료를 작은 값부터 크기순으로 나열하였을 때, 한 가운데 있는 값이다.  위의 보기 중 맞는 것을 모두 고른 것은?  A. 모두 정답임  B. ①, ④  C. ①, ②, ④  D. ①, ②  E. ③, ②  <b>Answer Key:</b> A  <b>Feedback</b></p>
@23@	<pre> &lt;style&gt; fieldsetroundedCorners {border: 1px solid Black; border-radius: 10px; box-shadow: 2px 2px; padding: 5px;} &lt;/style&gt; &lt;fieldset class="roundedCorners"&gt; &lt;legend&gt;보기&lt;/legend&gt; @13[1]@ &lt;/fieldset&gt; 위의 보기 중 틀린 것을 모두 고른 것은? @15[1]@ @15[2]@ @15[3]@ @15[4]@ @15[5]@                     </pre>	<p><b>Item instance with seed 112 [ ]</b>  <b>Stem:</b> 보기  ① 분산: 편차를 제공한 값의 평균이다.  ② 평균: 자료를 작은 값부터 크기순으로 나열하였을 때, 한 가운데 있는 값이다.  ③ 최빈값: 분산의 용이 아닌 제공근이다.  ④ 분산: 자료의 값 중에서 가장 많이 나타나는 값이다.  위의 보기 중 틀린 것을 모두 고른 것은?  A. ②, ③, ④  B. 없음  C. ③  D. ①, ②  E. ③  <b>Answer Key:</b> A  <b>Feedback</b></p>

Figure 6. An example of item coding and instances about a combined-response type and a negative type

### 매체 유형에 따라 생성된 문항 사례

대푯값과 산포도의 내용적 특성으로 주어진 자료의 특성에 따라 적절한 대푯값과 산포도를 선택하여 계산할 수 있는지, 그리고 과정적 특성으로 어떤 상황에서 대푯값과 산포도가 유용하게 사용될 수 있는지와 관련한 적용을 평가하는 질문을 매체 유형에 따라 생성한 문항 사례는 Figure 7과 같다. Figure 7에 제시되어있는 막대그래프와 원그래프는 Google Chart ("[https://developers.google.com/chart/interactive/docs/quick\\_start](https://developers.google.com/chart/interactive/docs/quick_start)")를 활용하였다. 막대그래프는 Google Charts의 왼쪽에 제시되어있는 Chart Type에서 Bar Charts를 선택하였으며, Google Charts의 왼쪽에 제시되어있는 Quickstart에서 pie chart를 활용하여 원그래프 문항을 생성하였다. 또한 표 및 동영상은 HTML 기반의 Tables Generator ("[https://www.tablesgenerator.com/html\\_tables](https://www.tablesgenerator.com/html_tables)")에서 Table을 활용하였으며, 동영상은 문제에 대한 피드백 영상으로 해당 통계단원 관련된 YouTube 영상을 src="[https://www.youtube.com/embed/ebXE\\_STkVBo](https://www.youtube.com/embed/ebXE_STkVBo)"을 활용하여 삽입하였다. 동영상과 마찬가지로 그림파일의 경우에도 연구자가 직접 그림을 그려서 문항에 삽입하거나 CAFA에 내장되어 있는 함수를 활용하여 직접 그릴 수 있다. 본 연구에서는 "<https://images.pexels.com/photos>"에 공개되어있는 그림파일을 Tables Generator의 Image URL을 이용하여 문항을 생성하였다.

