

## 우리나라 학생평가 실태 연구: PISA 2015 참여국과의 국제비교를 중심으로<sup>1)</sup>

김 난 옥(경북대학교 박사과정)\*

손 원 숙(경북대학교 교수)\*\*

---

### <요 약>

평가는 학습의 질과 방향을 결정짓는 중요한 요소로서 미래 핵심역량을 키우기 위한 학교교육의 변화에서 매우 중요한 위치를 차지한다. 본 연구에서는 세계 각국의 학생평가 체제를 총체적 관점에서 경험적으로 비교하여 국제 수준에서 한국의 평가체제 특징을 확인하고 학생평가 개선을 위한 시사점을 도출하고자 하였다. 국가 간 학생평가 비교를 위해 평가 거버넌스, 교사의 평가역량, 평가절차, 평가결과의 활용의 네 가지 범주로 구성된 분석틀을 수립하였으며, PISA 2015의 학교, 학생, 교사 설문 자료를 활용하였다. 그 결과, PISA 2015 참여국 수준을 고려할 때 한국의 학생평가 특징은 다음과 같았다. 첫째 평가 거버넌스와 관련하여 한국에서는 표준화검사를 강조하고 총괄적 목적으로 평가를 사용하고자 하는 의도가 높게 나타난 반면, 학교의 평가 자율성은 참여국 평균보다 높았다. 둘째 교사의 평가역량 측면에서 우리나라의 현직교사 평가역량 개발을 위한 프로그램 제공과 과학교사의 평가효능감 수준은 전체 참여국의 평균보다 낮게 나타났다. 셋째, 평가절차에서 과학교사 간의 평가 관련 협력 수준은 높게 나타났다. 교실수준에서 교사가 개발한 평가나 관찰 및 숙제보다는 표준화검사나 구술 중심의 평가유형 사용 비중이 높았으며, 성적 산출 시 성장이나 능력 참조준거의 활용 정도가 상대적으로 낮았다. 넷째, 평가결과의 활용과 관련하여 우리나라 학교들의 평가결과 활용도는 다른 국가에 비해 낮은 경향을 보였다. 이러한 결과를 바탕으로 학생평가 체제 비교를 위한 분석틀의 각 측면에서 한국의 학생평가 체제에 대한 시사점을 논의하였다.

주제어 : 학생평가, 총체적 관점, PISA 2015, 평가 거버넌스, 교사의 평가역량, 평가절차, 평가결과의 활용

---

1) 본 논문은 2018 한국교육평가학회·한국교육과정평가원 공동 추계학술대회의 발표원고를 수정·보완함

\* 제1저자, ailyn@hanmail.net

\*\* 교신저자, wsohn@knu.ac.kr

## I. 서론

지식기반사회로 접어들면서 변동성과 불확실성을 수반하는 미래 사회에 행복한 삶을 영위할 수 있는 역량을 길러주기 위한 교육이 주된 목표가 되었다(교육부, 2016; 최상덕 외, 2013; Looney, 2009). 세계 각국에서는 이를 위해 교육과정, 교수·학습, 평가 등 교육 전반에 걸쳐 변화를 시도해왔다. 특히 평가는 학습의 질을 결정짓는 중요한 요소로서, 21세기 접어들면서 각국에서는 학생의 성장과 발달을 측정할 수 있는 평가 체제 마련에 집중하고 있다. 평가는 교사의 의사결정 및 수업의 모든 측면과 내적으로 연계되어 있는 활동(McMillan, 2014, p. 36)으로, 교수·학습의 핵심 요소이자 학교에서의 수행에 대한 질적 관리와 성장을 파악하는 근거가 된다. 평가방법에 따라 수업 및 학습방법이 달라질 수 있다는 평가의 교육조형기능(김성훈 외, 2018)이나 수업활동 계획에 앞서 평가를 위한 학습증거를 결정하는 백워드 교육과정 설계 모형(McTighe & Wiggins, 1999) 등은 교수·학습과정에서 평가의 역할을 강조하고 있다. 이에 미래 사회의 핵심역량을 키워주기 위한 학교교육의 변화에서 평가방안 개선은 매우 중요한 관심사라 할 수 있다.

수업과 평가를 연계하고 평가의 형성적 기능을 강조하는 ‘학습을 위한 평가(Assessment For Learning, AFL)’는 전 세계적으로 교실평가 변화의 핵심이라고 볼 수 있다. 특히 영국의 평가개혁단(the Assessment Reform Group)(ARG, 2002)은 AFL을 학교 현장에서 실천할 수 있도록 지침을 마련하여 제공하는 등 선구적인 역할을 하였다. 이후 이스라엘, 캐나다, 미국, 노르웨이, 아일랜드, 호주, 뉴질랜드 등에서도 AFL에 기반한 교실평가 개선 방안이 제안되었다(Birenbaum et al., 2015; OECD, 2005). 한편 우리나라에서도 성취기준에 맞춰 학생의 성취 수준을 평가하고 학습과정도 평가하도록 하는 등 교육과정-교수학습-평가의 연계성을 강화하고 있으며, 서술형 및 논술형 평가의 확대 시행 등(교육부, 2015) 국가수준에서 학생의 학습과 성장을 지원하는 과정중심평가가 추진되고 있다. 또한 최근의 과정중심평가에 대한 논의는 개별 교사가 가르친 학생들을 평가하는 교사별 평가로 확대되고 있다(임은영 외, 2017).

우리나라에서 추진되는 과정중심평가는 학습결과를 중심으로 서열화하고 경쟁을 유발하던 기존 평가의 문제점을 고려할 때 평가의 본래 기능을 회복하려는 노력으로 볼 수 있다. 하지만 과정중심평가의 의미나 운영방안의 모호성(박지현 외, 2018), 수업-평가 연계에 대한 학교의 비적극성(조수영, 2017), 평가의 공정성이나 타당성 확보(신혜진, 안소연, 김유원, 2017), 교사의 평가전문성(반재천 외, 2018) 등 과정중심평가 시행 과정 속에서 다양한 우려가 제기되고 있다. 하지만 학습을 위한 평가가 강조되고 있는 세계적 추세를 고려할 때 세계 각국의 학

생평가 동향을 확인하고 국제수준에서 한국의 평가 실제를 파악하는 것은 학생평가의 실질적 수행 수준 및 강·약점을 객관적으로 확인할 수 있고, 보다 나은 학생평가 체제 구축을 위한 유의미한 시사점을 가져다줄 것으로 판단된다.

학생평가 체제 비교에 대한 선행 연구를 살펴보면, 주로 일부 국가의 학생평가 정책이나 실제 등을 질적으로 검토하거나(김순남 외, 2013; 김천홍, 홍수진, 2017; 최상덕 외, 2013), 전반적 교육정책이나 교육과정 연구에 연계되어 학생평가 체제가 부분적으로 다뤄지거나(김석우 외, 2006; 정미경 외, 2013), 성취평가제(김유정, 홍훈기, 2019) 등 일부 평가요소에만 초점을 두었다는 제한점이 있었다. 한편 OECD(2013)에서는 25개국을 대상으로 학생평가, 교사평가, 학교 및 학교장 평가, 교육시스템 평가 등을 거시적 시각에서 비교하였으며, UNESCO 방콕 지부 아시아-태평양지역 교육사무국(Network on Education Quality Monitoring in the Asia Pacific, NEQMAP)에서는 아시아 8개국을 중심으로 학교 단위의 학생평가를 조사하였다. 이들 연구는 국가 간 학생평가 비교를 위한 개념적 틀(conceptual framework)을 제공하였다는 점에서 의의가 있지만, 연구대상 수가 비교적 제한적이며 경험적인 자료를 이용한 분석보다는 질적인 방식으로 국가 간 비교를 시도하였다는 한계점이 파악되었다. 따라서 보다 객관적인 자료와 다양한 국가를 포함하여 종합적인 관점에서 학생평가 체제를 분석할 필요성이 제안된다. 이에 본 연구에서는 양적 자료를 이용하여 국가별로 작동되고 있는 학생평가 체제를 하나의 종합적인 분석틀로 바라보는 총체적인 관점(holistic approach)을 활용하고자 한다. 이러한 관점은 평가의 구성요소들이 서로 간의 중복이나 목적의 불일치를 피하고 하나의 일관된 시스템으로 움직일 때 비로소 효과적으로 기능(OECD, 2013)한다는 점에서 유용하다.

이에 본 연구에서는 첫째, 국가 간 학생평가 비교를 위한 분석틀을 수립하였다. 분석틀은 OECD(2013)에서 제안한 평가체계 개념틀을 기반으로 하여 평가 거버넌스, 교사의 평가역량, 평가절차, 평가결과의 활용의 4가지 범주로 구성하였다. 둘째, 분석틀에 따라 PISA 2015 자료를 활용하여 참여국의 학생평가 체제 양상을 비교하였다. PISA 2015는 이전 주기와 달리 교사설문이 함께 실시되어 학생평가 현황을 학생 및 학교뿐 아니라 교사수준에서도 파악할 수 있으며, 세계 72개국이 일관된 절차에 따라 조사에 참여하였다는 점에서 다양한 국가의 자료를 포함한다는 장점이 있다. 본 연구에서는 세계 각국의 학생평가 체제를 경험적으로 비교하고 국제수준에서 한국의 평가체계 특징을 확인하여 보다 나은 평가 체제 구축을 위해 필요한 시사점을 도출하고자 한다. 또한 경험적 자료에 기반한 학생평가 체제 및 실제에 대한 국제비교 연구는 그동안 질적 연구 중심으로 이루어져 온 기존 선행연구를 보완할 뿐 아니라, 학생평가 체제 및 실제에 대한 국제적 흐름을 경험적으로 점검하고, 더불어 우리나라 학생평가 정책의 개선을 위한 시사점을 도출하는데 기여할 것으로 기대한다.

