

청소년의 읽기, 수학 및 과학 성취에 영향을 미치는 주요 예측변수 탐색¹⁾

정혜원 (충남대학교 교육학과 교수)*
박소영 (충남대학교 교육학과 박사수료)**

요약

본 연구는 청소년의 인지적 역량 중 읽기, 수학 및 과학 성취 관련 주요 예측변수를 탐색하고, 주요 예측변수가 세 영역에 미치는 영향력을 검증하고자 하였다. 이를 위해 PISA 2018 자료에 다층자료 구조를 반영할 수 있는 혼합효과 랜덤포레스트(mixed-effect random forest: MERF)와 다층모형을 적용하였다. 주요 연구결과는 다음과 같다. 첫째, 예측성도가 좋은 혼합효과 랜덤포레스트를 적용한 결과, 학생과정 요인(읽기 전략 방법, 읽기에 대한 능력 지각 및 즐거움, PISA 검사와 관련된 변수) 및 ICT 요인(ICT 사용에 대한 지각된 자율성)이 읽기, 수학 및 과학 성취에 공통적인 주요 예측변수로 도출되었다. 영역에 따라 상이한 결과도 도출되었는데, 특히 수학 성취에서만 연간 가계 소득, 정규 수업 외 추가 교육, 방과 후 공부 시간 등이 주요 예측변수로 탐색되었다. 둘째, 다층모형 분석 결과, 혼합효과 랜덤포레스트에서 도출된 주요 학생 및 부모관련 요인은 청소년의 읽기, 수학 및 과학 성취에 유의한 영향을 미치는 것으로 나타났다. 이러한 연구결과를 토대로 청소년의 인지적 성취 수준 제고를 위한 다각적인 시사점을 논의하였다.

주제어 : 읽기 성취, 수학 성취, 과학 성취, PISA 2018, 혼합효과 랜덤포레스트, 다층모형

1) 이 논문은 2020년 대한민국 교육부와 한국연구재단의 인문사회분야 중견연구지원사업의 지원을 받아 수행된 연구임 (NRF-2020S1A5A2A01041976)

* 제1저자, chw7@cnu.ac.kr

** 교신저자, psy7522@hanmail.net

I. 서 론

한국은 최근 실시된 PISA(Programme for International Student Assessment) 2018에서 조사에 참여한 전체 79개국 중 읽기 영역 6~11위, 수학 영역 5~9위, 과학 영역 6~10위를 기록하며, 상위권의 인지적 성취 수준을 보였다(조성민, 구남옥, 김현정, 이소연, 이인화, 2019). 한국은 2000년부터 만 15세 청소년을 대상으로 3년 주기로 실시되는 PISA에 매년 참가하고 있으며, PISA 2000에서 PISA 2018까지 7번 시행되는 동안 순위 변화는 있었으나 지속적으로 상위의 인지적 성취 수준을 유지하고 있다. 그러나 PISA 2009 결과에서 한국의 인지적 성취의 비약적인 성장에 대한 OECD 보도 자료가 배포될 만큼 최상위 수준이었던데 비해(이은하, 최은정, 2015), 이후 점차 하락하고 있는 추세이다.

그간 국내에서는 PISA 자료를 활용한 수많은 연구가 수행되어 왔으며, 다양한 학생관련 변수와 학교관련 변수가 읽기, 수학 및 과학 성취에 영향을 미치는 것으로 밝혀졌다. 구체적으로 학생관련 변수인 학생배경 요인, 학생과정 요인, 진로관련 요인, ICT 요인(구자옥, 한정아, 김성숙, 2015; 김수혜, 2018; 김혜숙, 2012; 윤정일, 이범수, 2006; 이인화, 구남옥, 2019; 이정민, 정혜원, 2019; 정혜원, 박소영, 김정인, 김아름, 2021), 사회·환경적 요인과 학교관련 변수(구자옥 외, 2015; 김수혜, 2018; 김혜숙, 2012; 윤정일, 이범수, 2006; 정혜원 외, 2021)가 청소년의 읽기 성취, 수학 성취 또는 과학 성취에 영향을 미치는 것으로 나타났다. 그러나 PISA 자료를 활용한 대부분의 선행연구에서는 회귀분석, 구조방정식, 다층모형 등의 전통적인 통계기법을 적용해왔으며, 선행연구나 연구자가 설정한 모형에 따라 예측변수를 선택적으로 분석에 포함하였다.