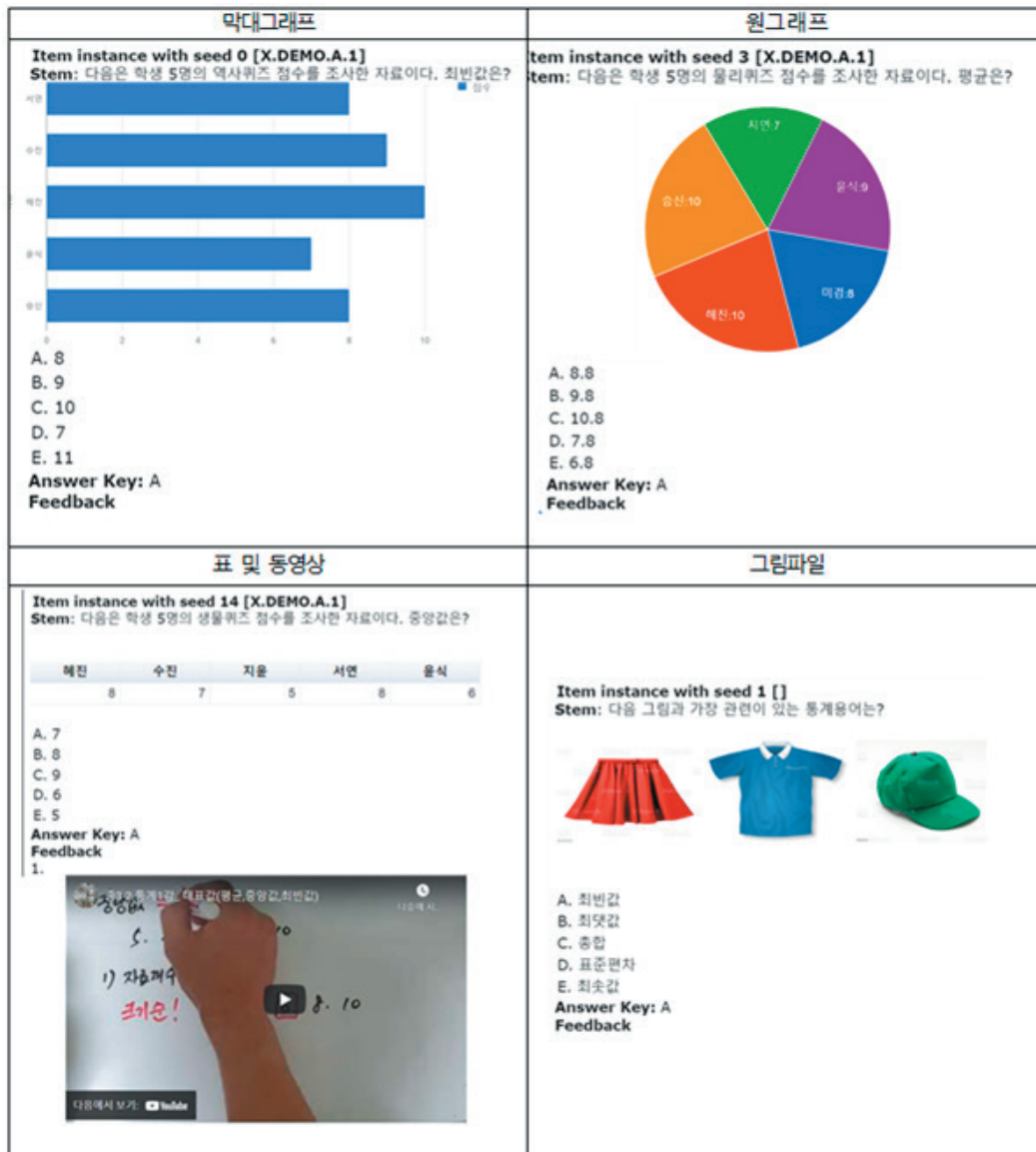


Figure 7. An example of item coding and instances using various media

Figure 8은 Figure 2의 문항모형으로부터 생성되었으며, @17@은 Google의 TTS(text-to-speech)를 활용하여 소리내어 읽어주는 문항의 예시이다. Figure 8의 문항모형 코딩 값에서 Korean Female로 제시되어있는 바와 같이 여성의 목소리로 질문과 선택지를 한국어로 읽어주는 문항이 생성된다.







