

교육과정평가연구

The Journal of Curriculum and Evaluation

2021, Vol. 24, No. 1, pp. 101~122

DOI: <https://doi.org/10.29221/jce.2021.24.1.101>

통합적 접근에 기반한 학생 평가 문항의 특성 분석 방안 탐색¹⁾

박지현 (한국교육과정평가원 연구위원)*
송미영 (한국교육과정평가원 선임연구위원)**
남민우 (한국교육과정평가원 연구위원)
최길찬 (CRESST Associate Director)

요약

최근 학생 평가 문항에 대한 새로운 분석 방법으로 문항 특성 분석(FA) 기법이 제안되었다. 이 기법은 평가 문항의 다양한 특성에 대한 질적 평정 및 평정한 특성과 평가 자료에 대한 양적 분석이 결합된 방법론이라 할 수 있으며, 문항 특성은 무엇을 어떻게 평가하는가에 대한 내용적·형식적 정보를 반영하고 있다. 본 연구는 문항 특성 분석(FA) 기법에 따라 2019년 국가수준 학업성취도 평가(NAEA) 중학교 국어, 수학 평가도구의 문항 특성을 교과 전문가들이 정성적으로 평정한 결과를 살펴봄으로써 당초의 평가 계획대로 평가도구 개발이 이루어졌는지 점검할 수 있음을 밝혔다. 또한 문항 특성별로 난이도와 변별도에 미치는 영향에 대한 정량적 분석을 통해 문항 특성 차원에서 학생 성취 결과에 대한 해석과 활용도 가능성을 확인하였다. 이러한 결과를 바탕으로 평가의 타당도 측면과 평가 결과의 환류 측면에서 갖는 문항 특성 분석(FA) 기법의 유용성이 학교 현장에서 심화되는 기반이 될 수 있는 후속 연구를 제안하였다.

주제어 : 문항 특성 분석, 학생 평가의 타당도, 교육과정 기반 평가, 국가수준 학업성취도 평가

1) 본 논문은 박지현 외(2020)의 일부 내용을 발췌하여 재구성함.

* 제1저자, pjh210@kice.re.kr

** 교신저자, mysong@kice.re.kr

I. 서 론

최근 학교 교육의 내용과 방법 측면에서 다양한 변화가 이루어지고 있는데, 학생 평가에서는 학생의 학습 과정과 현 상태를 진단하고, 후속 학습을 지원하는 기능이 주목받고 있다(박지현 외, 2018, p. 15). 교과별 교육과정의 성취기준에 도달한 정도를 파악하고, 학생의 성장을 돕기 위해 평가 결과를 활용하는 것을 강조하고 있으며, 서·논술형 평가와 수행평가 활성화 방안을 마련하도록 하는 등(교육부, 2018, p. 14; 서울특별시교육청, 2020, p. 18), 학생의 성장을 지원하는 기능을 강화하는 방향으로 학생 평가 정책이 변화하고 있다. 학생의 수준에 대한 판단과 성장 지원을 위한 자료로서 평가가 유의미하게 활용되기 위해서는 정책 변화에 따른 형식적 측면의 개선뿐만 아니라 내용적 측면의 점검이 이루어져야 한다. 특히 평가의 목표와 내용의 일치, 즉 평가의 타당도가 전제되어야 평가의 결과가 학생 성장을 돕는 피드백 자료로 적절하게 활용될 수 있다.

타당도(validity)는 신뢰도와 함께 질 높은 평가를 위해 필수적으로 갖추어야 할 조건으로 평가 도구가 의도한 평가 목표를 반영하는 정도를 의미하며 평가 결과의 해석·추론·활용의 정확성·적절성과 직결된다(AERA et al., 2014, p. 11; McMillan, 2018, p. 79). 따라서 학생 평가를 위해 시행되는 다양한 평가의 타당도 확보를 중요하게 인식하여야 하며, 평가 계획 및 평가 도구 개발 단계에서부터 타당도를 점검할 필요가 있다. 평가 문항이 가지는 다양한 특성들은 문항으로 구성되는 평가 도구의 성격을 규정할 뿐만 아니라 문항에 대한 학생들의 반응에 영향을 미칠 수 있기 때문에, 평가 목적에 부합하게 문항 제작이 이루어졌는지를 면밀히 살펴야 한다.

국가 및 시도 수준의 대규모 평가나 단위학교의 학생 평가는 교과별 교육과정에 대한 학생들의 도달 수준을 파악하기 위한 목적에서 실시되는데, 일반적으로 평가 도구에 대한 전반적인 출제 계획을 수립하여 각각의 문항에 대한 정보를 설정한 후 출제하고, 출제한 문항과 사전에 수립한 계획 간 일치 여부를 점검하기 위해서 몇 단계의 검토를 거친다. 예를 들어, 단위학교의 경우 평가 계획 수립 시 평가의 영역, 내용 등의 정보가 포함된 문항정보표를 작성하고, 이를 중심으로 평가 도구를 개발할 뿐만 아니라(교육부, 2020a, p. 26), 평가 도구에 대한 상호 검토를 실시하도록 하고 있다(교육부, 2020b, p. 130). 이러한 방식에 의한 평가 도구의 질 점검은 출제자의 전문성이 확보되었다는 가정 하에 이루어진다고 볼 수 있으며, 결과적으로 평가의 타당도는 출제자의 전문적 판단에 영향을 받는다(McMillan, 2018, pp. 80-81). 따라서 평가 도구 개발 시 이루어지는 출제자의 전문적 판단을 지원하는 실증적 점검 방안이 마련될 필요가 있다.

평가 문항에 대한 최신의 분석 방법인 문항 특성 분석(Feature Analysis; 이하 FA, Baker et al., 2015a; 2015b; Choi et al., 2020) 기법에 근거하여 각 문항의 특성(item feature)이 평가 계획에 부합하는지를 실증적으로 점검할 수 있다. 문항 특성이란 문항이 무엇을 어떻게 평가하는가와 관련된 정보를 포괄하는 개념이며, 문항이 가지고 있는 내적 속성과 외적 형식 측면의 특성은 난이도와 변별도 등의 통계적 속성에 반영된다. 문항 특성 분석(FA) 기법은 평가 목표에 대한 분석을 통해 평가 도구에 포함해야 할 문항 특성을 상정한 후, 평가 문항의 다양한 특성에 대한 질적 평정 및 평정한

특성과 난이도의 관계 등에 대한 양적 분석을 통해 평가의 타당도를 점검하는 통합적 접근법을 취한다. 문항 특성 분석(FA) 결과는 평가가 시행되기 이전에 평가 도구의 검토와 수정을 위해 활용 가능하며, 평가가 시행된 이후 교수학습 및 평가 도구 개선을 위한 기초 자료를 제공할 수 있다. 따라서 문항 특성 분석(FA) 기법을 적용하여 평가 문항을 분석하는 것은 국가 수준뿐만 아니라 학교 수준에서 시행되는 학생 평가의 타당도를 높일 수 있는 유용한 방법이 될 수 있다.

본 연구는 학생 평가의 타당도 관점에서 통합적 접근에 기반한 문항 특성 분석(FA) 기법의 유용성을 밝히고자 한다. 이를 위하여 교육과정 기반의 학생 평가에 적용할 수 있는 문항 특성 분석틀을 활용하여 교육과정에 기반을 둔 평가를 대상으로 문항의 특성을 평정하고, 문항 특성과 학생의 성취 특성과의 관계를 분석한다. 문항 특성에 대한 통합적 분석을 통해 문항 특성의 분포 양상을 파악하고, 학생의 성취도를 예측하는 문항 특성과 각 특성의 영향력에 대한 정보를 도출한다. 이와 같은 분석 결과는 평가 목표 - 평가 내용 - 평가 결과 해석 및 활용의 일관성을 확보하고, 학생 평가와 교수학습 간 환류 체계를 개선하기 위한 방안을 마련하는 데 기여할 수 있을 것이다.