이와 달리 최근에는 PISA 자료에서 포함하고 있는 많은 변수를 동시에 활용할 수 있는 머신러닝 기법을 적용해 청소년의 읽기 성취를 예측하는 주요 변수를 탐색한 연구가 시도되었다(손윤희, 박현정, 박민호, 2020; 정혜원 외, 2021). 이를 통해 청소년의 읽기 성취를 예측하는 데 주요한 변수를 확인하였으며, 선행연구에서는 상대적으로 주목받지 못한 새로운 주요 변수를 확인하였다는 점에서 머신러닝 기법의 활용 가능성이 대두되었다(손윤희 외, 2020; 정혜원 외, 2021). 그러나 이러한 선행연구들은 인지적 성취의 세 영역 중 PISA 2018의 주 영역²⁾인 읽기 성취만을 분석하였으며(손윤희 외, 2020; 정혜원 외, 2021), PISA 자료가 학생이 학교에 내재된 다층자료 구조이지만 다층자료의 구조를 분석시 반영하지 못하는 전통적인 랜덤포레스트(random forest: RF) 기법을 적용하였다.

이에 본 연구는 다층자료 구조를 반영할 수 있는 혼합효과 랜덤포레스트 기법(mixed-effect random forest: MERF)(Hajjem, Bellavance & Larocque, 2014)을 소개하고, 전통적인 랜덤포레스트와 예측성과를 비교하여 PISA 2018 자료에 더 적합한 기법을 확인하고자 한다. 이후 예측성과가 더 우수한 기법을 적용하여 청소년의 읽기, 수학 및 과학 성취에 영향을 미치는 주요 예측변수를 탐색하고, 세 영역에서 공통적 또는 차별적으로 도출되는 예측변수를 확인하여 영역별 균형 있는 학습 발

2) PISA에서는 매년 시행되는 평가에서 읽기, 수학 및 과학 영역 중 주 영역을 순환적으로 설정하고 있으며, PISA 2018에서는 읽기 영역이 주 영역이었음.

달을 위해 범교과적 또는 영역별 방안을 모색하고자 한다. 또한 머신러닝 기법을 적용해 도출된 주요 예측변수가 청소년의 읽기, 수학 및 과학 성취에 어떠한 영향을 미치는 지 검증하기 위해 다층모형 분석을 수행하고자 한다. 본 연구에서 설정한 연구문제는 다음과 같다.

연구문제 1. 혼합효과 랜덤포레스트와 일반적인 랜덤포레스트 간 청소년의 읽기, 수학 및 과학 성취 예측성과는 차이는 어떠한가?

연구문제 2. 청소년의 읽기, 수학 및 과학 성취에 영향을 미치는 주요 예측변수 중 영역별 공통적 또는 차별적 예측변수는 무엇인가?

연구문제 3. 주요 예측변수는 청소년의 읽기, 수학 및 과학 성취에 어떠한 영향을 미치는가?

II. 선행연구 고찰

1. 읽기 성취 영향변수

PISA 2018에서 읽기 소양의 개념은 “자신의 목적을 성취하고 지식과 잠재적 능력을 계발하며, 사회에 참여하기 위해 텍스트를 이해·활용·평가·성찰하고, 다양한 텍스트 읽기 활동에 관여하는 능력”으로 정의되었다(조성민 외, 2019, p.39). 읽기 성취와 관련된 선행연구를 살펴보면, 청소년의 읽기 성취는 학생 또는 학교와 관련된 다양한 변수에 의해 영향을 받는 것으로 나타났다.