매개변수	값	문항 사례
@29@	다음 중 용어와 설명이 올바른 것은? @TTS("다음 중 용어와 설명이 올바른 것은?", "# 20 20 CP 0 1 0 1 1 1 0 1 0 FI Black SH 1299 0 1 1 1 0 ##, Korean Female, 1, 1)  @1[1]@: @2[1]@ @1[2]@: @2[3]@ @1[3]@: @2[4]@ @1[4]@: @2[5]@ @1[5]@: @2[2]@	Item instance with seed 2 [] Stem: 다음 중 용어와 설명이 올바른 것은?  A. 평균: 자료의 총합을 자료의 개수로 나눈 값이다.  B. 중앙값: 편차를 제외한 값의 평균이다.  C. 분산: 자료의 값 중에서 가장 많이 나타나는 값이다.  D. 최빈값: 각 자료의 값에서 평균을 뺀 값이다.  E. 편차: 자료를 작은 값부터 크기순으로 나열하였을 때, 한 가운데 있는 값이다.  Answer Key: A Feedback

Figure 8. An example of item coding and instances using TTS

## 결론 및 제언

본 연구에서는 수학 3의 대푯값과 산포도 중단원의 내용에 대한 형성평가로 자동문항생성을 활용하여 온톨로지모형 기반의 문항모형을 CAFA 시스템에서 제작하는 절차를 명세화하였다. 이는 상대적으로 우리나라 학생들의 성취도가 떨어지는 통계 영역에서(Kim & Kang, 2020; Park et al., 2015) 기존의 연구들이 한글 기반의 문제은행을 구성하는 데에만 초점(Jeong et al., 2008)을 맞추거나, 최선답형 및 정답형을 중심으로 한 제한적인 문항유형만 생성(Choi et al., 2018; Kang & Choi, 2020; Lim, 2017) 하거나, 또는 중3의 대푯값 소단원에서만 문항모형을 활용(Choi et al., 2022)한 연구들과 달리 산포도 소단원의 내용을 포함하여 합답형이나 배합형과 같은 다양한 문항유형을 생성하는 문항모형 제작 절차와 문항 사례를 제시하였다는 점에서 의의가 있다. 구체적으로 자동문항생성 기반의 디지털 평가 활용방안에 대한 시사점을 각각 학생, 예비교사, 일반교사, 그리고 특수교육 측면에서 제시하면 다음과 같다.

학생 측면에서 본 연구의 결과는 디지털 지식 기반의 맞춤형 학습을 가능하게 하는데 활용될 수 있다. 온톨로지를 기반으로 문항에서 문항모형으로 전환됨에 따라 같은 내용이라 하더라도 다양한 유형의 문항이 생성되므로 이를 기반으로 즉각적인 취약점을 진단할 수 있게 된다. 일례로 본 연구에서 핵심 개념을 묻는 정답형 문항의 문두는 '다음 설명을 나타내는 용어는?'이다. 이에 해당하는 용어인 온톨로지는 6개이고, 오답 선택지는 5개, 추가로 오답 선택지를 구성하는 용어는 최댓값과 최솟값 2개이므로, 선택지를 구성할 수 있는 조합은  $C_4$ 이므로 35개가 된다. 여기에 온톨로지별로 선택지의 조합이 생성되므로 문항 사례는 210개가 된다. 만일 추가 오답 선택지를 10개로 늘린다면 선택지를 구성할 수 있는 조합은 1,365개가 되며, 문항 사례는 8,190개가 된다. 여기에 각각의 문항 유형을 고려한다면 하나의 문항모형에서 적어도 현재 생성된 문항 사례의 7배(최선답형, 합답형, 미완성문장형, 부정형, 진위형, 배합형, TTS)가 생성됨을 알 수 있다. 또한 문항모형에서 오개념을 반영한 오답들을 선택지로 작성함으로써 학생들이 오답을 선택할 경우 각 오답 별로 미리 설계한 피드백 슬롯의 내용들과 연계하여 맞춤형 피드백을 제공할 수 있다. 문항을 틀린 경우 학생들의 선택에 따라 선수학습에 해당하는 전단계로 이동할 수도 있으며, 같은 개념을 묻는 다른 유형의 문항을 다시 풀 수도 있으며, 문항을 맞힌 경우에는 후속학습에 해당하는 후단계로 이동하는 등 학생들이 현재 수준에서 다음 수준으로 나아갈 수 있는 방향을 제시할 수 있다. 한편 텍스트뿐만 아니라 그림, 소리 및 동영상과 같은 매체를 포함한 문항을 생성함으로써 학생들의 흥미를 유발시킬 수 있으며, 학생들이 수학을 학습할 때 수학적 정보 인식의 선호 경향이 시각적 양식인지 언어적 양식인지에 따라 학생들이 효과적으로 학습할 수 있는 방식을 반영(Kang et al., 2013; Paik, 2013)한 맞춤형 피드백을 제공할 수 있다(Kim et al., 2000). 이러한 학생 중심의 맞춤형 피드백은 학생들이 자기주도학습 전략을 개발하는데 좀 더 도움이 되고 효율적일 수 있다. 궁극적으로 자기주도학습 전략을 개발함으로써 고차원 사고력으로 알려진 메타인지를 개발하고, 학습동기를 유발하고, 학습속도를 향상시키는데 활용될 수 있다(Choi et al., 2018; Kim et al., 2000).