## II. 이론적 배경

### 1. 학생평가의 정의와 구성요소

본 연구에서는 다양한 평가 대상 중 학생을 대상으로 하는 학생평가에 초점을 두며, 학생평가는 ‘학생의 학습과정이나 결과를 개선시키기 위해 필요한 교사의 의사결정을 돕기 위해 필요한 정보를 수집·해석·활용하는 것과 관련된 모든 활동’(McMillan, 2014, p. 37)으로 정의하고자 한다. 교사의 의사결정에는 학생들의 강·약점이나 오개념 진단, 숙달을 위한 학생의 노력 및 진전과정 점검, 학습기록, 동기 및 핵심역량 향상, 성적 부여, 학부모에 대한 피드백 제공 등이 포함된다. 한편 학생평가의 구성요소는 그 목적과 특성에 따라서 다양하게 구분될 수 있는데, 예를 들어 PISA 2015(OECD, 2017)와 McMillan(2014)은 미시적 관점에서, UNESCO NEQMAP(정혜경 외, 2016; Tom, 2016)와 OECD(2013)에서는 보다 거시적 관점에서 학생평가의 구성요소를 제시하고 있다. 이들 구성요소에 대한 내용을 요약하면 <표 1>과 같다.

<표 1> 학생평가의 구성요소

구분	저자	구성요소
미시적 관점	PISA 2015 (OECD, 2017)	① 목적 및 기준(purpose and criteria): 어떤 목적으로 평가하는지, 정해진 목표를 공유하는지를 의미하며, 성적 산출을 위한 채점 기준도 포함함. ② 실행(practices): 교사개발평가나 표준화검사 등 평가에 사용하는 도구를 말함. ③ 활용 및 결과(use and consequences): 미리 정해진 목표에 따라 평가결과를 활용하는 것을 의미하며, 활용 방식은 형성적 방법과 종합적 방법으로 구분함.
	McMillan (2014)	① 평가목적(purpose): 어떤 평가를 왜 시행하는지, 평가를 통해 어떤 정보가 수집될 것인지, 수집된 정보가 교사의 의사결정에 어떠한 도움을 줄지 등 명확한 계획을 의미함. ② 측정(measurement): 어떠한 기법을 이용하여 학생 정보를 수집할 것인지를 의미하며, 공식적 비공식적 절차를 모두 포함하고 다양한 기법(검사, 평정, 관찰, 면담 등)이 사용됨. ③ 해석(interpretation): 측정결과에 가치를 부여하는 것으로, 어떤 기준을 사용하여 어떻게 해석할 것인지가 포함됨. ④ 결과활용(use): 진단이나 성적부여, 교수적 의사결정 등 어떤 용도로 평가결과를 활용하는지 의미함.
거시적 관점	UNESCO NEQMAP (정혜경 외, 2016; Tom, 2016)	① 평가전반(assessment): 평가목적 및 체계를 의미하며, 평가의 타당성과 공정성을 포함. ② 채점(marking): 학생 작업을 채점하고 증거를 기록하는 것으로 참된 평가 과제, 루브릭 이용, 기준대비 학생의 성취보고 등을 포함. ③ 교사 및 학교 간 조정(moderation among teachers and schools): 교사 판단의 일관성을 위해 협동하여 평가계획을 세우거나 연간 공통된 과제 작성, 채점을 위한 공통 기준, 기준 공유 등을 의미함. ④ 수유자에 대한 피드백(feedback to learners): 교사나 학생에 대한 시의적절하고 관련성 있으며 발전적인 피드백 제시를 의미하며, 구두 및 서면 피드백을 모두 포함.

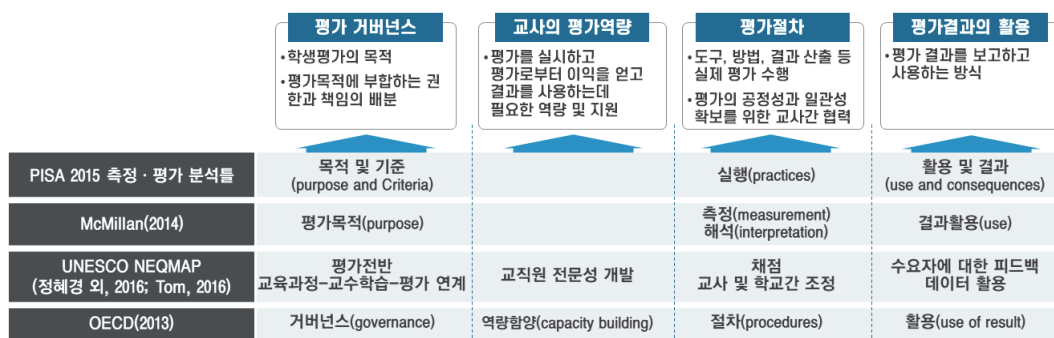
구분	저자	구성요소
		⑤ 데이터 활용(use of data): 교수 조정, 학습요구 충족, 목표 설정 등을 위한 결과 활용을 의미함. ⑥ 교육과정-교수학습-평가 연계(alignment): 교육을 위한 교육과정, 평가, 결과 제공의 연계성으로 각각에 대한 순환적 피드백 제공을 의미함. ⑦ 교직원 전문성 개발(professional development of staff): 평가를 실행하기 위한 전문성을 의미함.
거시적 관점	OECD (2013)	① 거버넌스(governance): 평가목적과 평가계획의 책임배분과 관련된 것으로, 평가목적은 형성적 방식과 총괄적 방식으로 구분함. ② 역량함양(capacity building): 평가에 참여하는 학생, 교사, 평가 전문기관 등이 가져야 할 평가전문가로서의 역량을 의미함 ③ 절차(procedures): 어떤 평가도구를 얼마나 사용하며 평가결과 산출에 어떤 기준을 활용하는지 말함. ④ 결과활용(use of results): 평가결과의 활용으로, 교육적 요구 식별이나 학생에 대한 피드백 제공 등 발달적 관점에서 활용하는 형성적 방식과 성적을 학생의 진급이나 반편성 등의 의사결정이나 학교의 학업 향상 정도, 교사의 효과성 등 책무성 관점에서 활용하는 총괄적 방식으로 구분함.

## 2. 학생평가 체제 분석을 위한 분석틀

각국의 학생평가 체제 분석을 위해 [그림 1]과 같이 분석틀을 수립하였다. 분석틀은 기본적으로 OECD(2013)에서 제안한 평가체제 개념틀에 초점을 두었다. 이는 OECD(2013)의 개념틀이 다른 선행연구에서 언급한 학생평가의 구성요소 및 원리를 대부분 포함하고 있으며, 본 연구가 총체적 관점에서 학생평가를 다루고자 할 때 고려한 다음 두 가지 조건을 만족하기 때문이다. 첫째는 학생평가가 무엇을 어떻게 평가할지 구체화되어 있고, 정확하게 목표에 부합되도록 정보가 수집되어 사용될 때 학습을 지원하는 강력한 도구가 될 수 있기에(Istance & Dumont, 2010) 목적, 절차, 활용의 평가 요소들을 총체적 시각에서 다룰 필요가 있다는 점이다. 둘째는 평가의 질을 높이기 위해서는 실용성과 효율성 높은 평가를 위한 교사 역량이 매우 중요하다(김성훈 외, 2018; McMillan, 2014)는 점이다.

분석틀은 크게 평가 거버넌스, 교사의 평가역량, 평가절차, 평가결과의 활용의 4가지 범주로 구분된다. 첫째, ‘평가 거버넌스’는 학생평가의 목적과 함께 평가목적에 부합하는 권한과 책임을 어떻게 배분할 것인가를 의미한다. 평가의 목적은 평가 시행의 근거로 구체적인 평가 설계 및 활용방안을 결정하는 핵심적 요소이다. 또한 형성평가의 특성상 교사가 교수·학습 진행 상황에 따라 평가의 내용, 방법, 형식 등 자유롭게 운영할 수 있는 평가의 자율성이 전제로 되어야하는 만큼(강태훈, 류성창, 2017), 최근 강조되고 있는 AFL에서 평가 권한에 대한 책임과 권한 배분은 중요한 요소라 할 수 있다. 이에 평가 거버넌스에서는 평가목적과 평가권한에 초점을 둔다. 평가목적은 평가유형에 따라 교수·학습 개선을 위한 형성적 의도와 성적부여나 효과성 평가 등의 총괄적 의도로 세분화하여 다루고, 평가의 권한과 책임은 학교의 평가

자율성으로서 평가계획 수립에 대한 권한이 학교·교사에게 있는지를 분석하였다. 둘째, ‘교사의 평가역량’은 평가를 실시하고 평가로부터 이익을 얻고 결과를 사용하는데 필요한 역량 및 지원을 의미한다. OECD(2013)에서는 평가에 참여하는 학생, 교사, 전문가 등의 평가역량을 모두 고려하고 있지만, 본 연구에서는 교사의 평가역량에 초점을 두었다. 이는 다양한 목적에 맞춰 평가방법을 선정하여 활용하고 타당한 의사결정을 내리기 위해서는 무엇보다 교사의 역량이 중요하기 때문이다(김성훈 외, 2018). 역량으로 교사의 평가효능감을 고려하였으며, 역량 함양을 위한 지원과 관련해서는 평가역량 개발 프로그램 제공여부와 전문성 개발 프로그램의 참여빈도를 분석하였다. 셋째, ‘평가절차’는 평가의 도구, 절차, 결과 산출 등 실제 평가 수행과 관련된 것을 말한다. 평가도구와 관련하여 성적 산출에 어떠한 참조준거를 활용하는지, 교실수준에서 사용하는 평가유형은 무엇인지 분석하였다. 한편 평가수행 과정에서 공정성 및 일관성 확보를 위한 노력이 고려될 필요가 있음을 고려하여, UNESCO NEQMAP에서 제안한 평가수행에서의 교사 간 협력을 포함하였다. 넷째, ‘평가결과의 활용’은 평가결과를 어떻게 보고하고 사용하는지와 관련된 것으로, 형성적 활용과 총괄적 활용으로 구분된다. 형성적 활용은 본질적으로 교육적 요구를 식별하고 학생에게 시의성 있는 피드백을 제공하며 교수전략을 조정하는 등 발달적 관점에서의 활용이며, 교사-학생 간 상호작용, 질문, 피드백과 같은 활동을 포함한다. 총괄적 활용은 산출된 성적을 학생의 진급이나 반 편성 등의 의사결정뿐 아니라 학교의 학업향상 정도, 교사의 효과성 등에 사용하는 것으로 책무성 관점에서의 활용이다.