PISA 2018 자료를 통해 읽기 성취에 영향을 미치는 변수를 분석한 정혜원 외(2021)의 연구에서는 학생배경 요인인 가정의 ESCS, 가정에서 모국어 사용 여부가 청소년의 읽기 성취에 긍정적인 영향을 미치는 것으로 나타났다. 또한 학생과정 요인인 읽기 전략(요약, 이해 및 기억), 디지털 읽기 전략, 읽기에 대한 능력 지각, 읽기에 대한 즐거움, PISA 검사 또는 응답에 대한 노력, 다양한 읽기 활동(픽션)은 읽기 성취에 정적인 영향을 미쳤다. 반면, 학생과정 요인 중 읽기에 대한 어려움 지각, PISA 검사 어려움에 대한 인식, 총 학습시간은 부적인 영향을 미쳤다. 진로관련 요인도 청소년의 읽기 성취에 영향을 미치는 것으로 나타났는데, 학생이 기대하는 직업 지위가 높을수록, 고등학교나 대학교 또는 대학원 이상의 학력에 대한 기대는 읽기 성취를 높이는 것으로 나타났다. ICT 요인으로는 ICT에 대한 흥미는 읽기 성취에 긍정적인 영향을 미친 반면, 가정에서 사용 가능한 ICT와 학교에서 일상적인 ICT 사용은 읽기 성취에 부정적인 영향을 미치는 것으로 나타났다. 또한 학교관련 변수인 학교 분위기에 영향을 미치는 학생 행동은 읽기 성취에 부정적인 영향을 미치는 것으로 나타났다.

PISA 2015 자료를 활용한 이인화와 구남욱(2019)은 성별에 따라 읽기 성취에 차이가 있는 것으로 나타났는데, 남학생에 비해 여학생의 읽기 성취가 상대적으로 더 높은 것으로 나타났다. 또한 학생배경 요인인 가정의 ESCS도 읽기 성취에 정적인 영향을 미쳤다. 학생과정 요인인 학업적 동기와 학습 시간(국어)은 읽기 성취에 정적인 영향을 미친 반면, 사회·환경적 요인인 학교 지각은 읽기 성취에 부정

인 영향을 미쳤다. 학교관련 변수 중에서는 학교 분위기에 영향을 미치는 학생 행동이 읽기 성취에 부적인 영향을 미치는 것으로 나타났다.

PISA 2009 자료를 분석한 김혜숙(2012)의 연구에서는 학생배경 요인인 성별과 가정의 ESCS에 따라 청소년의 읽기 성취에 차이가 나타났으며, 남학생에 비해 여학생이, 가정의 ESCS가 높을수록 읽기 성취가 높은 것으로 나타났다. 학생과정 요인인 읽기에 대한 즐거움과 온라인 읽기 활동은 읽기 성취에 정적인 영향을 미쳤다. 또한 ICT 요인이 읽기 성취에 영향을 미쳤는데, 학교에서 ICT 사용과 레저 목적(오락, 인터넷 등)으로의 ICT 사용은 읽기 성취에 부적인 영향을 미친 반면, 컴퓨터에 대한 태도는 읽기 성취에 정적인 영향을 미치는 것으로 나타났다. 학교관련 변인으로는 설립유형에 따라 청소년의 읽기 성취에 차이가 있었으며, 국공립에 비해 사립학교에 다니는 학생의 읽기 성취가 더 높은 것으로 나타났다.

2. 수학 성취 영향변수

PISA 2018에서는 PISA 2012와 동일하게 수학 소양을 정의하였으며, 수학 소양은 “다양한 맥락 속에서 수학을 형식화하고, 이용하고, 해석하는 개인적인 능력이다. 여기에는 현상을 기술하고, 설명하고, 예측하기 위해 수학적으로 추론하고 수학적 개념·절차·사실·도구를 사용하는 것이 포함된다. 수학 소양은 개인이 실세계에서 수학의 역할을 인식하고 건설적이고 참여적이며 반성적인 시민에게 요구되는 근거 있는 판단과 결정을 할 수 있도록 돕는다”고 하였다(조성민 외, 2019, p.87). 청소년의 수학 성취 역시 다양한 학생 및 학교와 관련된 변수에 의해 영향을 받는 것으로 나타났다.