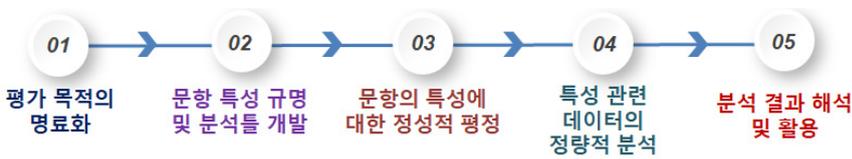
II. 문항 특성에 대한 정성적 평정과 정량적 분석

1. 문항 특성 분석 기법

문항 특성 분석(FA) 기법은 평가 문항의 특성을 평정하고 이를 평가 결과와 연계하여 살펴봄으로써 평가의 타당도를 점검할 뿐만 아니라, 평가 도구 및 교수학습 개선을 위한 정보를 산출할 수 있는 최신의 분석 기법이다(Baker et al., 2015a; 2015b; Baker & Choi, 2019). 여기서 문항 특성에는 평가 문항을 통해 측정하고자 하는 인지적 요소, 문항에 제시된 언어/도식/수리적 자료 유형, 문항에 응답하는 방식과 채점 방식 등 내용적·형식적 측면의 특성이 포함될 수 있다. 문항 특성 분석(FA) 기법은 다양한 측면의 문항 특성에 대해 질적 접근과 양적 접근을 결합하여 분석한다는 특징을 갖는다. 구체적으로 평가 도구를 구성하는 요소인 문항(item)이나 과제(task)에 대한 내용적·형식적 특성을 정의·규명하고, 각 문항이 어떤 특성을 갖는지 평정·분류하는 단계는 질적 접근이라 할 수 있고, 질적 분석에 따른 특성 데이터에 대한 통계 분석을 통해 평가 도구에서 두드러지는 특성은 무엇이며 사전에 설정한 목표에 부합하는지, 각 특성과 관련된 난이도 분포가 어떠한지 등을 파악하는 단계는 양적 접근에 해당한다.

문항 특성 분석(FA) 기법의 세부 절차를 살펴보면(그림 1) 참조, 가장 먼저 평가의 목적을 명료화하고 이에 부합하는 문항 특성을 규명한다. 학생 평가가 실시되는 목적과 시행 방식에 따라 문항의 내용이나 유형이 달라질 수 있으므로, 이를 고려하여 문항 특성이 선별될 수 있다. 문항

특성을 정의하고 범주화하여 분석틀을 구성한 후, 분석틀에 기초하여 각 문항의 특성을 평정한다. 문항 특성에 대한 평정은 문항별로 포함하고 있는 특성의 경우 1, 그렇지 않은 특성의 경우 0으로 표기한다. 각 문항의 특성을 평정한 결과에 특성별 빈도를 분석하여 각 특성이 전체 문항을 통해 고루 다루어졌는지, 어떠한 특성에 편중되어 있지 않은지, 상대적으로 덜 다루어진 특성은 무엇인지 등을 파악할 수 있다. 또한 평정 결과와 검사 자료를 연계한 통계 분석을 통해 문항 특성의 관점에서 학생의 성취에 미치는 영향력을 밝힐 수 있다(예, 조지민 외, 2018, p. 30). 이러한 분석 결과를 바탕으로 평가 도구의 전반적인 성격을 파악하고 평가 계획 시 목표로 설정한 핵심적인 내용을 충실히 평가하는지를 살펴볼 수 있다.



* 출처: Baker와 Choi(2019, p. 7, 재구성)

[그림 1] 문항 특성 분석(FA) 기법의 세부 절차

평가도구의 구성 단위원 문항이 가진 다양한 특성을 사전에 정의하고, 문항 특성에 비추어 각 문항을 종합적으로 검토하는 특성 분석(FA) 기법은 평가의 타당도를 높이기 위하여 효과적으로 활용될 수 있다. 첫째, 문항에 대한 전문가의 특성 평정을 통해 평가 도구가 측정해야 할 목표에 부합되게 문항이 측정하고 있는지에 대한 실증적인 자료를 제공한다. 둘째, 평가 도구 혹은 개별 문항을 수정하거나 향후 새로운 평가 도구를 제작할 시 필요한 기초 자료를 얻을 수 있다. 셋째, 문항 특성에 따른 학생들의 성취 결과를 분석하여 교수학습에 대한 피드백 제공이 가능하다. 따라서 문항 특성 분석(FA)을 통해 평가도구 개발자·사용자로서의 교사, 교수학습 안내자로서의 교사에게 교육적 시사점을 제공할 수 있다(Madni et al., 2015, pp. 43-44).

지금까지 학생 평가의 타당도는 평가 도구를 개발하는 단계에서 출제자의 전문성에 의존하여 확보되는 경우가 대부분이었다. 이에 반하여 문항 특성 분석(FA) 기법을 적용하면, 실증적인 자료를 기반으로 평가의 타당도를 체계적으로 검토할 수 있다. 학생 평가의 타당도 제고를 위해 다음과 같이 문항 특성 분석(FA) 기법을 적용할 수 있다. 구체적으로 평가 설계 및 개발 단계에서는 평가 계획 시 목표로 설정한 특성들이 출제된 평가 도구에 어떤 양상으로 분포되어 있는지를 파악하고, 평가 도구 구성 과정에서 평가 목표에의 부합성이나 특성 간 균형 등을 점검하며 차기 평가 도구 개발을 위해 피드백한다. 뿐만 아니라 학생의 성취와 성장에 영향력 있는 문항 특성을 확인하여 향후 교육과정 개발 및 운영의 개선을 위한 시사점을 마련한다. 이와 같이 평가 도구의 개발에서부터 결과의 활용에 이르기까지 학생 평가의 전반적인 운영 과정에서 평가 문항의 질을 점검하고 그 결과를 평가 도구 개발과 교수학습 계획에 환류할 수 있다. 평가 자체의 타당도 검증은 위해 문항 특성에 대한 분석 결과를 평가 결과의 예측이나 분석과 함께 활용한다는 점에서 전문가의 판단에 의존하는 기존의 타당도 점검 방법과 차별화된다.

2. 문항 특성 분석(FA) 기법 관련 국제 연구 동향

실제로 문항의 특성 분석(FA) 기법을 적용한 연구는 비교적 최근에 수행되기 시작하였다. 미국 SBA(Smarter Balanced Assessment) 문항의 특성을 분석한 연구(Madni et al., 2015), 대학-직업 준비도 측면에서 문항 특성을 추출하고 대규모 평가 결과를 분석한 연구(조지민 외, 2018; Baker et al., 2018; Choi et al., 2018), 웹 개발자에게 요구되는 특성이나 교육용 게임의 특성을 추출하고 분석한 연구(Choi, 2019; Chung et al., 2018) 등이 수행되었다. 또한 교육과정 기반 평가 문항의 국제 비교 연구를 위한 문항 특성을 추출하고, 여러 국가의 학업성취도 평가 문항에 대한 특성을 비교하는 연구가 진행 중에 있다(Baker & Choi, 2019, p. 3).

Madni 외(2015)는 미국의 SBA 초·중·고 문항에 문항 특성 분석(FA) 기법을 적용한 결과, 선다형 상호작용 유형(Multiple choice problem interaction type)이 문항 난이도를 줄이는 데 가장 핵심적인 특성인 것으로 나타났다. 이 연구는 향후 검사 개발이나 교육 방향 마련을 위해 평가 결과를 활용할 수 있도록 학생 평가 문항의 특성을 체계적으로 분석할 수 있는 기법을 개발하고, 교육과정과 관련한 특성을 분석할 수 있을 뿐만 아니라 그 외 다양한 특성을 종합적으로 파악할 수 있음을 밝혔다.

조지민 외(2018)는 대학-직업 준비도 측면에서 우리나라 2017년 국가수준 학업성취도 평가(이하 NAEA) 중학교 3학년 수학 문항과 미국 SBA의 11학년 수학 문항의 특성을 비교·분석하였다. 연구 결과, NAEA는 대학-직업 준비도 측면의 문항 특성 중 ‘자료 및 정보 분석’ 특성의 비율이 높은 반면, SBA는 ‘연역적 추론’의 비율이 가장 높아 차이가 있었으며, ‘의사결정과 문제 해결’ 등의 특성이 문항 난이도와 정적인 관련이 있음을 보였다. 이 연구에서 활용한 문항 특성 분석틀은 교육과정 기반 학생 평가의 타당도 제고와 교수학습의 질 관리를 위해서 활용하는 데에는 한계가 있다.

미국 CRESST를 중심으로 핀란드, 독일을 비롯한 9개국에서 연구네트워크를 결성하여 각국의 교육과정에 기반한 대규모 학업성취도 평가 문항에 대해 문항 특성 분석(FA) 기법을 적용하고, 문항 특성과 학생의 성취수준 간 관계뿐만 아니라 각국에서 시행하는 학업성취도 평가의 유사성 및 차별성을 살펴보는 국제 비교 연구가 진행 중이다. 각 참여국에서는 공통의 문항 특성 분석틀을 설정하여 평가 도구의 특징을 살펴보는데, 내용 영역(Content/Domain), 인지(Cognition), 과제(Task), 언어적(Linguistics) 범주의 특성을 설정하고 있다(Baker & Choi, 2019, pp. 7-11). 이 연구의 국가 간 공통의 분석틀은 타국의 평가 도구와 자국의 평가 도구의 특징을 파악할 수 있다는 장점이 있지만, 각국의 교육과정에서 상정하고 있는 교육목표와 평가 도구에 대한 출제 방식의 차이 때문에, 자국의 학생 평가에 대한 타당도를 점검하기 위한 분석틀은 교육목표와 출제 방식 등의 평가 상황에 적합한 문항 특성을 추출하여 마련할 필요가 있다.