예비교사 측면에서 본 연구의 결과는 교과지식, 교수 및 평가 전반에 걸친 디지털 역량을 강화하는 데 활용될 수 있다. 이와 관련하여 최근 정부는 학생들이 다양하고 질 높은 디지털 교육을 학교에서 받을 수 있도록 예비교원들에게 디지털 역량을 함양할 수 있도록 교원양성기관의 교육과정 개편을 지원하겠다고 밝힌 바 있다(Ministry of Education, 2022). 디지털 역량이란 다양한 작업을 수행하기 위해서 ICT와 디지털 미디어를 사용할 때 요구되는 지식, 기술, 태도의 집합으로(Ala-Mutka, 2011) 이를 교육과정에 반영하기 위해서는 코딩과 같은 컴퓨팅 도구를 활용한 정보처리 수행 능력과 인공지능, 소프트웨어 및 정보화 디지털 영역의 컴퓨팅에 대한 기본 개념 및 원리 학습이 요구되어진다(Ministry of Education, 2022). 본 연구에서 활용한 CAFA 시스템을 통해 예비교사들은 교과 지식을 기반으로 한 온톨로지모형으로부터 수학적 논리를 바탕으로 하는 규칙 기반 프로그래밍의 종류의 문항모형을 제작하는 경험을 하게 된다. 이 과정은 수학 교육과 마찬가지로 논리적이고 구조적인 사고력이 요구되므로, 예비교원들의 이러한 사고력을 강화하는 데 활용할 수 있다(Choi et al., 2022). 또한 예비교사들이 인지하고 있는 추상적인 성격의 지식들을 온톨로지모형으로 작성하고, 문항모형을 바탕으로 디지털 객체인 문항 예시들을 생성하여 이를 경험적으로 관측할 수 있게 하기 위해서는 CAFA 시스템을 구현시키는 코딩 절차가 필요하다. 특히 텍스트, 소리, 이미지 및 동영상 등 다양한 유형의 디지털 데이터를 활용하여 문항모형을 제작하는 활동은 단순한 코딩 기술 습득에 그치는 것이 아니라 미래 학생들의 학교급에 따라 학생들이 쉽고 재미있게 학습할 수 있도록, 그리고 실생활의 문제 해결에 적용할 수 있도록 활용할 수 있다.