PISA 2015 자료를 활용해 청소년의 수학 성취를 분석한 김수혜(2018)는 학생배경 요인인 부모의 교육수준이 높을수록 수학 성취가 높은 것으로 밝혔다. 학생과정 요인 중 성취동기, 학업시간, 부모의 정서적 지원은 수학 성취에 정적인 영향을 미치는 것으로 나타났다. 또한 진로관련 요인으로 직업에 대한 기대 수준이 높을수록 수학 성취가 높은 것으로 나타났다. 한편, 학교관련 변인 중 학부모가 인식하는 평균 학교교육의 질은 학생의 수학 성취에 정적인 영향을 미친 반면, 학교 평균 교사의 차별은 수학 성취에 부적인 영향을 미치는 것으로 나타났다.

PISA 2009를 활용하여 분석한 김혜숙(2012)의 연구에서도 학생배경 요인인 가정의 ESCS가 높을수록 수학 성취가 높은 것으로 나타났다. ICT 요인인 학교에서 ICT 사용과 레저 목적(오락, 인터넷 등)으로의 ICT 사용은 수학 성취에 부적인 영향을 미친 반면, 고차원 과제에 대한 자신감과 컴퓨터에 대한 태도는 수학 성취에 정적인 영향을 미치는 것으로 나타났다.

PISA 2003 자료를 활용해 수학 성취에 영향을 미치는 변수를 살펴본 윤정일과 이범수(2006)의 연구에서는 학생배경 요인인 성별에 따라 수학 성취에 차이가 있었으며, 여학생에 비해 남학생의 수학 성취가 더 높은 것으로 나타났다. 또한 가정의 ESCS와 최종 학력에 대한 기대 수준(대학교 및 대학원)은 수학 성취에 긍정적인 영향을 미쳤다. 사회·환경적 요인인 교사와 학생의 관계도 청소년의 수학 성취에 정적인 영향을 미쳤다. 학교관련 변수도 수학 성취에 영향을 미쳤는데, 학교위치, 교사부족, 학교장이 인식한 학생의 사기는 수학 성취에 긍정적인 영향을 미친 반면, 수학교사 1인당 학생 수는 수학 성취에 부정적인 영향을 미치는 것으로 나타났다.

3. 과학 성취 영향변수

PISA 2018에서는 PISA 2015와 마찬가지로 과학 소양을 “성찰적인 개인(시민)으로서 과학적인 사고를 발휘해 과학과 관련된 문제에 참여할 수 있는 능력”으로 개념화하였다(조성민 외, 2019, p.125). 과학 소양을 위해서는 현상에 대한 과학적인 설명, 과학 탐구의 평가와 설계, 자료와 증거에 대한 과학적인 해석 등 세 가지의 역량이 필요하다고 하였다(조성민 외, 2019). 선행연구에서 청소년의 과학 성취는 학생 또는 학교와 관련된 다양한 변수에 의해 영향을 받는 것으로 밝혀져 왔다.

PISA 2015 자료를 활용해 청소년의 과학 성취를 연구한 이정민과 정혜원(2019)의 연구에서는 학생과정 변인인 학업적 자기효능감이 과학 성취에 정적인 영향을 미치는 것으로 나타났다. 또한 사회·환경적 요인 중 교사의 관심과 지원을 높게 인식할수록 과학 성취가 높아진 반면, 수업에서 피드백을 활발히 제공할수록 오히려 청소년의 과학 성취가 낮아지는 것으로 나타났다.