[그림 1] 학생평가 체제 분석틀의 구성요소



### III. 연구방법

#### 1. 분석자료

본 연구에서는 국가별 학생평가 체제를 비교하기 위하여 PISA 2015의 학생, 학교장, 교사 설문 자료를 활용하였고, 이들 설문은 주로 PISA 2015의 주영역에 해당하는 과학교과 중심으로 구성되었다. PISA 2015에는 OECD 회원국 35개국, 비회원국 37개국으로 총 72개국이 참여하였으며 교사설문에는 총 19개 국가만 참여하였다. 교사설문의 경우 표본 학생이 소속된 학교에 재직 중인 교사 중 주영역인 과학 교사 10명과 그 외 일반교사 15명을 표집하여 설문을 실시하였다. 본 분석에는 학교장과 학생 자료 분석의 경우 PISA 2015에 참여한 72개국 중 키프로스(Cyprus)를 제외한 71개국을 대상으로 총 17,565개교, 총 508,676명의 학생 응답값이 포함되었으며, 교사 자료 분석의 경우 총 19개국의 약 113,751명의 교사 응답값이 포함되었다.

#### 2. 분석변수

국가별로 수행되고 있는 학생평가 실태를 경험적으로 파악하기 위해 앞서 제시한 분석틀을 사용하였다. PISA 2015의 조사틀(OECD, 2017)에 제시된 변인과 문항내용을 고려하여 PISA 2015의 학교장, 교사, 학생 설문에서 평가 거버넌스, 교사의 평가역량, 평가절차, 평가결과의 활용의 4가지 범주 및 각 하위요소에 해당하는 문항을 선별하여 선정하였다(<표 2> 참조).

<표 2> 본 연구의 분석틀 및 분석문항

범주	하위요소	변인 설명	설문
평가 거버넌스	형성적 의도	<ul style="list-style-type: none"> <li>표준화검사 또는 교사개발평가의 형성적 사용 의도 여부(3문항)</li> <li>-학생들의 학습안내</li> <li>-수업이나 교육과정의 개선점 확인</li> <li>-학생들의 요구에 따른 교수법 적용</li> </ul>	학교
	평가 목적	<ul style="list-style-type: none"> <li>표준화검사 또는 교사개발평가의 총괄적 사용 의도 여부(5문항)</li> <li>-학부모에게 학생의 학업향상정도 통보</li> <li>-학교성적을 지역 또는 전국성적과 비교</li> <li>-연도에 따른 학교 학업향상정도 점검</li> <li>-교사의 효과성에 대한 평가</li> <li>-다른 학교와의 비교</li> </ul>	
	총괄적 의도		
	학교의 평가 자율성	<ul style="list-style-type: none"> <li>학생평가 계획 수립에 대한 주 책임자가 학교인 경우</li> </ul>	1=교장·교사 0=교육청·교육부

범주	하위요소	변인 설명	설문
교사의 평가 역량	제공여부	· 교원양성과정(=예비교사 대상), 12개월 이내 전문성 계발활동(=현직교사 대상)에 학생평가 실습이나 평가결과 사용관련 주제 포함 여부	1=포함 0=비포함 교사 (일반)
	역량 개발 프로 그램 참여빈도	· 지난 12개월 동안 전문성 계발 활동(자격프로그램, 전문성계발 목적의 교사모임, 전문적으로 관심을 갖고 있는 주제에 대한 개인·공동연구, 공식적 학교 계획에 따른 멘토링·동료관찰, 전문서적 읽기, 동료 교사와의 교수향상을 위한 방법에 대한 비공식적 논의)에 참여한 횟수	교사 (공통)
	평가효능감	· 평가방법을 활용할 수 있는 능력(2문항 평균값) -성취수준이 가장 높은 학생에서부터 가장 낮은 학생에 이르기 까지 맞춤형 과제를 제작하고 부여하기 -다양한 평가 방법을 사용하기	1=전혀 못함 2=거의 못함 3=어느 정도 함 4=매우 잘 함 교사 (과학)
평가 절차	참조준거	· 학기말 성적 산출 시 기준 ① 기준참조(동일 수업 내 다른 학생의 성취수준) ② 기준참조(국가 또는 지역의 성취수준) ③ 능력참조(학생의 인지적 역량) ④ 성장참조(학생의 학업향상 정도나 노력, 수업참여 정도)	1=전혀 사용안함 2=거의 사용안함 3=약간 사용함 4=많이 사용함 교사 (일반)
	교사협력	· 시험출제 시의 성취기준, 과제, 채점기준, 다양한 난이도를 포함한 숙제 자료 공유에 대한 과학교사 간 협력 정도(4문항 평균값)	1=전혀 동의 안함 2=동의 안함 3=동의함 4=매우 동의함 교사 (과학)
	평가유형	· 평가유형별 사용 정도 ① 표준화검사, ② 구술, ③ 교사개발검사, ④ 관찰, ⑤ 숙제, ⑥ 학생의 자기평가	1=전혀(거의) 안함 2=가끔 3=자주 4=항상(거의) 항상 교사 (일반)
평가 결과의 활용	피드백	· 과학수업에서 교사의 피드백(5문항 평균값) -과학 과목에서 보이는 성취수준에 대해 말씀해주신다 -과학과목에서 나의 강점이 무엇인지에 대한 피드백을 주신다 -나에게 어떤 영역에서 성취도를 더 높일 수 있는지 말씀해주신다	1=전혀(거의) 없다 2=일부 수업에서 3=대부분 수업에서 4=모든(거의 모든) 수업에서 학생
	형성적 활용	· 나의 성취수준을 향상시킬 수 있는 방법을 말씀해주신다 · 나의 학습목표에 어떻게 도달할지에 대해 조언해주신다	
	수업 조정	· 과학수업에서 교사의 수업조정(3문항 평균값) -학생의 요구와 지식수준에 맞추어 수업을 조정하신다 -주제나 과제를 이해하지 못하는 학생들을 위해 개별적 도움을 제공하신다 -대부분의 학생이 이해하기 어려운 주제에 대한 수업을 할 때 수업 내용이나 방식을 바꾸신다	
	총괄적 활용	· 평가결과의 총괄적 활용 여부(4문항) -성취도 결과를 일반 언론에 공개 -행정당국에 의해 지속적으로 성취도 결과 추이 파악 -학부모에게 직접 성취도 결과가 제공 -학생평가를 교수활동 점검을 위해 사용	1=예 0=아니오 학교



### 3. 분석방법

본 연구의 분석을 위해 SPSS 18.0과 Mplus 8.1(Muthén & Muthén, 1998-2017)을 사용하였으며 구체적인 분석절차는 다음과 같다. 첫째, 본 연구의 분석단위는 학교수준으로, 모든 학생 및 교사 설문 응답은 학교수준으로 통합하였다. 둘째, 학교의 평가 자율성(평가 거버넌스), 역량 개발 프로그램 제공여부 및 참여빈도, 평가효능감(이상 교사의 평가역량), 평가수행에서의 교사협력(평가절차)의 국가별 수준을 확인하기 위해 기술통계분석을 실시하였다. 셋째, 평가목적(평가 거버넌스), 참조준거 및 평가유형(이상 평가절차), 평가결과의 활용에 따른 잠재집단 유형을 확인하기 위해 잠재프로파일분석(Latent Profile Analysis, LPA)을 실시하였다.

잠재프로파일분석은 투입된 연속적 관찰변인을 통해 특정 모집단에 존재하는 잠재적 이질성을 추정하는 방법으로 본 연구에서는 2~6개의 잠재프로파일 모형을 비교하였다. 모형 적합도 판정 기준으로 정보지수, 모형비교검증, 분류의 질, 집단 내 분류비율, 해석가능성을 고려하였다. 정보지수는 AIC(Akaike Information Criterion), BIC(Bayesian Information Criterion), 수정된 BIC(Sample-size Adjusted BIC)로, 각 지수는 값이 작을수록 좋은 모형임을 의미한다. 모형비교검증은 LMRT(Lo-Mendell-Rubin Adjusted Likelihood Ratio Test)을 사용하였다. 본 연구에서는 잠재집단의 수 결정을 위하여 관련 선행연구(Nylund, Asparouhov, & Muthén, 2007)에 기반하여 다양한 정보지수 중 BIC를 가장 우선적으로 고려하였고, BIC의 감소가 둔화되는 지점(Petrás & Masyn, 2010)에 근거하였다. 그리고 5% 미만의 소규모 집단의 해석 가능성 정도를 검토하였다. 한편 국가 간 비교의 용이성을 위하여 투입된 변인의 z점수 값을 활용하였다. 또한 PISA 2015에서 제공하는 모집단 가중치 중 Senate 가중치를 적용하였는데, 가중치를 적용함으로써 국가별 표본크기가 동등화되어 각 국가별 표본의 응답이 분석 결과에 동등한 효과를 갖도록 하였다(OECD, 2017).

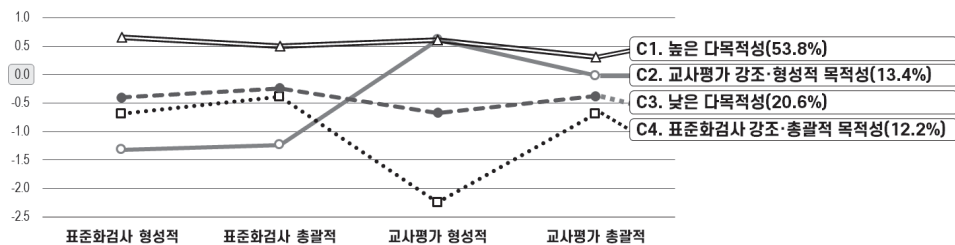
## IV. 연구결과

### 1. 평가 거버넌스

#### 가. 평가목적

평가목적은 크게 형성적 대 총괄적 기능으로 구분하여 표준화검사와 교사가 개발한 교실평가별 평가 목적을 측정하였다. 평가목적 양상은 잠재프로파일분석을 토대로 탐색하였다([부록 표 1] 참조). 앞서 제시한 통계적 기준과 이론적 기준을 모두 고려하여 잠재집단의 수가 4개