3. 평가 문항의 특성에 관한 국내 연구 동향

우리나라에서 수행된 평가 문항의 특성에 관한 연구는 교과별 교육과정, 교수학습 방법, 평가 도구 등을 개선하기 위한 목적과 적정 난이도에 맞추어 평가 도구를 구성하기 위한 목적으로 나누어 볼 수 있다. 먼저 전자의 목적으로 대규모 평가 문항의 특성을 분석한 연구(구자옥 외, 2019a; 권태현, 이정찬, 김승현, 2017; 권태현, 김승현, 2019; 남민우, 민재원, 이인화, 2016)에서 설정한 문항 특성을 살펴보면 다음과 같다. 2015년~2018년 NAEA 문항을 질적으로 분석하고 성취기준별 학생의 성취 특성을 도출한 구자옥 외(2019a)의 연구에서는 내용의 난이도, 자료나 소재의 친숙도, 문항 유형(답안 작성 방식, 문항 구성이나 답지 제시의 복잡성) 등과 같은 특성을 기준으로 문항을 분석하였다. 한편 대학수학능력시험(이하 수능) 국어 문항에 대한 양호도를 분석한 남민우 외(2016), 권태현 외(2017), 권태현과 김승현(2019)의 연구에서는 문항별 평가 내용의 교육과정상 중요도, 문항의 완성도, 난이도, 참신성, 지속성의 5개 범주로 구성된 문항 양호도 평정틀을 설정하고, 교육과정과 평가 문항 간 연계성 강화 등을 제안하였다.

다음으로 목표 난이도에 맞춘 평가 도구의 구성을 위해 평가 문항의 특성을 분석한 연구(고호경, 이현숙, 2007; 박문환, 2004; 송현정, 2004; 이종승 외, 2003)에서는 평가 도구의 적정 난이도 예측 모형을 개발하기 위해 문항 특성을 고려하였다. 예를 들어, 2002학년도 및 2003학년도 수능 영역별 최적의 난이도 예측 모형을 도출한 이종승 외(2003)의 연구에서는 언어 및 수리 영역과 관련하여 각각 10개, 6개의 특성이 선별되었다. 이 연구는 내용 전문가의 질적 판단에만 의존하지 않고 실증적 검증 방법의 결합을 통해 예측 모형을 제안하였다는 의의를 지니지만 평가 도구의 난이도 측면에서 문항 특성을 조정할지에만 주목하였다.

〈표 1〉 선행 연구에 나타난 국어/수학 관련 문항 특성

분석 대상		문항 특성	관련 연구
평가	교과		
NAEA	수학	<ul style="list-style-type: none"> 성취기준에 따른 내용의 난이도 자료나 소재의 친숙도 문항 유형: 답안 작성 방식(서답형), 문항 구성/답지 제시의 복잡성(합답형 등) 	구자옥 외(2019a)
	국어	<ul style="list-style-type: none"> 교육과정상의 중요도 문항의 난이도 문항의 지속성 	<ul style="list-style-type: none"> 문항의 완성도 문항의 참신성 권태현 외(2017) 권태현, 김승현(2019) 남민우 외(2016)
수능	언어	<ul style="list-style-type: none"> 듣기 영역: 내용/제재의 생소성, 구조/내용전개의 복잡도, 유형의 생소성 쓰기 영역: 과제의 생소성, 과제 해결 소요 시간, 행동 영역 읽기 영역: 과제 해결 소요 시간, 자료로 인한 문제의 복잡도, 오답지 매력도, 문제 해결을 위해 요구되는 언어 단위 	박문환(2004) 송현정(2004) 이종승 외(2003)
	수리	<ul style="list-style-type: none"> 내용이나 제재의 생소성, 문제 해결에 필요한 개념의 수, 계산의 복잡성, 내용 영역, 행동 영역, 문항 유형 	
전국 연합 학력평가	수학	<ul style="list-style-type: none"> 내용 영역: 수학 교과 내용적 유사성에 따라 재설정 행동 영역: 계산, 이해, 발견적·연역적 추론, 내적·외적 문제 해결 문항 유형: 정답형, 합답형, 부정형, 완성형, 단답형 	고호경, 이현숙 (2007)

평가 문항의 특성에 관해 국내에서 수행된 선행 연구들에서 다루어진 문항 특성은 <표 1>과 같이 정리할 수 있다. 각 연구의 목적에 따라 문항의 주요 특성이 선별되었듯이 교육과정을 중심으로 학생들의 성취도를 파악하는 평가를 대상으로 할 경우에는 교육과정의 강조점을 함께 반영하여 문항 특성을 설정할 필요가 있다. 또한 제재의 생소성, 문제의 복잡도와 같이 주관성이 개입될 여지가 있는 특성은 교과 전문가의 검토가 필요하므로(구자옥 외, 2019a; 이종승 외, 2003), 주관적 평정에 따라 발생될 수 있는 차이를 보정할 방안을 마련하여야 한다.

III. 연구 방법

1. 분석 대상

학생 평가와 교육과정 간의 연계성 강화 맥락에서 타당도 제고 방안을 마련하기 위한 본 연구의 목적에 따라, 교육과정에 대한 성취도를 파악하는 평가를 분석 대상으로 삼았다. 우리나라에서 시행되는 교육과정 기반의 평가는 NAEA가 대표적이라 할 수 있다. 전국 규모로 시행되는 NAEA는 교육과정 질 관리를 목적으로 국가 수준의 교육과정에 근거한 평가 도구를 개발하여 교과별 성취기준에 학생들이 도달한 정도를 평가하고 모니터링한다. 이에 본 연구는 NAEA 문항의 특성을 분석함으로써 교육과정을 기반으로 하는 학생 평가 타당도 제고를 위한 방안으로서 문항 특성 분석(FA) 기법의 활용 가능성을 살펴보았다. NAEA 중학교 문항 중 가장 최근에 공개된 2019년 국어 및 수학 교과와 G형 검사지 공개 문항을 본 연구의 분석 대상으로 설정하였다. 2019년 NAEA 평가틀에 따라 중학교 교과별 문항은 2009 개정 교육과정을 기반으로 출제되었으며, 교과별 평가틀과 분석 대상 문항 수는 <표 2>와 같다.

<표 2> 2019년 NAEA 국어/수학 교과와 중학교 평가틀 및 공개 문항 수

구분	국어	수학
내용영역	화법/독서/작문/문법/문학 •화법 지식, 사실적 의사소통, 추론적 의사소통, 비판적 의사소통 •독서 지식, 내용확인, 추론, 평가와 감상 •작문 지식, 내용 생성, 내용 조직, 표현과 고쳐 쓰기 •음운과 어휘, 문장과 담화, 국어 규범, 국어의 변천 •문학의 수용, 문학의 생산, 문학의 생활화	수와 연산/문자와 식/함수/확률과 통계/기하 •계산 •이해 •추론 •문제해결
평가범위	중학교 1~3학년군 과정 범교과 소재	중학교 '수학1'과 '수학2' 전 범위, '수학3'의 '제곱근과 실수', '근호를 포함한 식의 계산', '다항식의 인수분해'
문항 수	(선다형) 28문항 (서답형) 6문항	(선다형) 29문항 (서답형) 4문항(하위 문항 기준 9문항)

* 출처: 구자옥 외(2019b, pp. 11-13), 동효관 외(2018, pp. 20-21)

2. 분석 절차와 방법

2019년 NAEA 문항을 대상으로 문항 특성 분석(FA) 기법을 적용하기 위해서 문항에 대한 다양한 측면의 정보를 체계화한 분석틀을 개발하였다. 교육과정 목표와 평가 목표 간 연계성과 평가의 실제성을 고려하여 문항 특성 분석틀 초안을 개발하고 2차에 걸친 전문가 검토와 수정을 거쳐 최종 분석틀을 확정하였다. 국어 교과는 교육과정 연계, 인지 특성, 지문 특성, 문항 유형과 과제 복잡도 등의 4개 범주, 수학 교과는 수학 개념, 수학 역량, 표현 형식, 문항 유형 및 점수 체제 등의 4개 범주로 구성된 분석틀을 구성하였다.