PISA 2012 자료를 활용한 구자옥 외(2015)는 학생관련 변수 중 학생배경 요인인 가정의 교육자원 지표, 가정의 문화적 자산 지표가 청소년의 과학 성취에 긍정적인 영향을 미친다고 하였으며, 학생과정 요인인 방과 후 공부시간(숙제)도 과학 성취에 정적인 영향을 미쳤다. 또한 사회·환경적 요인인 학습활동에 대한 태도는 과학 성취에 정적인 영향을 미쳤으며, 학교소속감은 오히려 한국 남학생의 과학 성취에 부적인 영향을 미치는 것으로 나타났다. 이 밖에도 학교관련 변수인 학교 분위기에 영향을 미치는 학생 행동은 과학 성취에 정적인 영향을, 학교 평균 교사와 학생의 관계는 여학생의 과학 성취에만 긍정적인 영향을 미치는 것으로 나타났다.

PISA 2009를 활용하여 분석한 김혜숙(2012)은 학생배경 요인 중 가정의 ESCS가 과학 성취에 정적인 영향을 미친다고 하였다. 또한 학생과정 요인인 학습시간(과학)은 과학 성취에 정적인 영향을 미쳤다. ICT와 관련하여 학교에서 ICT 사용과 레저 목적(오락, 인터넷 등)으로의 ICT 사용은 과학 성취에 부적인 영향을 미친 반면, 고차원 과제에 대한 자신감과 컴퓨터에 대한 태도는 과학 성취에 정적인 영향을 미치는 것으로 나타났다. 학교관련 변인 중 설립유형에 따라서도 과학 성취에 차이가 나타났으며, 국공립에 비해 사립학교에 다니는 학생의 과학 성취가 더 높은 것으로 나타났다. 특히 김혜숙(2012)은 읽기, 수학, 과학 성취를 모두 분석하였는데, 영역에 따라 영향을 미치는 예측변수가 일부 상이한 결과를 보였다.

III. 연구방법

1. 연구대상

본 연구는 OECD에서 회원국과 비회원국 총 79개국의 대상으로 실시한 국제 비교자료인 PISA 2018의 한국 자료를 활용하였다. PISA 2018은 중학교 또는 고등학교에 재학하고 있는 만 15세 청소년

년을 대상으로 조사를 실시하였으며, 본 연구에서는 한국의 188개 학교에서 표집한 6,650명을 연구 대상으로 선정하였다. 이 중 남학생은 3,459명(약 52%), 여학생은 3,191명(약 48%)이 분석에 포함되었다.

2. 측정도구

본 연구의 종속변수는 읽기, 수학 및 과학 성취로, PISA에서 각 인지적 성취 수준에 대한 평가 결과는 10개의 유의측정값(plausible values: PV)으로 제공된다. PISA의 경우 개별 학생이 모든 문항에 응답하는 것이 아니기 때문에 문항반응이론(item response theory: IRT)에 근거해 학생의 능력모수를 추정하고, 능력모수의 사후 분포로부터 임의로 10개의 값을 추출하게 된다(OECD, 2017). 이 때 유의측정값은 OECD 회원국 응시생들의 읽기, 수학 및 과학 성취 평균을 500점, 표준편차를 100점으로 척도화한 후 제공된다(OECD, 2017). 유의측정값을 분석에 활용할 때 평균을 사용하거나 한 개의 유의측정값만을 사용하게 되면 표준오차의 편이나 측정오차 등이 발생할 수 있어 10개 유의측정값을 모두 활용하는 것이 권장되고 있으므로(OECD, 2020), **본 연구는 10개의 유의측정값을 모두 이용하여 분석한 후 통합된 결과를 제시하였다.**