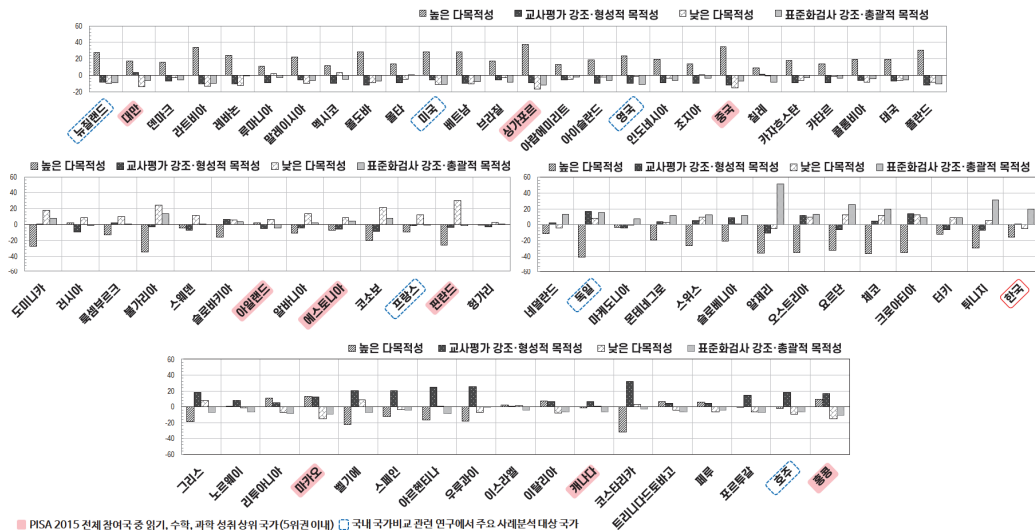
인 모형을 최종모형으로 선택하였으며, 이들 4개 집단의 프로파일 분포를 도식화하면 [그림 2]와 같다. 집단1(높은 다목적성, 53.8%)은 표준화검사나 교사평가의 형성적 총괄적 사용 의도가 모두 높았으며, 집단2(교사평가 강조·형성적 목적성, 13.4%)는 표준화검사의 형성적 또는 총괄적 사용 의도가 낮고 교사평가의 형성적 사용 의도가 높은 집단으로 관찰되었다. 집단3(낮은 다목적성, 20.6%)은 평가의 형성적 또는 총괄적 사용 의도 모두 평균보다 낮은 수준이었으며, 집단4(표준화검사 강조·총괄적 목적성, 12.2%)는 전반적으로 평가의 사용 의도가 낮았으나 특히 교사평가의 형성적 사용 의도가 매우 낮고 상대적으로 표준화검사의 총괄적 의도는 높게 관찰되었다.



[그림 2] 잠재프로파일 양상: 평가목적

학교의 평가목적 프로파일 분포를 국가별로 살펴보면, ‘높은 다목적성’은 13.06~91.45%, ‘교사평가 강조·형성적 목적성’은 1.15~45.34%, ‘낮은 다목적성’은 3.66~50.92%, ‘표준화검사 강조·총괄적 목적성’은 0.52~63.42%로 나타났다. 71개국의 프로파일은 그 양상에 따라 크게 4개의 유형으로 분류할 수 있다([그림 3] 참조). 첫 번째는 ‘높은 다목적성’ 프로파일의 비중이 평균보다 높고 다른 프로파일의 비중은 평균보다 낮은 국가로, 대만, 싱가포르, 중국, 뉴질랜드, 미국, 영국 등이 포함된다. 두 번째는 ‘낮은 다목적성’ 프로파일의 비중이 크고 ‘높은 다목적성’의 비중이 작은 국가로 에스토니아, 아일랜드, 핀란드, 프랑스 등이 포함된다. 이들 국가의 경우 ‘교사평가 강조·형성적 목적성’의 비중은 평균 수준이었다. 세 번째는 평균과 비교해 볼 때 ‘높은 다목적성’ 프로파일의 비중이 낮고 ‘표준화검사 강조·총괄적 목적성’의 비중이 평균보다 높은 국가로, 한국을 비롯하여 독일, 네덜란드, 스위스 등이 포함된다. 네 번째는 ‘교사평가 강조·형성적 목적성’의 비중이 높은 국가로 홍콩, 마카오, 캐나다, 코스타리카, 호주 등이 포함된다. 한편 한국의 경우 ‘높은 다목적성(38.00%)’ 프로파일의 비중은 평균보다 낮고, ‘표준화검사 강조·총괄적 목적성(31.91%)’은 평균보다 높게 나타났는데, 독일의 경우 한국과 마찬가지로 ‘표준화검사 강조·총괄적 목적성’ 프로파일의 비중이 높긴 했으나 ‘교사평가 강조·형성적 목적성’의 비중이 한국보다 높게 나타났다. 특히 싱가포르의 경우 ‘높은 다목적성(91.45%)’ 프로파일의 비중이 71개국 중 가장 높았다.

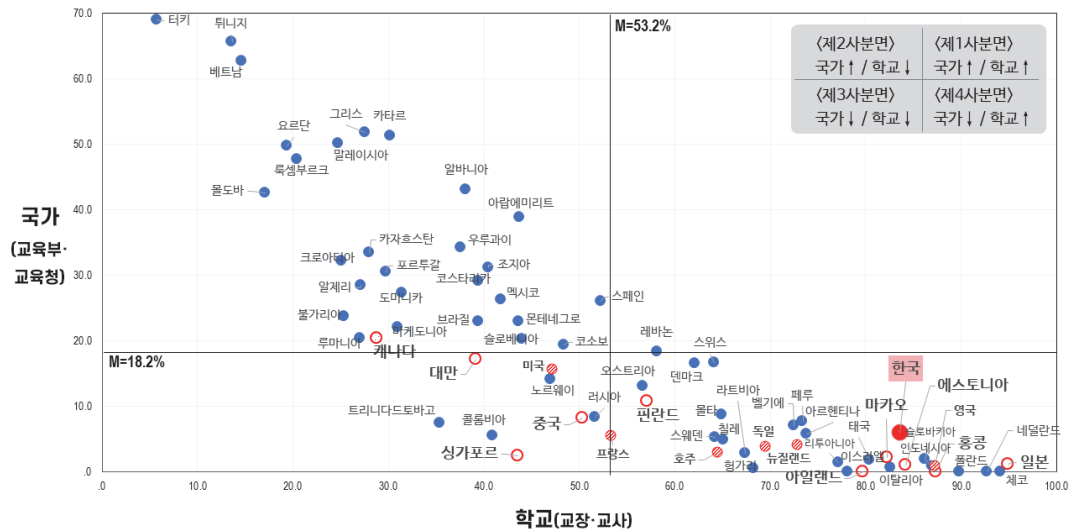
우리나라 학생평가 실태 연구: PISA 2015 참여국과의 국제비교를 중심으로



[그림 3] 국가별 평가목적 프로파일의 분포(국가별-전체평균과의 차이%p)

## 나. 학교의 평가 자율성

학교의 평가 자율성은 학생평가 계획 수립의 주책임자에 대한 학교장의 응답을 분석한 것으로, 국가별 비교 결과는 [그림 4]와 같다. 학생평가 계획 수립의 주책임이 어디에 있는지 교장, 교사, 운영위원회, 지역교육청, 교육부 중 중복으로 선택하도록 하였으며, 그 중 교장이나 교사만 선택한 경우 학생평가 계획 수립의 주책임이 학교(교장·교사)에 있다고 보고, 지역교육청이나 교육부만 선택한 경우 학생평가 계획에 대한 주책임이 국가(교육부·교육청)에 있다고 보았다. [그림 4]에서 가로축은 학교 책임, 세로축은 국가 책임으로, 특히 제2사분면은 학교가 주책임자라고 응답한 비율은 평균보다 낮고 국가가 주책임자라고 응답한 비율은 평균보다 많은 국가, 제4사분면은 학교가 주책임자라고 응답한 비율이 평균보다 많고 국가가 주책임자라고 응답한 비율이 평균보다 낮은 국가를 의미한다. 전체 분석대상 학교 중에서 학생평가 계획 수립의 주책임자가 학교라고 응답한 비율은 53.2%이며, 국가라고 응답한 비율은 18.2%였다. 한국의 경우 주책임이 학교에 있다고 응답한 학교는 83.5%로 높은 편이었다. 한국과 함께 일본(94.9%), 홍콩(87.3%), 영국(87.3%), 에스토니아(84.1%), 마카오(82.2%)에서 학교에 평가계획의 책임이 있다는 응답률이 80% 이상으로 높았으며, 핀란드, 호주, 독일, 뉴질랜드의 경우 평균보다 약간 높은 57.1~72.9%였다. 중국(50.3%), 미국(47.1%), 싱가포르(43.6%), 대만(39.2%), 캐나다(28.8%)에서는 학생평가 계획 수립의 주책임이 학교에 있다는 응답률이 평균보다 낮은 편이었다.



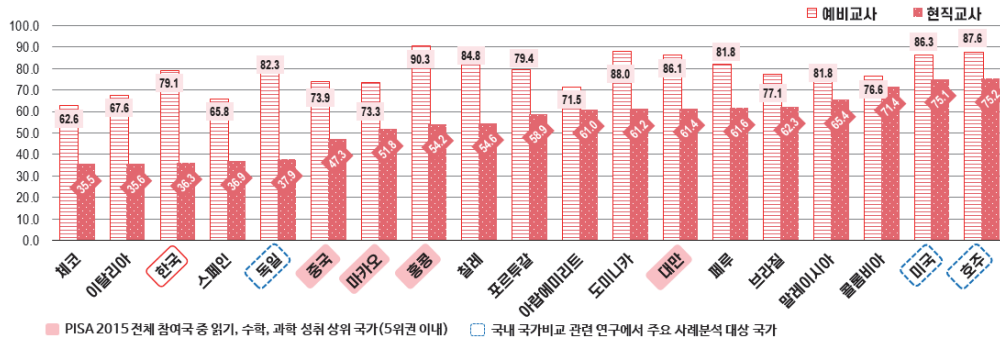
[그림 4] 학교의 평가 자율성: 학생평가 계획수립의 주책임자

## 2. 교사의 평가역량

### 가. 평가역량 개발 프로그램 제공여부

예비교사와 현직교사에게 학생평가 관련 역량을 개발하기 위한 프로그램이 제공되는지 일반 교사의 응답을 분석하였다([그림 5] 참조). 먼저 예비교사에게 학생평가 관련 역량 개발 프로그램이 제공된다고 응답한 비율은 전체 평균 78.4%이었다. 한국은 79.1%로 평균 수준이었으며, 홍콩(90.3%)에서 가장 높고 체코(62.6%)에서 가장 낮았다. 홍콩을 비롯하여 호주, 미국, 대만, 독일에서 80% 이상으로 평균보다 높은 응답률을 보였다. 다음으로 현직교사에게 제공된 역량 개발 프로그램 중 학생평가 주제가 포함되었다고 응답한 비율은 전체 평균 54.4%이었다. 한국(36.3%)을 포함한 이탈리아, 스페인, 독일에서 35.6%~37.9%로 평균보다 낮았으며, 호주(75.2%)에서 가장 높고 미국, 콜롬비아에서 70.0% 이상으로 평균보다 높은 응답률을 보였으며, 체코(35.5%)가 가장 낮게 나타났다.