다음으로 문항 특성 분석틀을 기초로 교과별 문항에 대한 평정 자료를 산출하였다. 문항 특성 평정은 교과별로 NAEA 출제 경험이 있는 국어/수학 교사 3인에 의해 이루어졌으며, 각각의 문항에 문항 특성이 포함되는 경우 1, 포함되지 않는 경우 0으로 평정하도록 하였다. 평정자간 일치율 산출, 의견 차이 조정을 위한 논의 및 재평정, 교과 전문가 검토 등을 통해 최종 분석틀에 따른 평정 결과를 얻었다.

NAEA 문항의 평정 결과를 바탕으로 문항 특성별 출제 경향과 평가 자료에 나타난 학생들의 성취 특성을 설명하는 문항 특성을 파악하였다. 먼저 문항 특성별 출제 경향은 평정 결과에 대한 빈도 분석을 실시하여 살펴보았다. 학생 성취에 대한 문항 수준의 정보라 할 수 있는 정량화한 문항 특성인 난이도와 변별도를 활용하여 분석틀의 문항 특성과 학생 성취 간 관련성을 탐색하였다. 이를 위해 문항 난이도 추정은 정답률을 활용하였고, 문항 변별도는 문항점수와 검사총점 간 상관계수로 추정하였다. 문항 특성을 지닌 문항(1로 평정)과 그렇지 않은 문항(0으로 평정)에 대해 각각의 평균 난이도와 평균 변별도를 산출하여 문항 특성 유무에 따른 난이도와 변별도의 평균 차이를 살펴보고, 다음과 같은 다중선형회귀모형(Multiple Linear Regression model)을 범주별로 적용하여 문항 난이도와 변별도에 미치는 문항 특성의 영향력을 분석하였다²⁾.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + e_i, e_i \sim N(0, \sigma^2)$$

Y_i : i 번째 문항의 난이도 또는 변별도

X_{ki} : i 번째 문항의 k 번째 문항 특성에 대한 평정값(1또는 0)

β_k : k 번째 문항 특성의 문항 난이도에 대한 영향력

문항 특성 분석틀의 범주 구성이 상호배타적인 속성을 갖는 특성으로 구성된 경우에는 해당 범주의 한 특성을 기준(reference level)으로 다른 특성을 더미(dummy) 변수 처리하여 회귀모형에 투입하였고, 문항 특성 간 완벽한 정적 혹은 부적 상관을 갖는 경우는 완전공선성(complete

2) 문항 난이도의 추정과 문항 특성의 영향력 분석은 문항반응이론(item response theory, IRT)의 선형로지스틱검사 모형(linear logistic test model, LLTM; random-effect LLTM) 등을 통해서도 확인해 볼 수 있지만, 문항반응자료가 있어야만 모형을 적용할 수 있을 뿐만 아니라 난이도만 고려하는 한계가 있음. 문항반응자료의 가용성, 관심 있는 종속변수의 다양성(난이도, 변별도 등) 측면에서의 실용도를 고려하여 본 연구에서는 문항 난이도와 변별도 추정 후에 다중선형회귀모형을 적용하는 2단계 순차적 분석 방법을 선택함.

collinearity) 문제가 발생하지 않도록 한 가지 특성만 모형에 투입하였다. 각 문항 특성의 상대적 영향력 확인이 가능한 표준화 회귀계수를 산출하였고, 그 절대값과 영향력은 비례하는데, 더미 변수에 대한 표준화 회귀계수는 기준 특성 대비 효과를 의미한다.

IV. NAEA 문항의 특성 분석(FA) 결과

1. 국어

가. 문항 특성 분석들에 따른 평정 결과

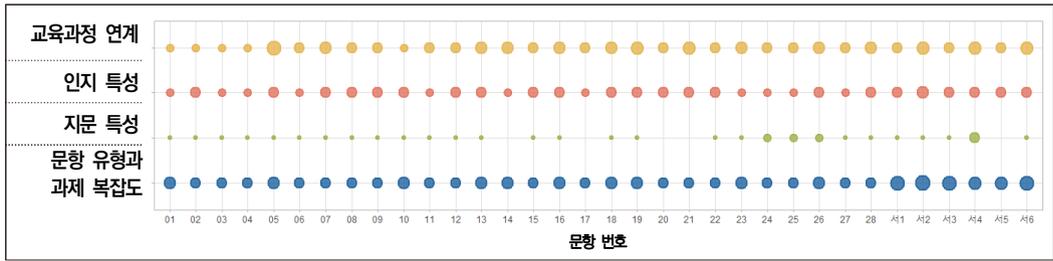
2019년 NAEA 중학교 수학 문항의 교육과정 연계, 인지 특성, 지문 특성, 문항 유형과 과제 복잡도 범주 특성에 대한 빈도 분석 결과는 <표 3>과 같다. 내용 영역은 문학 영역의 비율이 가장 높았으며, 비판적·창의적 역량과 의사소통 역량이 상당히 높은 반면 자기 성찰·계발 역량은 거의 다루어지지 않았다. 인지 유형은 이해 또는 분석에 분포하였고, 문제 (해결) 상황 유형은 회상 및 재구성이 압도적으로 많았다. 지문 특성은 단일 매체의 비율이 매우 높은 반면 복수 구성 또는 복합 구성 지문의 비율은 낮아, 대체로 단일 매체의 지문 1개를 활용하는 구조를 취하고 있었다. 문항 유형의 경우, 선택형 문항에 치중되어 있고, 단답형과 서술형 모두 부분점수가 부여되는 유형이었다. 보기(자료)를 활용하는 문항이 절반 이상이었으며, 지문 내용의 적용 상황 설정 유형이 가장 많았고, 추가 지문 제시 유형은 전혀 출제되지 않았다. 과제 수행 맥락은 학문적 탐구 상황 맥락이 상대적으로 많았고, 지문 조회 범위는 지문 전체를 조회하도록 하는 문항이 가장 많았고, 약 18%의 문항에 답안 작성 조건을 제시하였으며, 그 중 절반은 찾아 쓰기를 요구하였다.

<표 3> 2019년 NAEA 중학교 국어 교과 문항 특성의 분포

대범주	중범주	문항 특성	빈도(개)/비율(%)
교육 과정 연계	내용 영역	화법	5 / 14.7
		작문	4 / 11.8
		독서	8 / 23.5
		문법	5 / 14.7
		문학	12 / 35.3
	국어과 역량	비판적·창의적 역량	25 / 73.5
		자료·정보 활용 역량	12 / 35.3
		의사소통 역량	28 / 82.4
		공동체·대인 관계 역량	8 / 23.5
		문화 향유 역량	10 / 29.4
	자기 성찰·계발 역량	1 / 2.9	

대범주	중범주	문항 특성	빈도(개)/비율(%)		
인지 특성	인지 유형	기억	1	2.9	
		이해	13	38.2	
		적용	1	2.9	
		분석	16	47.1	
		평가	3	8.8	
		창안	0	0.0	
	문제 (해결) 상황 유형	회상 및 재구성	33	97.1	
		오류와 갈등 분석	2	5.9	
		절차적 탐구	3	8.8	
		전략적 사고와 의사결정	3	8.8	
논리와 추론		16	47.1		
지문 특성	매체	설계	1	2.9	
		단일 매체	28	82.4	
	지문 구성	복합 매체	1	2.9	
		복수 구성	4	11.8	
문항 유형과 과제 복잡도	문항 유형 및 점수 체제	복합 구성	1	2.9	
		문항(과제) 유형	선택형	28	82.4
			단답형	4	11.8
			서술형	2	5.9
	점수 체제	부분 점수	6	17.6	
	과제 수행 맥락	일상 및 학교생활 맥락	11	32.4	
		(미래) 직무 수행 맥락	0	0.0	
		학문적 탐구 상황 맥락	23	67.6	
	보기(자료) 유형	개념 정의 및 설명	5	14.7	
		추가 지문 제시	0	0.0	
지문 내용 적용 상황 설정		10	29.4		
지문에 대한 학생 반응 설정		6	17.6		
지문 조회 범위	특정 단어	1	2.9		
	특정 문장	4	11.8		
	특정 문단	7	20.6		
	지문 전체	16	47.1		
답안 작성 조건 유형	형식(길이 등) 및 규범 준수 요구	1	2.9		
	찾아 쓰기 요구	3	8.8		
	재구성하여 쓰기 요구	2	5.9		

NAEA 중학교 국어 문항별로 네 가지 문항 특성 범주에 포함된 특성들이 나타난 빈도를 도식화하면 [그림 2]와 같다. 그림의 원이 클수록 빈도가 높게 나타남을 의미하는데, 개별 문항은 각 범주의 특성을 다양하게 가지고 있는 편이며, 지문 특성은 상대적으로 다양하게 분포하지 않음을 알 수 있다. 다양한 문항 특성을 지닌 문항들은 서답형으로, 선다형 문항 28개의 평균 특성 수는 10.8개, 서답형 문항 6개의 평균 특성 수는 16.1개로 나타났다. 문항에 포함된 특성 수가 많을수록 다양한 능력을 평가하는 데 유용할 수 있다. 따라서 다수의 문항 특성을 가진 문항이 서답형이라는 점은 국어과 문항 유형의 방향성을 실증적으로 보여주는 결과이다.