일반교사 측면에서 본 연구의 결과는 교사 학습공동체를 활성화시키는 데 활용될 수 있다. 교사 학습공동체란 교사의 전문성 신장과 학생들의 학습을 증진시키기 위해 교사들이 서로 협력적으로 배우고 탐구하고 실천하는 교사들의 집단을 의미한다(Seo, 2009). 우리나라에서는 역량 중심의 2015개정 교육과정이 고시되면서부터(Ministry of Education, 2015) 교사의 핵심 역량이 지식의 전달에서 학생의 역량 개발로 변하게 됨에 따라 교사들이 협력적으로 교육할 수 있는 교육전문가로서의 관점이 더욱 강조되고 있다(김창환 외, 2015). 또한 교사 학습공동체가 수업 전문성 향상과 학습 효과에 긍정적인 영향을 준다는 연구 결과가 보고되면서(Vescio et al., 2008) 교육정책 차원에서 교사들의 학습공동체 활동을 장려하고 있다(Kim & Lim, 2020). CAFA 시스템에서는 문항모형을 제작하기에 앞서 온톨로지모형을 작성하는 것이 필요하다. 온톨로지는 합의된 지식으로 어느 개인에게 국한되는 것이 아니라 그룹 구성원이 모두 동의하는 개념이라는(Choi, 2019; Wikipedia, 2022) 점에서 수학 교사들뿐만 아니라 여러 교과 교사들과 협력을 통해서 온톨로지모형을 작성할 수 있다. 주어진 글의 핵심 개념을 찾기 위해 자주 등장하는 단어의 빈도수를 고려하는 최빈값, 나이와 망막까지 도달하는 빛의 양의 관계를 나타내기 위해 산점도(Lee, 2020), 또는 기체의 압력과 부피를 그래프로 작성할 때 자주 범하는 오류(Kim et al., 2009) 등 수학 교과 중심의 융합수업 개발 및 평가를 일제로 들 수 있다. 이러한 교사들의 실질적인 필요성이 동기가 되어 교사들의 경험적 지식을 공유하기 위해 학습공동체 참여를 독려하는 데 활용될 수 있다.

특수교육 측면에서 장애가 있는 학생 또는 다문화 학생 대상의 평가에 활용될 수 있다. 특수교육에는 장애와 학업부진, 언어, 문화적 배경, 빈곤 등 다양한 어려움을 포함하는 모든 학생들이 포함되어 있다(Chu, 2014). 선진국들의 사례에 따르면, 다양한 장애 특성과 요구에 맞추어 수학교육환경을 융통성 있게 제공할 때 장애학생에게도 보통의 학생들과 동등한 학습 기회와 풍부한 경험이 제공되어 수학교육의 효과를 증진시킬 수 있다고 알려져 있다. 특히 시각장애 학생들을 위한 수학교육 환경은 스크린을 읽어주는 프로그램이나 책을 모니터에 비추고 이 책을 문자인식 장치를 이용하여 말로 읽어주거나 점자책으로 바꾸어 주는 프로그램 등이 이미 현실에서 구현되고 있는 실정이다(Lee et al., 2016). CAFA 시스템의 TTS 옵션은 텍스트를 소리로 변환시켜주는 기능으로 이는 시각장애 학생들의 평가와 피드백에 유용하게 활용될 수 있다. 또한 구글 번역기와 CAFA 시스템의 LINK 옵션을 사용함으로써 한글을 다양한 언어로 번역하는 것이 가능하다. 이는 다문화가정 학생들의 평가와 Ministry of Education (2018)의 글로벌 역량개발을 지원하는 일환으로 방과후학교 또는 창의적 체험활동에서 이중언어교재를 보급하는 데에도 유용하게 활용될 수 있다.

상기에서 제시한 활용방안이 교육현장에서 실현되기 위해서는 CAFA 시스템과 같이 교육 분야에서 디지털 전환이 빠르게 이루어질 수 있도록 정부의 지원 및 교사, 학습자, 교육기업과 같은 모든 교육 참여자들이 편리하게 활용할 수 있는 국가 단위의 디지털 기반시설을 구축하는 것이 필요하다. 또한 디지털 평가의 중요성에 대한 교사들의 인식 제고를 통해 평가 경험이 풍부한 현장 교사들의 자발적인 협력을 바탕으로 한 다양한 연수가 제공되어야 한다.