우리나라 학생평가 실태 연구: PISA 2015 참여국과의 국제비교를 중심으로



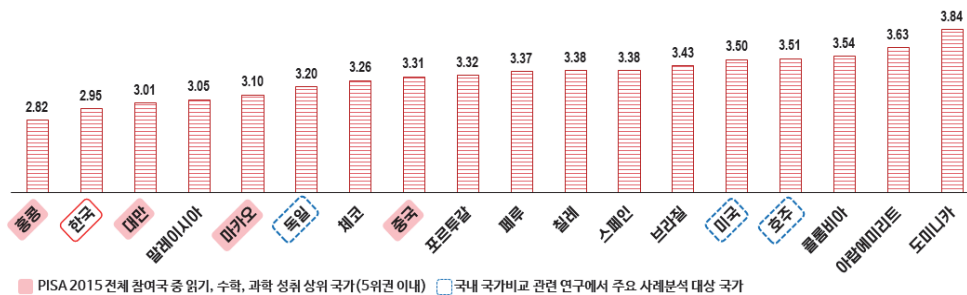
[그림 5] 평가역량 개발 프로그램 제공여부(%)

## 나. 역량 개발 프로그램 참여빈도

지난 12개월 동안 현직교사의 다양한 역량 개발 활동 참여빈도를 분석결과, 국제 수준의 평균은 연 3.34회였다. 한국의 경우 3.56회 참석하여 평균보다 높은 수준이었으며, 중국이 4.22회로 가장 높았고, 호주(3.81회)와 미국(3.59회)은 상대적으로 매우 높은 수준, 마카오(3.44회)와 대만(3.40회)은 평균을 약간 웃도는 정도, 홍콩(3.10회)과 독일(3.21회)은 평균보다 낮았다.

## 다. 교사의 평가효능감

평가효능감은 평가 전문성에 대한 자신감 및 효능감(이태경, 유신복, 손원숙, 2018)으로 정의되며, 과학교사의 응답을 학교수준에서 통합하여 국가별로 비교·분석하였다([그림 6] 참조). 교사의 평가효능감의 참여국 평균은 3.29(4점척도)로 다양한 평가방법을 ‘어느 정도 할 수 있는 수준’ 이상이였다. 한국의 경우 평균 2.95로, 가장 낮은 평가효능감을 보인 홍콩(2.82)에 이어 두 번째로 낮았다. 도미니카(3.84)에서 교사의 평가효능감이 가장 높았고, 미국(3.50), 호주(3.51)의 평가효능감은 평균보다 높았던 반면, 대만(3.01), 마카오(3.10), 독일(3.20)에서의 평가효능감 수준은 평균보다 낮은 것으로 나타났다.

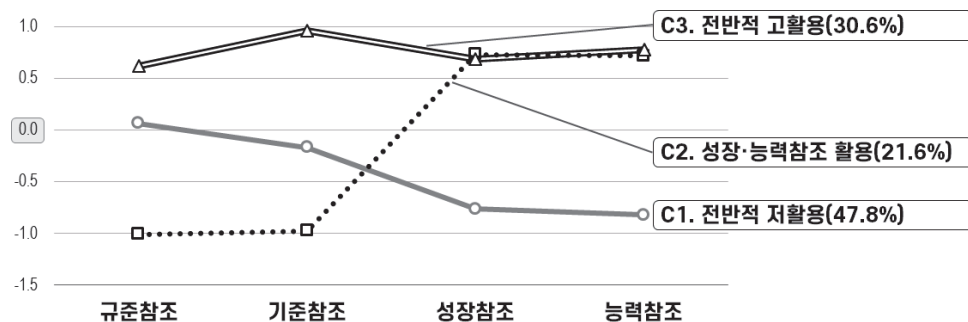


[그림 6] 과학교사의 평가효능감(4점척도)

### 3. 평가절차

#### 가. 참조준거

평가의 참조준거는 일반교사의 학기말 성적 산출 시 사용하는 평가기준에 대한 것으로, 잠재프로파일분석을 토대로 참조준거의 양상을 탐색하였다([부록표 1] 참조). 앞서 제시한 통계적 기준과 이론적 기준을 모두 고려하여 잠재집단의 수가 3개인 모형을 최종모형으로 선택하였다. 3개 집단의 프로파일 분포를 도식화하면 [그림 7]과 같다. 집단1(전반적 저활용, 47.8%)은 모든 산출 기준의 사용 정도가 ‘거의 사용하지 않는다’~‘약간 사용한다’ 사이의 비교적 낮은 경향을 보였으며, 집단2(성장·능력참조 활용, 21.6%)는 성적을 산출하기 위해 규준이나 기준의 참조준거를 활용하는 정도는 낮으나 성장이나 능력의 참조준거를 높게 활용하는 것으로 관찰되었다. 집단3(전반적 고활용, 30.6%)은 규준, 기준, 성장, 능력의 모든 참조준거를 ‘약간 사용한다’ 이상으로 비교적 많이 활용하는 것으로 나타났다.

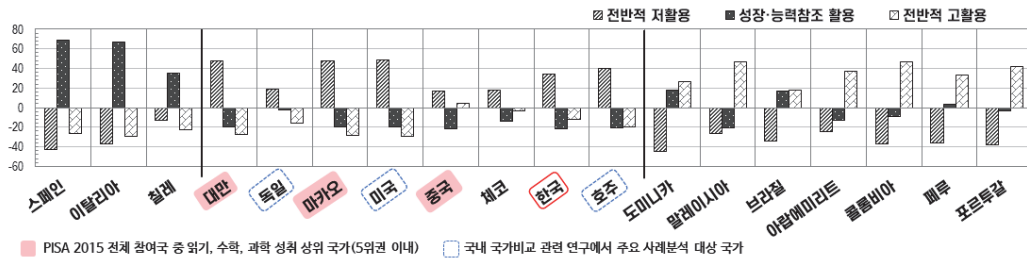


[그림 7] 잠재프로파일 양상: 참조준거

국가별로 참조준거의 프로파일 분포를 살펴보면, ‘전반적 저활용’은 2.82~96.73%, ‘성장·능력참조 활용’은 0.00~90.84%, ‘전반적 고활용’은 0.96~77.23%로 나타났다. 18개국의 프로파일은 그 양상에 따라 크게 3개의 유형으로 분류할 수 있다([그림 8] 참조). 첫 번째는 ‘성장·능력참조 활용’의 비중이 평균보다 높고 ‘전반적 저활용’ 및 ‘전반적 고활용’의 비중이 평균보다 낮은 국가로 스페인, 이탈리아 등이 포함된다. 두 번째는 ‘전반적 저활용’의 비중이 평균보다 높고 다른 프로파일의 비중이 평균보다 낮은 국가로, 한국을 비롯하여 대만, 마카오, 중국, 독일, 미국, 호주가 포함된다. 세 번째는 ‘전반적 고활용’의 비중이 평균보다 높은 국가로 콜롬비아, 말레이시아 등이 포함된다. 한편 한국의 경우 ‘전반적 저활용’의 비중(81.67%)이 평균보다 높고, ‘성장·능력참조 활용(0.00%)’ 및 ‘전반적 고활용(18.33%)’의 비중은 낮은 편이었으며, 이러한 경향은 대만, 마카오에서도 나타났다. 반면 스페인은 18개국 중 ‘성장·능력참조 활용(90.84%)’ 비중이 가장 높게 나타났으며, 콜롬비아에서 ‘전반적 고활용’ 비중이 가장 높았다.



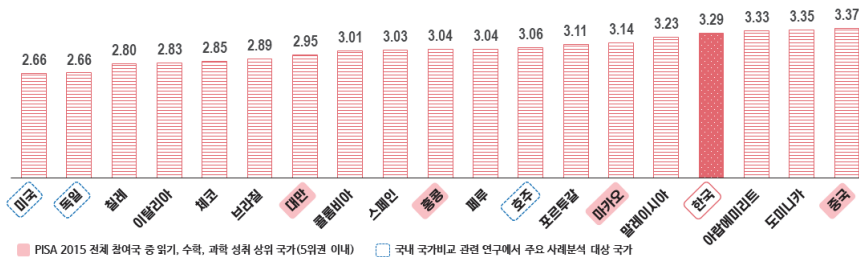
우리나라 학생평가 실태 연구: PISA 2015 참여국과의 국제비교를 중심으로



[그림 8] 국가별 참조준거 프로파일의 분포(국가별-전체평균과의 차이%p)

### 나. 평가수행을 위한 교사협력

동료 과학 교사들과 시험출제 시의 성취기준, 숙제, 채점기준, 다양한 난이도를 포함한 과제 자료 공유 등 학생평가와 관련하여 정기적으로 어느 정도 협력하는지 과학교사의 응답을 학교수준에서 통합하여 비교·분석한 결과는 [그림 9]와 같다. 참여국 전체의 과학교사 간 협력 평균은 3.03(4점척도)으로 ‘어느 정도 협력하는 수준’인 것으로 나타났다. 한국은 평균 3.29로 평균보다 높은 수준이었다. 중국이 평균 3.37로 가장 높았으며, 호주(3.06)와 홍콩(3.04)은 평균 수준, 독일(2.66)과 미국(2.66), 대만(2.95)은 평균보다 낮았다.