* 원의 크기에 따른 특성의 빈도: ● 1 ● 2 ● 3 ● 4 ● 5

[그림 2] 2019년 NAEA 중학교 국어 교과 문항 특성 범주의 문항별 분포

나. 문항 특성과 학생 성취 간 관계

NAEA 중학교 국어 문항에 대해 4개 범주의 문항 특성이 정답률과 변별도에 미치는 영향을 분석한 결과는 <표 4>~<표 7>과 같다. <표 4>에서 교육과정 연계 범주의 문항 특성이 미치는 영향을 살펴보면, 내용 영역 범주에서 화법 특성은 타 영역보다 정답률을 높이고, 변별도를 낮추는데 가장 큰 영향을 주며, 역량 범주에서 의사소통 역량은 정답률에 미치는 영향이 크지 않은 반면, 변별도를 높이는 것으로 나타났다. 다른 특성이 통제된 상태에서 변별도를 높이는 데 가장 영향이 큰 특성은 자료·정보 활용 역량으로 분석되었다. 교육과정 연계 범주(내용 영역, 국어과 역량)의 특성은 정답률과 변별도를 24~36% 가량 설명하며, 정답률에 대한 설명력이 변별도에 대한 설명력보다 다소 높았고, 내용 영역 범주보다 국어 교과 역량 범주가 설명하는 정도가 많았다.

<표 4> 국어 교육과정 연계 범주의 문항 특성이 정답률/변별도에 미치는 영향

중범주	특성	정답률		변별도	
		평균 차이*	표준화 회귀계수	평균 차이*	표준화 회귀계수
내용 영역†	화법	18.4	0.428	-0.13	-0.406
	작문	-4.3	-0.130	0.11	0.225
	독서	-5.7	-0.189	0.03	0.021
	문법	1.2	-	0.02	-
	문학	-4.3	-0.152	-0.01	-0.088
		회귀모형의 $R^2 = 0.308$		회귀모형의 $R^2 = 0.243$	
국어 교과 역량	비판적·창의적 역량	-10.3	-0.330	0.07	0.205
	자료·정보 활용 역량	-2.4	0.108	0.10	0.583
	의사소통 역량	8.8	0.090	-0.01	0.298
	공동체·대인 관계 역량	-3.0	-0.223	0.01	-0.058
	문화 향유 역량	-10.7	-0.379	0.01	-0.024
	자기성찰·계발 역량	14.4	0.230	-0.06	-0.262
		회귀모형의 $R^2 = 0.360$		회귀모형의 $R^2 = 0.313$	

* 각 특성에 해당하는 문항(평정=1)과 해당하지 않는 문항(평정=0)의 평균 정답률/변별도 차이를 의미함.

† 내용 영역 범주의 경우, '문법' 기준의 더미코딩으로 회귀계수는 '문법' 대비 영향력을 나타냄.

인지 특성 범주의 문항 특성은 <표 5>에 제시하였듯이 정답률과 변별도를 11~23% 가량 설명하며, 인지 유형 범주의 설명력에 비해 문제 상황 유형 범주의 설명력이 더 컸다. 인지 유형 범주는 변별도보다 정답률을 미미하게 더 설명하는 반면, 문제 상황 유형 범주는 정답률보다 변별도를 더 설명하는 것으로 분석되었다. 인지 유형에서는 기억 유형이 문항의 정답률을 높이고, 변별도를 낮추는 데 가장 큰 영향을 주는 것으로 나타났다. 문제 상황 유형에서는 절차적 탐구 유형이 변별도에 가장 큰 영향을 주는 특성이었다.

<표 5> 국어 인지 특성 범주의 문항 특성 특성이 정답률/변별도에 미치는 영향

중범주	특성	정답률		변별도	
		평균 차이*	표준화 회귀계수	평균 차이*	표준화 회귀계수
인지 유형†	기억	21.5	-	-0.16	-
	이해	-1.7	-0.876	-0.02	0.618
	적용	14.4	-0.096	-0.06	0.138
	분석	-2.5	-0.913	0.01	0.714
	평가	-0.1	-0.489	0.08	0.571
		회귀모형의 $R^2 = 0.134$		회귀모형의 $R^2 = 0.106$	
문제(해결) 상황 유형	회상 및 재구성	10.9	0.117	0.02	0.023
	오류와 갈등 분석	11.6	0.123	-0.02	0.036
	절차적 탐구	-2.7	-0.112	0.18	0.481
	전략적 사고와 의사결정	-5.4	-0.223	-0.01	0.084
	논리와 추론	-5.6	-0.279	0.01	0.157
	설계	-5.7	-0.142	0.03	0.105
		회귀모형의 $R^2 = 0.156$		회귀모형의 $R^2 = 0.230$	

* 각 특성에 해당하는 문항(평정=1)과 해당하지 않는 문항(평정=0)의 평균 정답률/변별도 차이를 의미함.

† 인지 유형 범주의 경우, '기억' 기준의 더미코딩으로 회귀계수는 '기억' 대비 영향력을 나타내며, '창안'은 해당 문항이 없어 분석에서 제외함.

지문 특성 범주의 문항 특성이 정답률과 변별도에 미치는 영향을 분석한 결과, <표 6>에 제시된 바와 같이 복합 매체 유형이 다른 특성에 비해 영향이 컸으며, 정답률과 변별도를 높이는 영향을 주는 것으로 나타났다. 지문 특성 범주는 정답률과 변별도를 2~15% 가량 설명하는 것으로 분석되었다. 정답률과 변별도에 대한 지문 특성 범주의 설명력은 교육과정 연계 범주, 인지 특성 범주에 비해 상대적으로 적었다.

〈표 6〉 국어 지문 특성 범주의 문항 특성이 정답률/변별도에 미치는 영향

중범주	특성	정답률		변별도	
		평균 차이*	표준화 회귀계수†	평균 차이*	표준화 회귀계수
매체	단일 매체	1.4	0.095	-0.05	-0.087
	복합 매체†	6.5	0.141	0.23	0.218
구성	(지문 수) 복수 구성	1.1	-0.031	0.11	0.197
	(장르나 매체) 복합 구성†	6.5	0.141	0.23	0.218
		회귀모형의 $R^2 = 0.016$		회귀모형의 $R^2 = 0.152$	

* 각 특성에 해당하는 문항(평정=1)과 해당하지 않는 문항(평정=0)의 평균 정답률/변별도 차이를 의미함.

† '복합 매체'와 '복합 구성' 특성은 동일하여(평정값 일치), 중범주(매체, 구성)를 통합하여 회귀분석을 실시하고, 두 특성 중에 1개만 회귀모형에 투입함.

문항 유형과 과제 복잡도 범주의 문항 특성이 정답률 및 변별도에 미치는 영향을 분석한 결과는 〈표 7〉과 같다. 문항 유형과 점수 체제 범주의 경우 정답률에 대해서는 특성별 차이가 크지 않았지만 변별도에 대해서는 단답형이 선다형보다 큰 영향을 주는 것으로 나타났다. 과제 수행 맥락은 학문적 탐구 상황 맥락 특성이 정답률을 낮추는 효과가 있었고, 보기(자료) 유형 중에는 개념 정의 및 설명 유형이 정답률을 낮추는 데 큰 영향을 주고, 지문에 대한 학생 반응 설정 유형이 정답률을 높이는 데 큰 영향을 주며, 지문 내용 적용 상황 설정은 변별도를 높이는 효과가 있었다. 지문 조회 유형 중에는 특정 단어 유형이 정답률을 높이는 데 가장 큰 영향을 주고, 지문 전체를 조회하는 유형은 변별도를 낮추는 효과가 있었다. 답안 작성 조건은 변별도를 높이는 영향을 주며, 재구성하여 쓰기 유형이 정답률을 높이는 데 가장 큰 영향을 주는 것으로 나타났다. 문항 유형과 과제 복잡도 범주는 문항의 정답률은 거의 설명하지 못하고, 변별도에 대해서는 과제 수행 맥락 범주를 제외하고, 15 ~ 34% 정도 설명하는 것으로 분석되었다.