예측변수는 PISA 2018에서 조사한 수백 개의 변수 중 분석에 활용할 수 있는 학생, 부모, 교사 및 학교와 관련된 변수를 활용하였으며, 이는 각 대상이 응답한 변수이다. 예측변수 전처리 과정 시 선행연구(정혜원 외, 2021)를 참고하여 파생변수(derived variable)를 우선적으로 포함하였다. 파생변수는 직접 관찰할 수 없는 잠재요인(예, ESCS(가정의 경제, 사회 및 문화적 배경 지표))을 측정하기 위해 결합된 변수로, 하나 또는 여러 개의 문항을 재코딩하여 변환하거나 문항반응이론을 활용해 스케일링하는 등의 방법으로 하나의 지표로 구성하게 된다(OECD, 2017). 따라서 파생변수를 생성하는 데 포함된 개별문항의 경우 분석에서 제외하였고, 파생변수 생성에 활용되지 않은 문항의 경우에는 개별문항으로 분석에 포함하였다(정혜원 외, 2021). 개별문항 중 범주형 변수의 경우 더미코딩하여 분석에 활용하였고(예, 남=0, 여=1), 교사관련 변수는 학교별 평균을 내어 활용하였다. 또한 결측이 30%를 넘는 예측변수는 분석에서 제외하였으며, 30% 미만의 결측에 대해서는 10회 대체하여 활용하였다. 결측 대체는 R 4.0.2의 mice(Ver.3.12.0) 패키지를 사용하였다(van Buuren & Groothuis-Oudshoorn, 2011).

이러한 데이터 정리 과정을 통해 **총 493개의 예측변수가 분석에 활용되었으며**, 구체적으로 학생관련 변수 196개, 부모관련 변수 93개, 교사관련 변수 106개, 학교관련 변수 98개가 포함되었다. 변수명과 변수 구분은 선행연구(정혜원 외, 2021; 한국교육과정평가원, 2019)를 참조하였으며, 자세한 예측변수는 다음의 <표 1>과 같다.