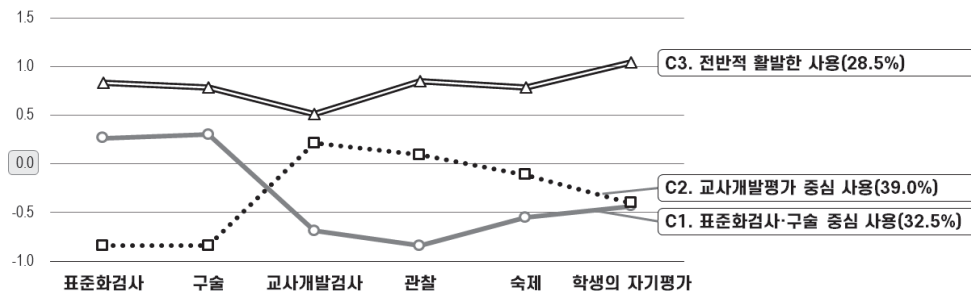


[그림 9] 평가수행을 위한 과학교사 간 협력 정도(4점척도)

### 다. 평가유형

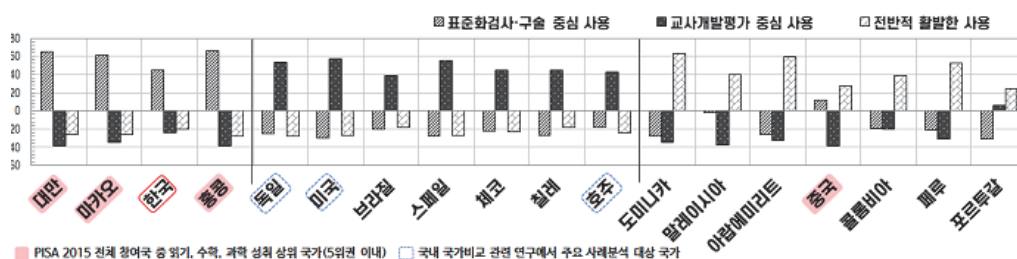
평가유형은 일반 교사(과학 교사를 제외한)가 사용하는 공식적 및 비공식적 학습평가의 사용 정도를 측정하였다. 평가유형의 양상은 잠재프로파일분석을 토대로 탐색하였다([부록표 1] 참조). 앞서 제시한 통계적 기준과 이론적 기준을 모두 고려하여 잠재집단의 수가 3개인 모형을 최종모형으로 선택하였다. 3개 집단의 프로파일 분포를 도식화하면 [그림 10]과 같다. 집단 1(표준화검사·구술 중심 사용, 32.5%)은 표준화검사와 구술의 사용 수준은 높으나 그 외 평가의 사용 수준은 낮고, 집단2(교사개발평가 중심 사용, 39.0%)는 교사개발평가 및 교사의 관찰, 숙제 등 교사개발평가를 중심으로 사용하는 것으로 관찰되었으며, 집단3(전반적 활발한 사용,

28.5%)은 검사, 구술, 관찰, 숙제, 학생의 자기평가의 모든 방법을 자주 사용하되 특히 학생의 자기평가 사용 수준이 높게 관찰되었다.



[그림 10] 잠재프로파일 양상: 평가유형

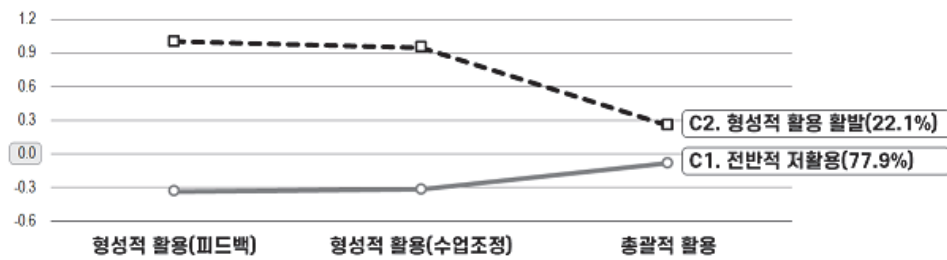
국가별로 교실수준 평가유형의 프로파일 분포를 살펴보면, ‘표준화검사·구술 중심 사용’ 프로파일은 1.57~98.24%, ‘교사개발평가 중심 사용’은 0.00~96.03%, ‘전반적 활발한 사용’은 0.46~91.86%로 나타났다. 18개국의 프로파일은 그 양상에 따라 크게 3개의 유형으로 분류할 수 있다([그림 11] 참조). 첫 번째는 ‘표준화검사·구술 중심 사용’ 프로파일의 비중이 평균보다 높고 ‘교사개발평가 중심 사용’과 ‘전반적 활발한 사용’의 비중은 평균보다 낮은 국가로, 한국을 비롯하여 대만, 홍콩, 마카오가 포함된다. 두 번째는 ‘교사개발평가 중심 사용’의 비중이 평균보다 높고 다른 유형의 비중은 평균보다 낮은 국가로 호주, 독일, 미국 등이 포함된다. 세 번째는 ‘전반적 활발한 사용’의 비중이 높은 국가로 콜롬비아, 중국 등이 포함된다. 한편 한국의 경우 ‘표준화검사·구술 중심 사용’의 비중(76.96%)이 높고, ‘교사개발평가 중심 사용’(15.15%) 및 ‘전반적 활발한 사용’(7.89%)의 비중은 많이 낮은 편이었다. 반면 ‘교사개발평가 중심 사용’의 비중은 18개국 중 미국에서 가장 높고 대만에서 가장 낮았으며, 독일의 경우 ‘전반적 활발한 사용’의 비중이 18개국 중 가장 낮은 특징을 보였다.



[그림 11] 국가별 평가유형 프로파일의 분포(국가별-전체평균과의 차이%p)

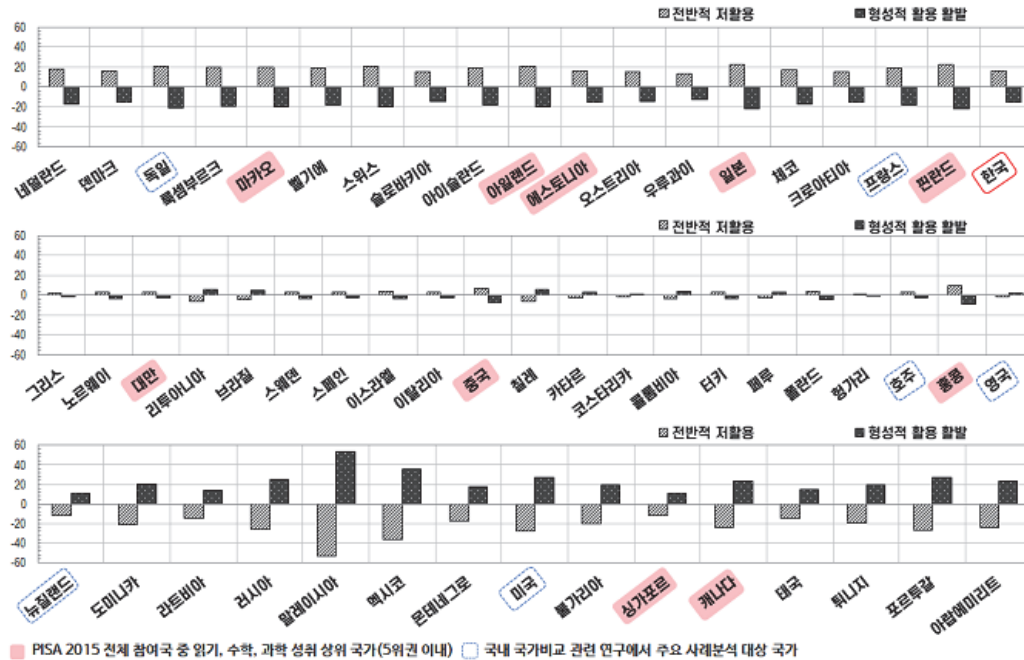
#### 4. 평가결과의 활용

평가결과의 활용은 크게 형성적 활용과 총괄적 활용으로 구분하였고, 형성적 활용은 학생이 지각한 평가실제와 교사가 보고한 피드백 실제 간에 일관성이 높다는 선행연구(김난옥 외, 2018)를 고려하여 학생이 지각한 피드백 제공과 수업개선을 위한 활용 정도를 사용하였으며, 총괄적 활용은 평가결과를 교사 및 학교의 질 점검과 학부모에 대한 정보 제공을 측정하였다. 평가결과 활용 양상은 잠재프로파일분석을 토대로 탐색하였다([부록표 1] 참조). 앞서 제시한 통계적 기준과 이론적 기준을 모두 고려하여 잠재집단의 수가 2개인 모형을 최종모형으로 선택하였다. 2개 집단의 프로파일 분포를 도식화하면 [그림 12]와 같다. 집단1(전반적 저활용, 77.9%)은 형성적 총괄적 활용 수준이 모두 낮은 편으로, 학생평가를 학생에게 피드백을 제공하거나 수업을 조정하는 형성적 활용 수준은 상대적으로 낮게 관찰되었으며 대부분의 학교가 여기에 포함되었다. 집단2(형성적 활용, 22.1%)는 형성적 총괄적 활용 수준이 모두 높은 편이었는데 특히 학생평가의 형성적 활용 수준이 높게 관찰되었다.



[그림 12] 잠재프로파일 양상: 평가결과의 활용

국가별로 평가결과 활용의 프로파일 분포를 살펴보면, ‘전반적 저활용’은 24.69~100.00%, ‘형성적 활용 활성화’는 0.00~75.31%로 나타났다. 55개국의 프로파일 양상에 따라 크게 3개의 유형으로 분류할 수 있다([그림 13] 참조). 첫 번째는 ‘전반적 저활용’의 비중이 평균보다 높은 국가로, 한국을 비롯하여 에스토니아, 핀란드, 프랑스, 독일, 아일랜드, 일본, 마카오 등이 포함된다. 두 번째는 ‘전반적 저활용’과 ‘형성적 활용 활성화’ 프로파일의 비중이 평균과 유사한 국가로, 호주, 대만, 홍콩, 영국, 중국 등이 포함된다. 세 번째는 ‘형성적 활용 활성화’의 비중이 평균보다 높은 국가로 캐나다, 뉴질랜드, 싱가포르, 미국 등이 포함된다. 한편 한국의 경우 ‘전반적 저활용’ 프로파일의 비중(93.54%)이 평균보다 높고, ‘형성적 활용 활성화’ 프로파일의 비중(6.46%)이 평균보다 낮다는 특징을 보였다. 반면 일본과 핀란드의 경우 ‘전반적 저활용’의 비중이 55개국 중 가장 높고 ‘형성적 활용 활성화’의 비중이 가장 낮다는 특징을 보였다.