〈표 7〉 국어 문항 유형과 과제 복잡도 범주의 문항 특성이 정답률/변별도에 미치는 영향

중범주	특성	정답률		변별도	
		평균 차이*	표준화 회귀계수†	평균 차이*	표준화 회귀계수†
문항 유형과 점수 체제†	선택형(=-부분점수)	-2.2	-	-0.16	-
	단답형	1.3	0.041	0.19	0.570
	서술형	3.4	0.069	0.06	0.180
		회귀모형의 $R^2 = 0.016$		회귀모형의 $R^2 = 0.338$	
과제 수행 맥락‡	일상 및 학교생활 맥락	7.8	0.300	-0.01	-0.048
	(=-학문적 탐구 상황 맥락)	회귀모형의 $R^2 = 0.090$		회귀모형의 $R^2 = 0.002$	
보기(자료) 유형	개념 정의 및 설명	-7.3	-0.182	-0.04	-0.069
	지문 내용 적용 상황 설정	-1.5	-0.048	0.09	0.365
	지문에 대한 학생 반응 설정	7.1	0.183	0.03	0.125
		회귀모형의 $R^2 = 0.082$		회귀모형의 $R^2 = 0.149$	

중범주	특성	정답률		변별도	
		평균 차이*	표준화 회귀계수†	평균 차이*	표준화 회귀계수†
지문 조회	특정 단어	13.4	0.198	0.07	0.031
	특정 문장	-0.6	0.016	0.04	-0.039
	특정 문단	-6.3	-0.127	0.07	0.026
	지문 전체	3.7	0.128	-0.10	-0.424
		회귀모형의 $R^2 = 0.083$		회귀모형의 $R^2 = 0.187$	
답안 작성 조건 유형	형식 및 규범 준수 요구	4.4	0.065	0.24	0.385
	찾아 쓰기 요구	-3.8	-0.071	0.13	0.360
	재구성하여 쓰기 요구	9.0	0.172	0.12	0.297
		회귀모형의 $R^2 = 0.040$		회귀모형의 $R^2 = 0.324$	

* 각 특성에 해당하는 문항(평정=1)과 해당하지 않는 문항(평정=0)의 평균 정답률/변별도 차이를 의미함.
 † 문항 유형과 점수 체제 범주의 경우, '선택형' 기준의 더미코딩으로 회귀계수는 '선택형' 대비 영향력을 나타내며, '부분점수'는 '선택형' 과 부적으로 일치하여(평정값 정반대), 회귀모형에 두 특성 중에 1개만 투입함.
 ‡ 과제 수행 맥락 범주의 경우, (미래) 직무 수행 맥락은 해당 문항이 없어 분석에서 제외되었고, '일상 및 학교 생활 맥락'과 '학문적 탐구 상황 맥락' 특성은 부적으로 일치하여(평정값 정반대), 회귀모형에는 두 특성 중에 1개만 투입함.

2. 수학

가. 문항 특성 분석들에 따른 평정 결과

2019년 NAEA 중학교 수학 문항의 수학 개념, 수학 역량, 과제 유형, 문항 유형 및 점수 체제 범주 특성에 대한 빈도 분석 결과는 <표 8>과 같다. 수학 개념 특성은 연산법칙, 다항식, 일차함수, 그래프(함수 영역)의 비율이 높은 편이었으며, 각각의 개념 특성은 0~4개의 문항에서 활용된 것으로 나타나 NAEA에서는 특정 수학 개념을 다루지 않는 경향이 있음을 확인할 수 있다.

<표 8> 2019년 NAEA 중학교 수학 교과 문항 특성의 분포

대범주	중범주	문항 특성	빈도(개)/비율(%)		문항 특성	빈도(개)/비율(%)	
수학 개념	수의 체계	소인수분해	2	5.3	순환소수	1	2.6
		정수	0	0.0	제곱근	2	5.3
		유리수	2	5.3	무리수	0	0.0
	수와 식의 연산	연산법칙	4	10.5			
	다항식	다항식	4	10.5	인수분해	1	2.6
		지수법칙	1	2.6			
	방정식과 부등식	방정식	1	2.6	연립방정식	1	2.6
		부등식	3	7.9	이차방정식	0	0.0
	함수와 그래프	좌표평면	0	0.0	일차함수	4	10.5
		정비례와 반비례	1	2.6	이차함수	0	0.0
함수의 개념		1	2.6	그래프(함수영역)	4	10.5	
평면도형	점/선/면	1	2.6	사각형의 성질	2	5.3	

대범주	중범주	문항 특성	빈도(개)/비율(%)		문항 특성	빈도(개)/비율(%)	
수학 역량		각	1	2.6	도형의 닮음	2	5.3
		삼각형의 합동	2	5.3	피타고라스 정리	0	0.0
		부채꼴	1	2.6	삼각비	0	0.0
		삼각형의 성질	1	2.6	원의 성질	0	0.0
	입체도형	입체도형	1	2.6			
	통계	자료의 정리와 해석	2	5.3	산포도	0	0.0
		대푯값	0	0.0	상관관계	0	0.0
	확률	경우의 수	1	2.6	확률	2	5.3
	문제 해결	문제 해결 전략	4	10.5	수학적 모델링	4	10.5
	추론	관찰과 추측	3	7.9	논리적 추론	9	23.7
창의·융합	창의적 사고	0	0.0	융합적 사고	6	15.8	
의사소통	수학적 표현	6	15.8	수학적 소통	0	0.0	
정보 처리	자료와 정보 처리	2	5.3	공학적 도구 및 교구 활용	1	2.6	
태도 및 실천	수학적 태도	0	0.0	수학적 시민의식	1	2.6	
표현 형식	수학적 표현	식	13	34.2	그래프/도식	4	10.5
		표	4	10.5	도형	10	26.3
	문항 제시 형식	수학기호(표기)로만 구성	3	7.9	시각적 자료	8	7.9
문항 유형 및 점수 체제	문항 유형	문장제	10	26.3	단서 포함	7	26.3
		선택형	29	76.3			
		단답형	8	21.1			
	점수 체제	서술형	1	2.6			
		부분 점수 여부	4	10.5			

수학 역량 특성은 논리적 추론의 비율이 24%로 가장 높았고, 수학적 표현과 융합적 사고도 16%로 높아 빈번하게 출제되는 편임을 나타내었으나 대규모 지필평가에서 측정하는 데 한계가 있는 창의적 사고나 수학적 소통을 갖는 문항은 없었다. 또한 수학 교과 특성 상 표현 형식 특성 중식의 비율이 34%로 가장 높았고, 지시문이 수학 기호(표기)로만 구성된 문항은 적었고, 문장제가 26%의 높은 비율로 상대적으로 많이 출제된 경향을 보였으며, 문항 유형 및 점수 체제 특성은 NAEA의 출제 방식에 따라 선택형의 비율이 높게 나타났다.

NAEA 중학교 수학의 네 가지 문항 특성 범주에 포함된 특성들이 개별 문항에 나타난 빈도를 도식화하면 [그림 3]과 같다. 원의 크기와 빈도는 비례하는데, 전체 문항에 대해 표현 형식 범주와 수학 역량 범주는 특성이 다양하게 분포하는 반면, 수학 개념 범주와 문항 유형 및 점수 체제 범주는 대다수의 문항에 포함된 특성이 많지 않은 것을 알 수 있다. 특히 마지막 문항에 해당하는 서답형 중 단답형과 서술형이 복합된 문항에서 수학 역량과 표현 형식이 더 다양하게 활용되었음을 확인할 수 있다.

〈표 10〉 수학 역량 범주의 문항 특성이 정답률/변별도에 미치는 영향

종범주	문항 특성	정답률		변별도	
		평균 차이*	표준화 회귀계수 [†]	평균 차이*	표준화 회귀계수 [†]
문제 해결	문제 해결 전략	-0.3	-0.103	-0.04	-0.676
	수학적 모델링	-4.7	-0.139	0.03	-0.232
추론	관찰과 추측	-5.1	0.026	-0.08	-0.162
	논리적 추론	-0.4	-0.004	-0.04	-0.091
창의·융합	융합적 사고	-0.2	0.190	0.01	0.430
의사소통	수학적 표현	-7.7	-0.195	0.07	0.156
정보 처리	자료와 정보 처리	-8.6	-0.154	-0.18	-0.340
	공학적 도구 및 도구 활용	-24.7	-0.282	-0.15	-0.147
태도 및 실천	수학적 시민의식	-20.9	-0.261	0.15	0.337
		회귀모형의 $R^2 = 0.182$		회귀모형의 $R^2 = 0.375$	

* 각 특성에 해당하는 문항(평균=1)과 해당하지 않는 문항(평균=0)의 평균 정답률/변별도의 차이를 의미함.