마지막으로 본 연구의 제한점과 향후 연구방향을 제시하면 다음과 같다. 본 연구에서는 자동문항생성 기반의 디지털 평가를 개발하는 일례로 수학 3의 대푯값과 산포도 중단원의 내용을 형성평가의 일부로 설계하여 제시하였지만, CAFA에서 제공하는 수학 함수 및 그래프, 그리고 구글차트 같은 오프 소스 패키지를 수학교사들의 교과지식과 접목함으로써 향후에는 다양한 학교급 또는 수학 영역별로 분류하여 문항모형을 제작하는 것이 수행되어야 할 것이다. 또한 본 연구에서는 CAFA 활용방안에 초점을 맞추어 생성된 예시문항들을 제시하는 것에서 그쳤지만, 실제 생성된 문항들을 학생들에게 응답하게 함으로써 문항과 학생들의 반응 데이터를 CAFA 시스템 내에서 누적하고 실시간으로 분석하는 연구가 수행되어야 할 것이다.

## Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No.2022R1A2C1010310).

I would like to thank three anonymous reviewers and the editor for their valuable comments and suggestions to improve this paper. I would also like to thank Professor Jaehwa Choi (The George Washington University) and Kyongil Yoon (Notre Dame of Maryland University) for their dedicated support in using the CAFA system.

## References

- Ala-Mutka, K. (2011). Mapping digital competence: Towards a conceptual understanding. *Sevilla: Institute for Prospective Technological Studies*, 7-60.
- Bejar, I. I. (2002). Generative testing: From conception to implementation. In S. H. Irvine & P. C. Kyllonen (Eds.), *Item generation for test development* (pp.199-217). Erlbaum.
- Bejar, I. I., Lawless, R. R., Morley, M. E., Wagner, M. E., Bennett, R. E., & Revuelta, J. (2002). A feasibility study of on - the - fly item generation in adaptive testing. *ETS Research Report Series*, 2002(2), 1-44. <https://doi.org/10.1002/j.2333-8504.2002.tb01890.x>
- Boo, J., & Vispoel, W. (2012). Computer versus paper-and-pencil assessment of educational development: A comparison of psychometric features and examinee preferences. *Psychological Reports*, 111(2), 443-460. <https://doi.org/10.2466/10.03.11.PR0.111.5.443-460>
- Carroll, S. N. (2021). *Fourth grade student experiences in mathematics assessment: Computer versus paper-pencil testing modality*. [Unpublished Doctoral Dissertation, Carson-Newman University].
- Choi, J. H. (2017). *Introduction to Assessment Engineering. Assessment, Testing, and Measurement Technical Report Series*. The George Washington University.
- Choi, J., Kim, H., & Pak, S. (2018). Evaluation of Automatic Item Generation Utilities in Formative Assessment Application for Korean High School Students. *Journal of Educational Issues*, 4(1), 68-89. <https://doi.org/10.5296/jei.v4i1.12630>
- Choi, J. H. (2019). Assessment engineering for learning analytics, *Educational Development*, 210, 42-45.
- Choi, J. H. (2020). Educational innovation in the digital age: A plan to implement an intelligent learning analysis platform based on big data. *Educational Development*, 214, 44-50.
- Choi, J. H., Oh, K. S., Youn, K. I., Lee, D. Y., Joung, J. M., Kim, S. Y., Youm, S. K., Lee, Y. E., Lee, E. J., Park, W. H. K & Lee, S. H. (2022). *CAFA Model Example Book*, Pubple.
- Choi, J. H., Kim, S. H., & Yoon, K. I. (2012-2022). *CAFA Modeling Manual: Computer Adaptive Formative Assessment User's Guide [System Manual]*, 2nd edition, CAFA Lab, Inc.
- Chu, Y. G. (2014). Approach toward multicultural education and special education as inclusive education. *Journal of Intellectual Disabilities*, 16(2), 333-353.
- Dardick, W., & Choi, J. (2016). Teacher empowered assessment system: Assessment for the 21st century. *Journal of Applied Educational and Policy Research*, 2(2), 87-98.