[그림 13] 국가별 평가결과의 활용 프로파일의 분포(국가별-전체평균과의 차이%p)

## V. 요약 및 논의

본 연구에서는 PISA 2015 참여국의 학생평가 체제 양상을 경험적으로 비교하여 국제수준에서 한국의 평가체제의 특징을 확인하고 보다 나은 평가체제 구축을 위해 필요한 시사점을 도출하고자 하였다. 이를 위해 OECD(2013)의 평가체제 분석틀에 기반하여 평가 거버넌스, 교사의 평가역량, 평가절차, 평가결과의 활용의 4가지 범주로 구성된 분석틀을 제시하고 이 분석틀에 근거하여 PISA 2015 데이터를 분석하였다. 평가가 그 목적에서부터 활용까지 유기적으로 연결되어있음에도 불구하고 기존 선행연구에서는 일부의 평가요소에만 초점을 둔 반면, 본 연구는 평가의 구성요소나 운영절차를 함께 고려하여 총체적 관점에서 각국의 학생평가의 실질적 수행 수준을 살펴보았다는 점에서 의미가 있다. 우리나라의 학생평가 체제 특성을 본 연구 결과에 기반하여 요약하고 시사점을 논의하면 다음과 같다.

첫째, 평가 거버넌스 측면에서 분석 결과를 살펴보면, 대다수의 참여국에서는 책무성뿐만 아니라 교수·학습 개선을 위해 평가를 사용하고자 하는 의도가 두드러지고 있었다. 반면 우리나라 학교에서는 대체로 총괄적인 의도로 평가가 시행되고 있으며, 형성적인 의도는 제한적

임을 보여주었다. 최근 우리나라에서도 과정중심평가와 함께 수업과 평가의 일체화를 통한 평가의 형성적 기능을 강화하기 위한 다양한 노력이 경주되고 있다. 하지만 세계적인 수준과 비교해 보았을 때 여전히 우리나라 학교에서는 평가의 형성적 기능이 낮은 수준임을 확인하였다. 반면 우리나라 학교의 학생평가 계획 수립을 위한 학교의 평가 자율성은 참여국 평균보다 높은 수준으로 나타났다. 이는 학습 중심 교육과정을 강조한 7차 교육과정 이후 평가를 포함하여 수업의 자율성은 계속적으로 확대되는 경향을 보였으며, 이러한 노력으로 어느 정도 학교의 평가 자율성에 대한 인식이 높게 나타났다고 볼 수 있다. 그럼에도 대학입시를 앞두고 있는 중·고등학교에서는 평가의 공정성이나 신뢰성 확보 문제, 평가에 대한 불신이나 민원 증가와 같은 사회적 문제 등으로 인하여 학교수준에서 평가의 자율성을 발휘하기에는 제한적일 수 있다. 평가목적은 평가대상이나 상황에 따라 다양한데, 평가목적에 따라 평가의 성격, 평가의 시행 및 채점방법, 결과 활용 등이 결정되게 되며, 평가목적에 부합하도록 시행되고 활용될 때 학습에 도움이 될 수 있다(Istance & Dumont, 2010). 즉, 평가의 다양한 목적을 인식하고 각 목적에 부합하는 평가의 활용방안을 구안하는 것은 필요할 것이며 특히 교수·학습과정에서 교실 평가의 형성적 기능을 강화할 수 있는 다양한 시도가 요구될 것이다.

둘째, 교사의 평가역량과 관련하여 우리나라 과학교사들의 평가효능감은 다른 참여국과 비교해 보았을 때 매우 낮은 수준이었으며, 역량 개발을 위한 프로그램 제공 빈도 측면에서 예비교사의 경우에는 평균 수준, 그러나 현직교사의 경우에는 평균보다 낮게 제공되고 있는 것으로 파악되었다. 이는 예비교사와 현직교사의 평가전문성 개발을 위해 각국에서 높은 관심을 갖고 있으며, 전문성 개발을 위한 프로그램 제공 등 많은 노력을 기울이고 있다는 OECD(2013)의 보고와는 차이를 보였다. 다양한 평가목적에 맞춰 평가방법을 선정하여 실시하고 타당한 의사결정을 내리기 위해서는 무엇보다 교사의 평가역량이 중요하며, 특히 교수·학습 과정에서 학습을 돕는 평가로서 시행되는 과정중심평가에서 그 중요성은 더욱 강조된다. 우리나라의 모든 현직교사로 일반화하기에 제한적이기는 하나 과학교사의 낮은 평가효능감은 교사의 다양하고 활발한 평가활용을 제한할 수밖에 없다. 현직교사들은 평가 교육을 토대로 최신 평가 동향을 학습할 수 있으며, 본인에게 필요한 평가역량을 함양시킬 기회를 제공받을 것이다. 따라서 이러한 교육 기회의 제공은 교사의 평가효능감을 높이는데 기여할 것이란 점에서 현직교사를 위한 평가역량 개발 프로그램 및 참여기회를 확대할 필요성이 제안된다. 이를 위해 우리나라 교사들이 교실 평가를 시행하는데 있어서 필요한 역량이 무엇인지 탐색할 필요가 있으며, 이들의 평가효능감을 저해하는 요소들을 체계적으로 분석하고 이를 교육 내용에 반영할 수 있을 것이다.

셋째, 평가절차와 관련된 분석 결과를 살펴보면, 우리나라에서는 평가의 참조준거는 표준참조평가에 집중되어 있었고, 다양한 평가방법들 중 표준화검사의 사용의 비중이 높았다. 한편 평가 과정에서 과학교사 간 협력은 평균보다 높은 수준이었다. 이는 본 연구 대상의 대다수가

속한 고등학교 교실평가의 방향이 학습을 위한 평가보다는 학습결과에 대한 평가에 맞추어져 있다는 면에서 현행 우리나라 교실평가 정책이 반영된 결과라고 해석된다. 대다수의 동양권 국가(대만, 마카오 등)들은 우리나라와 비교적 유사한 양상의 평가절차 특성을 갖고 있는 것으로 파악되었다. 그러나 참여국의 전체 학교 중 약 52.2%는 다양한 평가참조를 활용하거나 성장 및 능력 참조의 활용 비중이 높은 것으로 나타났으며, 더불어 다양한 평가방법들을 균형 있게 활용하거나 표준화검사보다는 교사개발평가의 비중이 높다는 결과는 주목할 만하다. 이와 유사하게 25개국의 교육평가정책을 검토한 OECD(2013)에서도 여러 국가에서 교사들이 전통적인 평가방법을 넘어 보다 광범위한 평가를 사용하는 추세를 보고하고 있으며, 교사의 다양한 평가방법 사용을 위한 각국의 노력을 언급하고 있다. 이는 많은 국가에서 평가목적은 책무성보다는 학습향상에 무게를 두는 평가 방안에 주력하고 있고, 또한 수업과 평가의 연계성을 중시함으로써 수업과정에서 교사의 판단에 따라 자율적으로 평가를 시행할 수 있는 권한이 부여되고 있음을 시사한다. 평가는 평가대상이 갖고 있는 모든 자료를 종합적으로 수집하여 평가할 수 있도록 다양한 평가방법을 활용해야하며, 연속적인 평가를 통해 대상의 변화 및 발달을 점검할 수 있어야한다(성태제, 2016). 이에 우리나라의 학생평가 체제 개선 시 수업과정 속에서 지속적이고 종합적인 평가를 통해 학생의 수준을 파악할 수 있는 다양한 평가방법에 대한 활용을 장려하는 방안이 제안된다.

넷째, 우리나라의 경우 평가결과를 활용하는 정도가 다른 참여국에 비해 낮은 것으로 나타났다. 특히 최근 평가의 형성적 기능을 강화하는 교실평가 정책이 강조되고 있음에도 불구하고 평가의 형성적 활용이 아직은 활발하지 않은 것으로 나타났다. 반면 학생평가 체제 개선을 위한 노력을 많이 기울이는 국가로 알려진 호주, 뉴질랜드, 싱가포르, 미국 등에서는 평가결과를 형성적으로 활발하게 사용하는 집단의 비중이 높게 나타났다. 또한 이들 국가에서는 평가목적에서 형성적 목적성 및 다목적성도 높게 나타나 평가의 사용 의도가 실제 활용으로 이어짐을 보여준다고 볼 수 있다. 평가를 계획하는 목적에서부터 절차, 활용에 이르기까지 평가의 형성적 활용이 구체적으로 계획되어 있을 때 비로소 평가의 효과적 활용도 가능하게 되므로, 다양한 형태로 평가를 활용하기 위해서는 평가목적에서부터 다양한 평가 의도를 포함하도록 강조할 필요가 있을 것이다.

전반적으로 PISA 2015 참여국의 학생평가 체제는 학습 결과에 대한 평가보다는 학습을 위한 평가 방안에 초점을 두고 있는 것으로 파악되었다. 이러한 국제적인 수준과 비교해 볼 때, 우리나라의 학생평가 체제는 평가의 목적과 활용 측면에서 평가의 형성적인 기능이 강화될 필요성이 제안되었다. 이와 연동하여서 표준화검사나 수업 내의 질의응답과 같은 평가방법 뿐 아니라 학습의 과정과 향상에 초점을 두는 실제성이 높은 대안적인 평가방법들의 활용 방안을 모색할 필요성이 제안된다. 이를 위하여 현직교사 대상의 평가교육 프로그램에 대한 점검이 필요하며, 교육내용 개선을 토대로 교사들의 평가전문성 및 평가효능감을 향상시킬 수



있는 방안이 보완될 필요가 있을 것이다. 한편 본 연구를 통하여 파악된 세계 각국의 학생평가 동향과 국제수준에서 한국의 실질적 수행수준 정보가 추후 학생의 성장과 발달을 돕기 위한 학생평가 체제 구축을 위한 기초자료로 활용될 것을 기대한다.

마지막으로 본 연구에서는 총체적 관점에서 국가별 학생평가 체제를 경험적으로 살펴보았으나 다음과 같은 제한점이 있다. 첫째, 본 연구는 각국의 학생평가 체제를 비교하기 위하여 PISA 2015 참여국의 학교수준에서 잠재프로파일분석을 실시하였다. 통계적 기준과 이론적 기준을 고려하여 잠재집단의 수를 결정하였지만, 향후 잠재프로파일 결과에 대한 교차타당화의 필요성이 제안된다. 둘째, 학생평가에서 교사의 역량이 중요한 요소임을 고려하여 분석들에 교사의 평가역량을 포함하였다. 하지만 PISA 2015의 교사설문이 과학교사와 일반교사로 분리·시행되어 교사의 평가역량 하위요소별로 분석대상이 상이하다는 제한점과 사용할 수 있는 문항에 한계가 있었다. 셋째, 학생평가의 내부 구성요소에 주목하여 국가별 교육정책이나 기타 외부 요인까지는 고려하지 못하였다는 한계가 있다. 각 국가별 평가는 그것을 둘러싼 교육적, 사회적, 정책적 맥락과 연계되어 서로 영향을 주고받게 된다. 따라서 추후 외부 맥락적 요소를 포함하여 분석할 필요가 있다. 나아가 본 연구에서 파악한 학생평가 분석들의 구성요소별 특징에 추가하여 추후 연구에서는 이들 구성요소 간의 관련성에 대한 국제 비교분석을 추가할 필요성이 제안된다.