표현 형식 범주의 문항 특성이 정답률과 변별도에 미치는 영향은 〈표 11〉과 같이 분석되었다. 표현 범주 특성은 모두 정답률을 낮추는 경향이 있었으며, 특히 그래프/도식이 가장 큰 영향을 주는 것으로 나타났다. 또한 시각적 자료 특성이 변별도를 높이는 데 가장 큰 영향을 주고, 문장제 특성이 변별도를 낮추는 데 가장 큰 영향을 주는 것으로 나타났다. 표현 형식 범주는 정답률과 변별도를 각각 24%, 26%를 설명하는 것으로 나타났으며, 수학 교과의 네 가지 범주 중 정답률에 대한 설명력이 가장 높았다.

〈표 11〉 수학 표현 형식 범주의 문항 특성이 정답률/변별도에 미치는 영향

종범주	문항 특성	정답률		변별도	
		평균 차이*	표준화 회귀계수 [†]	평균 차이*	표준화 회귀계수 [†]
(문항에 포함된) 수학적 표현	식	0.8	-0.193	-0.04	-0.357
	표	-8.6	-0.177	0.00	0.071
	그래프/도식	-13.3	-0.409	0.08	0.104
	도형	-0.5	-0.222	-0.02	-0.258
문항 제시 형식	수학 기호(표기)로만 구성	2.5	-0.036	-0.04	-0.010
	문장제	-3.8	-0.111	-0.03	-0.520
	시각적 자료	-11.4	-0.351	0.05	0.239
	단서 포함	0.0	-0.130	-0.05	-0.257
		회귀모형의 $R^2 = 0.237$		회귀모형의 $R^2 = 0.259$	

* 각 특성에 해당하는 문항(평균=1)과 해당하지 않는 문항(평균=0)의 평균 정답률/변별도의 차이를 의미함.

마지막으로 문항 유형과 점수 체제 범주의 문항 특성이 정답률 및 변별도에 미치는 영향을 분석한 결과는 〈표 12〉에 제시하였다. 문항 유형과 점수 체제 범주 특성 중 서술형 특성이 정답률을 낮추는 데 가장 큰 영향을 주는 것으로 나타났고, 이 범주에 해당하는 문항과 해당하지 않는 문항의 평균 정답률 차이는 -21%p로 매우 컸다. 또한 단답형 특성이 변별도를 높이는 데 가장 큰 영향을 주는 것으로 나타났다. 문항 유형과 점수 체제 범주는 정답률을 거의 설명하지 못하는 것으로 나타난

반면, 변별도는 50% 가량 설명하는 것으로 나타나 네 가지 범주 중 변별도에 대한 설명력이 가장 높았다.

〈표 12〉 수학 문항 유형과 점수 체제 범주의 문항 특성이 정답률/변별도에 미치는 영향

종범주	문항 특성	정답률		변별도		
		평균 차이*	표준화 회귀계수†	평균 차이*	표준화 회귀계수†	
문항 유형	선택형	6.5	-	-0.18	-	
	단답형	-3.9	-0.194	0.17	0.622	
	서술형	-20.9	-0.297	0.15	0.217	
점수 체제	부분 점수 여부	-3.8	0.160	0.17	0.100	
			회귀모형의 $R^2 = 0.068$		회귀모형의 $R^2 = 0.498$	

* 각 특성에 해당하는 문항(평정=1)과 해당하지 않는 문항(평정=0)의 평균 정답률/변별도의 차이를 의미함.

† '선택형'은 독립변수 간 공선성 문제로 다중회귀모형에서 제외됨.

V. 결론 및 제언

학생 평가가 학교 교육의 성과에 대해 점검할 뿐만 아니라 학생의 성장을 견인하는 데 중요한 역할을 하는 기제로 의미있게 쓰려면 평가에 대한 타당도 확보가 요구된다. 평가 문항의 특성을 정성적·정량적 방법에 의해 분석하는 문항 특성 분석(FA) 기법은 타당도를 입증할 경험적 자료(evidence)를 수집하며, 이를 평가 문항의 수정 및 교수학습의 피드백 자료로 활용하는 것을 주목적으로 한다. 본 연구에서는 2019년 중학교 NAEA 국어, 수학 평가도구에 문항 특성 분석(FA) 기법을 적용하여 교과별 문항 특성을 살펴본 결과, 교육과정과 출제 계획(III장 1절 참조)에 근거하여 타당하게 출제가 이루어지고 있으며, 특히 교육과정과의 연계가 적절하게 이루어지고 있음을 확인할 수 있었다. 이를 통해 우리나라 교육과정 기반 문항 특성 분석틀을 국가 수준에서 시행되는 교육과정 기반 평가의 타당도 점검을 위한 도구로 활용할 수 있음을 보이고, 문항 특성 분석(FA) 결과를 통해 평가 계획 단계에서 의도한대로 평가도구가 제작되었는지 반성적으로 평가할 수 있음을 밝혔다.

2019년 중학교 NAEA 교과별 문항 특성을 분석한 결과, 국어과 인지 특성 범주에서 출현 빈도가 낮은 특성이 존재하는 것은 평가 시간 제한이 있는 대규모 평가의 특성과 중학생의 일반적인 사고 특성을 고려하여 문항 풀이에 소요되는 시간과 고차적인 사고가 필요한 인지 특성들이 포함되지 않는 경향이 있는 것으로 보인다. 수학의 경우 지시문이 문장으로 구성된 문항이 약 4분의 1정도로 나타났는데, 2020년 평가부터 중학교 문항에 수학 역량이 반영되면서 실생활 맥락을 다루는 문장제 특성의 비중이 더욱 증가할 것으로 보인다. 국어와 수학 모두 출제 계획에 따라 선다형보다 서답형이 적은 수의 문항이 출제되는데, 서답형 문항에 해당하는 특성의 수가 더 많은 결과가 나타나 다양한 특성을 서답형 문항 유형에서 다루고 있음을 알 수 있다. 이러한 결과는 최근의 역량

평가와 서·논술형 평가가 강조되는 평가 정책과 맥을 같이 한다.

학생 평가 문항에 대한 새로운 분석 방법을 적용함으로써 본 연구는 학생 성취 결과에 대한 해석과 활용 다양화 가능성을 확인하였다. 국어과 및 수학과 문항 특성별로 난이도와 변별도에 미치는 영향 분석을 통해 학생 성취 특성을 세부적으로 살펴보았다. 정답률을 높이는 특성이나 낮추는 특성은 각각 어떤 특성에 대한 학생들의 성취가 높거나 낮은지를 나타내며, 변별도를 높이는 특성이나 낮추는 특성은 각각 성취수준이 높은 학생과 낮은 학생 간 성취도 차이가 크거나 낮은지를 나타내는 것이다. 이러한 분석을 통해 문항 특성 분석(FA) 기법이 평가 도구의 타당도 점검뿐만 아니라 평가 결과의 환류 측면에서도 유용성이 있음을 알 수 있었다. 성취기준의 숙달에 어려움을 주는 요인을 파악하여 차후 평가 도구 개발의 기초 자료, 교수학습 개선을 위한 피드백 자료로 활용할 수 있으며, 나아가 교육과정 개정을 위한 자료로도 활용할 수 있다. 다만 특정 연도에 국한한 평가도구를 분석하였기 때문에 대규모 평가 결과라 할지라도 본 연구의 분석 대상 평가 문항 수가 갖는 한계를 고려하여 분석 결과를 해석해야 함에 유의해야 한다.

본 연구에서 보여준 바와 같이, 문항 특성 분석(FA) 기법을 실제 문항에 적용한 결과는 평가 도구의 타당도에 관한 다양한 증거를 수집하는 분석 방안의 하나로 쓰일 수 있다. 이러한 맥락에서 본 연구는 학생 평가의 타당도 관련 연구 방법론을 확장해 나가는 의의가 있다. 앞으로 본 연구에서 시도한 분석이 체계적·지속적으로 이루어질 때 문항 특성 분석(FA)의 유의미성이 심화될 것이다. 이 때 본 연구 결과에서 교과마다 내용 영역이나 중점을 두는 문항 특성에 차이가 있었음을 고려하여 평가 목적과 평가 도구의 용도에 맞게 분석틀을 재구성할 뿐만 아니라 국어, 수학 교과 이외 타 교과에 대해서도 분석틀의 전형을 개발하고 실제 문항 분석을 적용하는 연구가 후속될 필요가 있다. 최근 학생 평가가 컴퓨터 기반 평가 체제로 전환하고 있는데, 평가 시행 체제에 따라 문항 특성 요소를 설정하거나 새 시대에 요구되는 역량 평가에 포함해야 할 문항 특성 요소를 선별하여 분석틀을 구성해야 한다. 본 연구에서 활용한 문항 특성 분석틀의 구성을 살펴보면, 평가틀에 규정되어야 하는 요소와 문항 특성 분석틀을 위해 규명되어야 할 문항 특성이 상호 공유되는 부분이 많음을 알 수 있다. 본 연구는 기 개발된 평가 도구와 실시된 평가 자료에 문항 특성 분석(FA) 기법을 실험적·예시적으로 적용하였기 때문에, 평가 종료 이후에 문항 특성을 규명한 다음 평정을 진행하였지만, 평가 계획을 수립할 때 평가틀과 함께 문항 특성 분석틀을 개발하는 것이 바람직하다.