- Deane, P., Graf, E. A., Higgins, D., Futagi, Y., & Lawless, R. (2006). Model analysis and model creation: Capturing the task - model structure of quantitative item domains. *ETS Research Report Series*, 2006(1), 1-63. <https://doi.org/10.1002/j.2333-8504.2006.tb02017.x>
- Embretson, S. E., & Kingston, N. M. (2018). Automatic item generation: A more efficient process for developing mathematics achievement items?. *Journal of Educational Measurement*, 55(1), 112-131. <https://doi.org/10.1111/jedm.12166>
- Fridenfolk, M. (2013). *System for automatic generation of examination papers in discrete mathematics*. International Association for Development of the Information Society.
- Gierl, M. J., & Haladyna, T. M. (2012). *Automatic item generation: Theory and practice*. Routledge.
- Gierl, M. J., & Lai, H. (2013). Instructional topics in educational measurement (ITEMS) module: Using automated processes to generate test items. *Educational Measurement: Issues and Practice*, 32(3), 36-50. <https://doi.org/10.1111/emip.12018>
- Gierl, M. J., Zhou, J., & Alves, C. (2008). Developing a taxonomy of item model types to promote assessment engineering. *Journal of Technology, Learning and Assessment*, 7(2), 1-51.
- Holling, H., Bertling, J. P., & Zeuch, N. (2009). Automatic item generation of probability word problems. *Studies in Educational Evaluation*, 35(2-3), 71-76. <https://doi.org/10.1016/j.stueduc.2009.10.004>
- Jeong, J. C., Shin, K. A., Lee, S. H., & Yoo, W. H. (2008). *Design and implementation of iterative contents based on SCORM in mathematics*. Proceeding of the Winter Conference of the Korea Society of Computer and Information, 16, 153-158.
- Kang, N. R., Lim, D. K., & R, H. A. (2013). Analyzing the styles and types of math learning for middle school students. *Journal of the Korean School Mathematics Society*, 16(2), 363-381.
- Kang, S. K., & Choi, S. P. (2020). *Research on mathematical automatic item generation based on dynamic knowledge-base*. Proceeding of the Korean Information Science Society Conference, 302-304.
- Kellogg, M., Rauch, S., Leathers, R., Simpson, M. A., Lines, D., Bickel, L., & Elmore, J. (2015, April). *Construction of a dynamic item generator for K-12 mathematics*. In National Council on Measurement in Education Conference, Chicago, IL..
- Kim, C. H., Kim, B. Y., Moon, S. B., Kim, I. K., Kim, J. S., Park, J. H., Park, H. J., Lee, K. H., & Chae, J. E. (2015). *A study on the development of educational indicators and indexes (IV)*. Korean Educational Development Institute.
- Kim, H. I., Heo, J. W., & Kim, K. S. (2000). *Development of web-based mathematics learning model for self-directed learning*. Proceeding of the Korean Information Science Society Conference, 27(1B), 678-680.
- Kim, H. R. (2021). New directions of english education course management for pre-service primary teachers in the post corona era. *Educational Research*, 81, 119-139. <http://doi.org/10.17253/swueri.2021.81.006>
- Kim, S. H., & Kang, H. Y. (2020). Exploring Ways to Improve Statistics Curriculum by Analyzing Secondary Curriculum in Foreign Countries. *School Mathematics*, 22(3), 631-650. <https://doi.org/10.29275/sm.2020.09.22.3.631>
- Kim, Y., Choi, G., & Noh, T. (2009). High school students' errors in constructing and interpreting science graph. *Journal of The Korean Association For Science Education*, 29(8), 978-989.
- Kim, W., & Lim, W. (2020). Research trends in mathematics teacher learning community: Literature review. *Communications of Mathematical Education*, 34(4), 439-464. <https://doi.org/10.7468/jksmee.2020.34.4.439>
- LaDuca, A., Staples, W. I., Templeton, B., & Holzman, G. B. (1986). Item modeling procedure for constructing content - equivalent multiple choice questions. *Medical education*, 20(1), 53-56. <https://doi.org/10.1111/j.1365-2923.1986.tb01042.x>
- Lee, S. G., Park, K. E., & Ham, Y. M. (2016). A study on 3D printed tactile mathematics textbook for visually impaired students. *Communications of Mathematical Education*, 30(4), 515-530. <https://doi.org/10.7468/jksmee.2016.30.4.515-530>
- Lee, E. H. (2010). *A summary-concept based analysis of the representative values and the measures of dispersion in the 9<sup>th</sup> grade textbook of Korean middle school* [Master Thesis, Ewha Womans University].
- Lee, S. M. (2020). *Mathematical connection in teaching correlation-Analysis of 2015 Revised Mathematics Curriculum and textbooks-* [Master Thesis, Seoul National University].
- Lim, H. J. (2017). Study on the design and properties of automatically generated items: focusing on polynomial factorization [Unpublished doctoral dissertation, Seoul National University].
- Luecht, R. M. (2012). *An introduction to assessment engineering for automatic item generation*. In *Automatic item generation* (pp. 69-86). Routledge.