## 참 고 문 헌

- 강태훈, 류성창(2017). 2015 개정 교육과정에 따른 교실 수업에서의 교육평가 개선 방향 탐색. **교육연구**, 70, 255-277.
- 교육부(2015). 2015 개정 교육과정 총론. 교육부 고시 제2015-74호 【별책 1】.
- 교육부(교육부 보도자료, 2016. 12. 23.). 2030 인재강국 실현을 위한 대한민국 미래교육 청사진 -‘지능정보사회에 대응한 중장기 교육정책의 방향과 전략’시안 발표. 자료출처(검색일 2019. 4. 25.): <https://moe.go.kr/boardCnts/view.do?boardID=294&boardSeq=65229&lev=0&searchType=null&statusYN=W&page=1&s=moe&m=0501&opType=N>
- 김난옥, 박민애, 이빛나, 손원숙(2018). 교사의 특성과 피드백이 초등학교의 정의적, 인지적 성취에 미치는 영향. **교육과정평가연구**, 21(2), 129-151.
- 김석우, 김석호, 김운용, 정홍식(2006). 한·미 고등학교 간 수준별 교육과정 및 학생 평가에 관한 비교 연구. **한국교육학연구**, 12(1), 5-35.
- 김성훈, 김신영, 김재철, 반재천, 백순근, 서민원(2018). **예비교사를 위한 교육평가**. 서울: 학지사.
- 김순남, 임현정, 손희권, 권재기, 한상미, 우지현, 홍나래(2013). **학생평가 방법 개선 연구**. 서울: 한국교육개발원 현안보고 OR 2013-04.
- 김유정, 홍훈기(2019). 한국과 핀란드의 고등학교 내신평가체제 비교분석. **교육과정평가연구**, 22(1), 77-100.
- 김천홍, 홍수진(2017). 영국과 한국 교사의 학생평가 체제 비교 분석. **교육문화연구**, 23(2), 211-237.
- 박지현, 진경애, 김수진, 이상아(2018). **과정 중심 평가 내실화를 위한 교사의 평가 전문성 신장 방안 연구**. 세종: 한국교육과정평가원 연구보고 RRE 2018-5.
- 반재천, 김선, 박정, 김희경(2018). 교사별 과정 중심 평가에 대한 교사의 인식. **교육과정평가연구**, 21(3), 105-130.
- 성태제(2016). **교육평가의 기초**. 서울: 학지사.
- 신혜진, 안소연, 김유원(2017). 과정 중심 평가 활용의 정책적 분석. **교육과정평가연구**, 20(2), 135-162.
- 이태경, 유신복, 손원숙(2018). 공업계열 고등학교 교사의 학습을 위한 평가실제와 학교의 학업낙관성과의 관계: 교사의 평가자율성, 평가효능감 및 평가태도의 매개효과. **중등교육연구**, 66(2), 525-550.
- 임은영, 김유향, 이인화, 심현표, 김성식(2017). **교사별 학생평가 실태 조사 및 현안 진단**. 서울: 한국교육과정평가원 연구자료 ORM 2017-66-24.
- 정미경, 양승실, 김경애, 김정민, 류성창(2013). **초·중등교육체제 개선 방안 연구-교육내용, 방법**

- 및 평가 체계를 중심으로-. 서울: 한국교육개발원 연구보고 RR 2013-32.
- 정혜경, 조지민, 홍미영, 박지현(2016). UNESCO NEQMAP의 교수·학습 개선을 위한 학생평가 운영 한국 사례 분석. 서울: 한국교육과정평가원 연구보고 RRE 2016-6.
- 조수영(2017). 고등학교 현장에서의 역량기반 교육과정-수업-평가 연계의 현실화 방안. *중등교육 연구*, 65(1), 255-281.
- 최상덕, 서영인, 황은희, 최영섭, 장상현, 김영철(2013). 미래 인재 양성을 위한 핵심역량 교육 및 혁신적 학습생태계 구축(Ⅰ). 서울: 한국교육개발원 연구보고 RR 2013-20.
- ARG (2002). Assessment for learning: 10 principles. <https://www.aaia.org.uk/news/article/assessment-for-learning-10-principles>(검색일: 2019. 7. 1.)
- Birenbaum, M., Deluca, C., Earl, L., Heritage, M., Klenowski, V., Looney, A., Smith, K., Timperley, H., Volante, L., & Wyatt-Smith, C. (2015). International trends in the implementation of assessment for learning: Implications of policy and practice. *Policy Futures in Education*, 13(1), 117-140.
- Istance, D., & Dumont, H. (2010). Future directions for learning environments in the 21st century, 317-388. In Dumont, H., Istance, D., & Benavides, F.(eds.), *The Nature of Learning: Using Research to Inspire Practice*, OECD Publishing, Paris, <https://doi.org/10.1787/9789264086487-15-en>.
- Looney, J. W. (2009). *Assessment and Innovation in Education*. OECD Education Working Papers, No. 24, OECD Publishing. <http://dx.doi.org/10.1787/222814543073>
- McMillan, J. H. (2014). *Classroom assessment: pearson new international edition: principles and practice for effective standards-based instruction*. Pearson Higher Ed. [손원숙, 박정, 강성우, 박찬호, 김경희 역 (2015). 교실평가의 원리와 실제: 기준참조수업과의 연계. 서울: 교육과학사].
- McTighe, J., & Wiggins, G. (1999). *Understanding by design: Handbook*. Alexandria. VA: Association for Supervision & Curriculum Development.
- Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user's guide*. Eighth edition. Los Angeles, CA: Muthén & Muthén.
- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A monte carlo simulation study. *Structural Equation Modeling*, 14(4), 535-569.
- OECD. (2005). *Formative Assessment: Improving Learning in Secondary Classrooms*. OECD Publishing, Paris.

- OECD. (2013). *Synergies for better learning: An international perspective on evaluation and assessment*. Paris: OECD Publishing.
- OECD. (2017). *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving, revised edition*. PISA, OECD Publishing, Paris. <http://dx.doi.org/10.1787/9789264281820-en>
- Petras, H., & Masyn, K. (2010). General Growth Mixture Analysis with Antecedents and Consequences of Change, 69-100. In Piquero, A. R., & Weisburd, D. (eds.). *Handbook of Quantitative Criminology*. New York, NY: Springer.
- Tom, C. (2016). School-Based Assessment Framework. <https://bangkok.unesco.org/content/orientation-meeting-regional-study-school-based-assessment-policies-and-practices-asia>(검색일: 2019. 4. 22.)

· 논문접수 : 2019.07.04. / 수정본접수 : 2019.08.02. / 게재승인 : 2019.08.14.

## ABSTRACT

### The Investigation of Student Assessment System of Korea: International Comparative Analysis of PISA 2015 Data

Nan-ok Kim

Graduate Student, Kyungpook National University

Won-Sook Sohn

Professor, Kyungpook National University

Education system around the world have been required changes to develop the students' competencies. Specially student assessment is important in those changes as a factor that determines the quality and direction of learning. The purpose of this study is to empirically compare student assessment systems of the countries participating in PISA 2015 and to draw lessons from the results to improve student assessment system in Korea. For analyses, we used the analytical framework consisting of four categories: assessment governance, teachers' assessment competency, assessment procedure and use of assessment results. As a result, the characteristics of the student assessment system in Korea were as follows. First, in terms of assessment governance, Korean schools would tend to put more stress on the use of standardized tests while the degree of school assessment autonomy was the higher than average. Second, teachers' competency development program was offered lower in Korea compared to other countries. And science teachers' assessment efficacy was significantly lower than average. Third, assessment cooperation among science teachers was higher than average. Also the proportion of latent profile using standardized test or oral test in the class was high in Korea. Fourth assessment results in Korea, the degree of use of assessment results for either formative or summative was the lower than in other countries. Finally, implications for to build better student assessment system in Korea are discussed.

*Key Words: student assessment, holistic approach, PISA 2015, assessment governance, teachers' assessment competency, assessment procedure, use of assessment results*

**[부록표 1] 잠재프로파일(집단) 수에 따른 모형 적합도**

구분	모형	(집단수)	Log Likelihood	AIC	BIC	SSA-BIC	Entropy	pLMP	집단크기 5%미만
평가 거버넌스									
평가목적	M1	2	-69545.144	139116.288	139214.036	139172.723	.953	.000	0
	M2	3	-66907.928	133851.856	133987.199	133929.997	.962	.000	0
	M3	4	-60239.712	120525.423	120698.362	120625.270	.968	.043	0
	M4	5	-59784.208	119624.416	119834.950	119745.969	.855	.418	0
	M5	6	-62882.827	125831.653	126079.783	125974.912	.907	.867	0
평가절차									
참조준거	M1	2	-27301.119	54628.238	54713.180	54671.871	.752	.000	0
	M2	3	-26425.757	52887.514	53005.127	52947.929	.782	.001	0
	M3	4	-25876.751	51799.502	51949.785	51876.699	.832	.146	1
	M4	5	-25620.247	51296.494	51479.447	51390.473	.766	.304	1
	M5	6	-25442.012	50950.024	51165.647	51060.785	.761	.334	1
평가유형	M1	2	-38473.228	76984.457	77107.468	77047.093	.809	.000	0
	M2	3	-37469.786	74991.571	75159.903	75077.284	.789	.133	0
	M3	4	-36949.856	73965.713	74179.364	74074.502	.779	.532	0
	M4	5	-36503.864	73087.729	73346.700	73219.595	.798	.541	1
	M5	6	-36145.863	72385.727	72690.018	72540.669	.824	.214	2
평가결과의 활용									
평가결과 의 활용	M1	2	-52494.050	105008.099	105082.549	105050.770	.606	.214	0
	M2	3	-51479.498	102986.995	103091.225	103046.734	.683	.007	1
	M3	4	-51171.838	102379.675	102513.684	102456.482	.733	.732	2
	M4	5	-50910.068	101864.135	102027.924	101958.010	.666	.320	2
	M5	6	-50675.049	101402.098	101595.666	101513.041	.704	.430	3