〈표 1〉 예측변수

구분	요인	변수명	변수 수
학생 배경		ESCS, 아버지의 사회경제적업수준, 어머니의 사회경제적업수준, 아버지의 교육수준, 어머니의 교육수준, 부모님의 교육수준, 가정의 문화적 보유자산, 가정의 교육적 자산, 가정의 부유함, ICT 자산, 이민 상황 지표(2문항), 학력변경 횟수, 학년유예, 전학횟수, 나이, 가정에서 사용 가능한 ICT, 성별, 집에서 주로 사용하는 언어, 초등학교 재학 중 전학 여부, 중학교 재학 중 전학 여부, 학습 계열 변경 여부, 연간 가계 소득	22
		읽기 전략-요약, 읽기 전략-이해 및 기억, 디지털 읽기 전략-질과 신뢰성에 대한 평가, 읽기에 대한 즐거움, 읽기에 대한 능력 지각, 읽기에 대한 어려움 지각, 긍정적 감정, 행복감, 실패에 대한 두려움, 숙달 목표 지향, 회복탄력성, 작업 완성도, 총 공부시간, 공부시간(각 영역), PISA 검사에 대한 노력, PISA 응답에 대한 노력, PISA 검사 어려움에 대한 인식, 삶의 만족도, 따돌림 경험(3문항), 무단 결석/결과/지각 여부(3문항), 따돌림에 대한 의견(5문항), 학습기회-텍스트 길이, 다양한 자료 읽기 활동(5문항), 학생의 책 읽는 방식, 뉴스를 읽는 방식, 읽기 시간, 온라인 읽기 활동(6문항), 작업 완성도(1문항), 지능의 가변 이론(1문항), 주관적 삶의 안녕감(6문항), 기술습득 여부(15문항), 정규 수업 외에 추가로 받는 교육(9문항), 등교 전 공부시간(2문항), 방과 후 공부 시간(2문항), 등교 전이나 방과 후 공부 이유(7문항), 해외 교환학생 프로그램 참여 희망 여부	88
학생 관련	진로 관련	학생이 기대하는 직업 지위, 직업에 대한 정보, 학교에서 제공하는 노동시장에 대한 정보, 학교 밖에서 제공되는 노동시장에 대한 정보, 최종 학력에 대한 기대(5문항), 미래 직업 결정에 중요한 요소(11문항)	196
		ICT에 대한 흥미, ICT역량 인지, ICT 사용에 대한 지각된 자율성, 학교에서 사용 가능한 ICT, 학교에서 일상적인 ICT 사용, 학교 밖에서의 ICT 사용(학교 활동 목적), 학교 밖에서의 ICT 사용(레저 목적), 사회적 상호작용의 주체로서의 ICT, 수업에서 과목과 관련된 ICT 사용, 수업이외의 주제와 관련된 ICT 사용, 디지털 기기를 처음 사용한 시기, 인터넷에 처음 접속한 시기, 학교에서 인터넷 사용 정도, 학교 밖에서 인터넷 사용 정도(2문항), 학교 수업에서 학생 또는 교사의 디지털 기기 사용 여부(5문항), 학교 밖에서 학교 관련 활동을 위한 디지털 기기 사용 빈도	21
사회· 환경적		학교 소속감, 경쟁력, 학교에서의 경쟁에 대한 인식, 학생들의 협동, 학습활동에 대한 태도, 교사의 흥미 인식, 교사의 직접적 지시, 수업에서의 피드백, 지시에 대한 순응, 학생이 인식한 읽기 참여에 대한 교사 자극, 국어 수업에서의 징벌적 분위기, 국어 수업에서의 교사의 지지, 부모의 정서적 지지, 유아기 교육과 돌봄, 학교 공부에 대한 주변의 지원(5문항), 교사 지원(3문항), 학습기회-자료(4문항), 학습 기회-국어 수업 과제 및 숙제(9문항), 디지털 국어 학습 기회(7문항)	44
		현재 가정에서 학습에 대한 부모의 지원, 과거 가정에서 학습에 대한 부모의 지원, 부모의 정서적 지원, 부모의 읽기에 대한 즐거움, 학부모 참여를 위한 학교 정책, 부모가 인식하는 학교의 질, 자녀가 10살 무렵 자료를 읽은 빈도(4문항), 부모가 취미로 독서하는 시간, 부모가 다양한 읽기 자료를 읽은 빈도(11문항), 부모가 책 읽는 방식, 부모의 뉴스 읽는 방식, 12개월 동안 지출한 교육비, 자녀의 학교 선택권(2문항), 자녀의 학교 선택 시 중요한 요소(14문항), 학부모의 지난 학년 학교활동 참여도(10문항), 학부모의 학교활동 불참 사유(11문항), 자녀의 영유아 교육 및 보육 프로그램 참여 여부(3문항), 자녀의 영유아 교육 및 보육 프로그램 참여 나이(8문항), 자녀의 영유아 교육 및 보육 프로그램 참여 이유(3문항), 자녀의 영유아 교육 및 보육 프로그램 참여 정도, 자녀가 초등학교 1학년일 때 한 활동에서 사용한 언어, 다양한 사회 이슈나 역사에 대한 관심 정도(6문항), 자녀에게 기대하는 학력 수준(4문항), 자녀의 초등학교 시기 추가 국어 수업(2문항), 학부모 설문 응답자(3문항)	93
부모 관련	-	교사만족도(현재 근무환경), 교사만족도(교직), 교사효능감(학급운영), 교사효능감(교수환경), 교사효능감(학생들과의 긍정적 관계 유지), full time 여부, 교수를 위한 교환 및 조정, 교사들에 의해 제공되는 피드백, 교사(엄격한 정의-표준 교사 훈련), 교사(광범위한 정의-표준, 현직 또는 직업 기반 교사 훈련), 교사의 견해(교육자료 부족), 교사의 견해(교직원 부족), 교사의 ICT 응용 프로그램 사용, 독해력을 비율 기회, 교사의 전문적 지식과 효능감(2문항), 교사의 신념과 효능감(4문항), 교직에 대한 견해(2문항), 12개월 동안 참여한 전문성 개발 활동(5문항), 전문성 개발 활동 참여 여부(6문항), 전문성 개발 활동 참여 의무 여부, 전문성 개발 활동 주제(36문항), 학습 평가 방법(7문항), 읽기관련 수업 여부(2문항), 학생들이 수업에서 읽었던 가장 긴 지문, 디지털 읽기 역량 관련 수업 여부(7문항), 교사의 디지털 읽기 활동(6문항), 학교의 수업용 디지털 기기 사용에 관한 규정 여부, 교사의 책 읽는 방식, 교사의 뉴스 읽는 방식, 수업시간 외 업무관련 독서 시간, 출생 국가(현재 국가 여부), 다른 국가에서 공부 경험, 교사 근무 연수(2문항), 나이, 교사 자격증 취득 방법(4문항)	106