- Ministry of Education (2015). *The guide of revision of educational curriculum*. Ministry of Education Notice No. 2015-80. Ministry of Education.
- Ministry of Education (2018). *2018 Multicultural education support plan*. Ministry of Education
- Ministry of Education (2022). In order to strengthen digital public education, we will promote teacher competency enhancement and teacher recruitment. <https://www.moe.go.kr/boardCnts/viewRenew.do?boardID=72771&lev=0&statusYN=W&s=moe&m=0315&opType=N&boardSeq=92581>
- Mislevy, R. J., Behrens, J. T., Bennett, R. E., Demark, S. F., Frezzo, D. C., Levy, R., Robinson, D. H., Rutstein, D. W., Shute, V. J., Stanley, K., & Winters, F. I. (2010). On the roles of external knowledge representations in assessment design. *Journal of Technology, Learning, and Assessment*, 8(2), 1-56. <https://doi.org/10.1037/e643752011-001>
- Oh, J. S., Chu, S. W., Cho, W. J., Kim, Y.S., & Lee, J. Y. (2006). A design of dynamic question generation system using Korean wordnet. *Journal of Korean Institute of Information Technology*, 4(5), 37-44.
- Oh, K. S. (2022). A study on the development of reading assessment using automatic item generation -focused on reading comprehension item model-. *Journal of CheongRam Korean Language Education*, 87, 7-34.
- Paik, H. S. (2013). Categorization of middle school students' math learning style preferences and comparison of academic characteristics. *School Mathematics*, 15(1), 15-35.
- Park, K. M. et al., (2015). *2015 Revision mathematics curriculum proposal development study II*. Ministry of Education. Research Report of Korea Foundation for the Advancement of Science and Creativity.
- Park, J. H. (2017). *A study on teaching method for mathematics underachiever through step-by-step learning card-focused on 'number and operations' unit of 9th grade* [Master thesis, Korea National University of Education].
- Seo, K. H. (2009). Teacher learning communities and professional development. *The Journal of Korean Teacher Education*, 26(2), 243-276.
- Schwinning, N., Schypula, M., Striewe, M., & Goedicke, M. (2014). *Concepts and Realizations of Flexible Exercise Design and Feedback Generation in an e-Assessment System for Mathematics*. In CICM Workshops (pp. 1-10).
- Sinharay, S., & Johnson, M. S. (2012). *Statistical modeling of automatically generated items*. In Automatic item generation (pp. 193-205). Routledge.
- Singley, M. K., & Bennett, R. E. (2002). Item generation and beyond: Applications of schema theory to mathematics assessment. In S. Irvine & P. Kyllonen (Eds.) *Item Generating for test development* (pp.361-384). Lawrence Erlbaum.
- Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education*, 24(1), 80-91. <https://doi.org/10.1016/j.tate.2007.01.004>
- Wilson, J., Morrison, K., & Embretson, S. E. (2014). *Automatic item generator for mathematical achievement items: MathGen3.0*. Technical report IES1005A-2014 for Institute of Educational Sciences Grant R305A100234. Cognitive Measurement Laboratory, Georgia Institute of Technology.
- Wikiedia (2022). Ontology, *Wikipedia, Wikimedia Foundation*, 23, February, 2022, <https://ko.wikipedia.org/wiki/%EC%98%A8%ED%86%A8%EB%A1%9C%EC%A7%80>
- Xinxin, Z. (2019). *Using Automatic Item Generation to Create Content for Computerized Formative Assessment* [Unpublished doctoral dissertation, University of Alberta].