정량적 분석의 통계모형은 본 연구에서 활용한 모형 외에도 분석 자료와 내용에 적합한 모형을 선택하여 유연하게 활용할 수 있다. 본 연구에서 실험적으로 적용한 NAEA와 같은 대규모 지필평가뿐만 아니라 평가 규모나 평가 방식이 다른 단위학교의 지필평가와 수행평가, 대규모 컴퓨터 기반 평가 등에도 문항 특성 분석(FA) 기법을 적용해 보고, 분석틀 구성과 통계 분석 모형 등의 적합성, 유연성, 확장성 등을 다각적으로 검토해 볼 필요가 있다. 예를 들어, 문항 특성의 영향에 대한 분석 모형은 실제 적용 용이성을 고려하여 다중선형회귀모형을 선택하였지만, 측정 이론과 통계 모형의 정교성에 따라 LLTM/RE-LLTM, 인지진단모형, 랜덤포레스트 기법 등의 다른 방법론에 의해 분석할 수 있다. 또한 통계적 분석에 익숙하지 않은 교사를 위해 문항 특성에 대한 정성적 평정과 정량적 분석을 통합한 문항 특성 분석(FA) 기법의 활용 지침이나 나이스와 연동 가능한 프로그램이 개발·제공된다면 학교 현장에서 실증적인 타당도 점검이 활성화 될 수 있을 것으로 기대한다.

참고문헌

- 고호경, 이현숙(2007). 고등학교 수리영역 시험의 난이도 예측 요인 분석. **한국학교수학회 논문집**, 10(1), 113-127.
- 교육부(2018). **모두를 포용하는 사회 미래를 열어가는 교육**. 2019 교육부 업무보고 (2018.12.11.).
- 교육부(2020a). **학교생활기록 작성 및 관리지침**. 교육부 훈령 321호[시행 2020.3.1.].
- 교육부(2020b). **2020 학교생활기록부 기재요령-중·고등학교**. 교육부(2020.2.).
- 구자옥, 김경주, 김준식, 박상복, 박지현, 성경희, 이광상, 이소라, 이재봉, 장의선, 강종훈, 신명경, 심규철, 최원호(2019a). **국가수준 학업성취도 평가 결과에 기반한 2009 개정 교육과정의 학업성취 특성 및 추이 분석**. 한국교육과정평가원 연구보고 RRE 2019-7.
- 구자옥, 김경주, 김준식, 박상복, 박지현, 성경희, 이광상, 이소라, 이재봉, 장의선, 강종훈, 신명경, 심규철, 최원호, 박수민, 박혜민, 주현욱, 홍경화(2019b). **2019년 국가 수준 학업성취도 평가 기술보고서: 출제 연구**. 한국교육과정평가원 연구보고 RRE 2019-8.
- 권태현, 김승현(2019). 수능 국어 영역 '화법과 작문' 평가 문항의 양호도 조사 연구. **청람어문교육**, 71, 161-194.
- 권태현, 이정찬, 김승현(2017). 수능 국어 영역 독서 평가 문항의 양호도 조사 연구. **독서연구**, 45, 131-159.
- 남민우, 민재원, 이인화(2016). 수능 문학 영역 평가 문항의 양호도 개선 방안 조사 연구. **문학교육학**, 51, 9-48.
- 동효관, 김경주, 강민경, 장의선, 성경희, 양성현, 김성경, 이재봉, 구자옥, 박상복, 김소연, 최원호, 김용진, 이기영(2018). **2015 개정 교육과정에 따른 국가수준 학업성취도 평가 출제 방안 연구**. 한국교육과정평가원 연구보고 RRE 2018-4.
- 박문환(2004). 대학수학능력시험 난이도 관련 변인 탐색. **수학교육학연구**, 14(1), 71-88.
- 박지현, 진경애, 김수진, 이상아(2018). **과정 중심 평가 내실화를 위한 교사의 평가 전문성 신장 방안 연구**. 한국교육과정평가원 연구보고 RRE 2018-5.
- 박지현, 송미영, 남민우, 최길찬(2020). **학생 평가의 타당도 제고를 위한 문항 특성 분석(Feature Analysis) 기법 활용 방안**. 한국교육과정평가원 연구보고 RRE 2020-1.
- 서울특별시교육청(2020). **2020 서울교육 주요업무**.
- 송현정(2004). 대학수학능력시험 언어 영역 난이도 관련 변인에 대한 연구. **국어교육**, 113, 233-262.
- 이종승, 김성훈, 김재철, 송현정, 박문환, 장경숙, 서재영(2003). **문항 난이도 추정 모형 개발 연구 - 대학수학능력시험의 언어, 수리, 외국어(영어) 영역을 중심으로**. 한국교육과정평가원 연구보고

RRE-2002-14.

조지민, 김명화, 김성경, 양성현, 정혜경(2018). **대학-직업 준비도 측면에서의 국가 수준 학업성취도 평가 분석 - 수학과를 중심으로**. 한국교육과정평가원 연구보고 RRE-2018-15.

- AERA, APA, & NCME (2014). *Standards for educational and psychological testing*. Washington, D.C.: American Educational Research Association.
- Baker, E. L., Cai, L., Choi, K., & Madni, A. (2015a). *Functional validity: Extending the utility of state assessments*. Presentation at the 2015 National Conference on Student Assessment (NCSA), San Diego, CA.
- Baker, E. L., Madni, A., Michiuye, J. K., Choi, K., & Cai, L. (2015b). *Mathematical reasoning project quantitative analyses results: Grades 4, 8, and 11*. LA: UCLA, CRESST.
- Baker, E. L., Choi, K., Madni, A., Kao, J. C., Rivera, N. M., & Cai, L. (2018). *Tasks 1 & 2: Feature analysis and Integrated psychometric modeling*. LA: UCLA, CRESST.
- Baker, E. L., & Choi, K. (2019). *Feature analysis approach: uses for national and international assessments*. Presented to the Chilean Delegates (2019.10.1.).
- Choi, K. (2019). *Quantitative Feature Analysis*. Presented at the workshop session at the annual meeting of the National Council of Measurement in Education. Canada: Toronto.
- Choi, K., Kao, J. C., Rivera, N. M., & Cai, L. (2018). *Career-readiness features in Korean assessment items(CRESST Report 859)*. LA: UCLA, CRESST.
- Chung, G., Redman, E., Michiuye, J. K., Feng, T., Eng, S., Madni, A., & Cai, L. (2018). *Innovative Item Pilot Study*. LA: UCLA, CRESST.
- Madni, A., Michiuye, J. K., Choi, K., Baker, E. L., & Cai, L. (2015). *Feature analysis qualitative and quantitative results for tests of mathematics and english language arts, Grades 4, 8, and 11*. LA: UCLA, CRESST.
- McMillan, J. H. (2018). *Classroom assessment-principles and practice that enhance student learning and motivation(7th ed.)*. NY: Pearson.

· 논문접수 : 2021.01.05. / 수정본접수 : 2021.01.29. / 게재승인 : 2021.02.16.

ABSTRACT

**Applying an integrated approach to analyse item features
for curriculum-based assessments**

Jihyun Park

Research Fellow, Korea Institute for Curriculum and Evaluation

Mi-young Song

Senior Research Fellow, Korea Institute for Curriculum and Evaluation

Minwoo Nam

Research Fellow, Korea Institute for Curriculum and Evaluation

Kilchan Choi

Associate Director, CRESST

This study explored how to apply the item feature analysis for ensuring the validity of student assessments. The item feature analysis provides evidences of ‘how and what to be assessed’. We considered item features for curriculum-based assessments, focusing on Korean and Math in middle schools. The results of implementing item feature analysis in Korean/Math tests from 2019 NAEA indicated that both Korean and Math tests consisted of valid items based on the national curriculum. Students’ strength and weakness were observed from analyzing the relationships between students’ achievement and features of Korean and Math test items. Item features based on the curriculum and test mode have been proved as critical applications for ensuring the validity of curriculum-based assessment as well as building a tool for student assessment. Follow-up study for more broadened use of the results of study was proposed to imply item feature analysis in schools and to widen the application across the feature categories for each subject. The results of this study would contribute to extend researches on the validity of student assessment and provide the theoretical basis for selecting item features for computer-based assessments.

Key Words: Item Feature Analysis, Validity of Student Assessment, Curriculum-based Assessment, National Assessment of Educational Achievement