구분	요인	변수명	변수 수
학교 관련	-	학교크기, 학생-교사 비율, 설립유형(2문항), 교사 전체 수, 정규직 교사 비율, 학사졸업 교사 비율, 석사졸업 교사 비율, 박사졸업 교사 비율, 학생 1명당 이용 가능한 컴퓨터 수, 인터넷 연결 가능한 컴퓨터 비율, 학습도구 부족, 학습 보조원 부족, 교과 외 활동 제공, 학교 분위기에 영향을 미치는 학생 행동, 학교 분위기에 영향을 미치는 교사 행동, 학교위치, 학교 선택, 교사 수(2문항), 학생 중퇴 비율, 입학 사정 시 고려 사항(7문항), 지난 3개월 간 전문성 개발 연수프로그램을 참여한 교사의 비율, 성취도 결과의 책무성 이행 절차에의 활용 여부(3문항), 질 점검 및 개선을 위한 대책(10문항), 1학년 학생들의 능력별 편성 여부(2문항), 1학년 학생들의 특성(3문항), 학생 학습 지원(3문항), 학교 행사에 참여한 학부모의 비율(4문항),한국어가 모국어가 아닌 학생들에 대한 교내 평등 정책(5문항), 정규 수업 외 추가 국어 수업, 국어 수업 추가 실시 목적, 학생들의 평가 결과 사용(11문항), 교수 학습을 위한 ICT 도구 준비 수준(11문항), ICT 도구 활용하기 위한 전략(8문항), 해외학교에서 교사 방문이나 교사 교환 프로그램 유무, 학교의 진로 지도 담당자(5문항), 진로 지도 방법	98

주. 파생변수는 기울임(이탤릭체) 표기하였음.

3. 연구방법

본 연구에서는 PISA 2018 자료가 학생이 학교에 내재된 다층자료 구조를 반영할 수 있는 혼합효과 랜덤포레스트(Hajjem et al., 2014)를 적용하였다. 본 연구에서 활용한 혼합효과 랜덤포레스트 식은 다음과 같다.

$$Y_{ij} = f(X_{ij}) + Z_j b_j + e_{ij} \quad (1)$$

위 식(1)에서 Y_{ij} 는 학교 j 에 속하는 학생 개인 i 에 대한 종속변수 벡터이며, X_{ij} 는 예측변수 행렬, Z_j 는 임의효과 예측변수의 행렬, b_j 는 학교수준 임의효과, e_{ij} 는 학생수준 오차 벡터를 의미한다. 이때, 본 연구에서 Z_j 는 [1]로 고정하였다. $f(X_{ij})$ 는 일반적인 랜덤포레스트(Breiman, 2001)를 통해 추정되며, b_j 와 e_{ij} 는 상호 독립적이며 정규분포를 따르는 것으로 가정한다(Hajjem et al., 2014). 다층자료 구조를 반영할 수 있도록 전통적인 랜덤포레스트를 확장한 혼합효과 랜덤포레스트를 제시한 Hajjem 외(2014)의 연구에서는 혼합효과 랜덤포레스트의 예측력이 전통적인 랜덤포레스트에 비해 우수하였으며, 특히 상위수준의 임의효과가 클 경우에는 예측력 차이가 보다 큰 것으로 나타났다.

본 연구에서는 전체 데이터의 70%를 학습용 자료로 30%를 검증용 자료로 구분하였고, 학습용 자료로 훈련한 후, 새로운 데이터인 검증용 자료로 예측오차를 계산하였다. 또한 혼합효과 랜덤포레스트와 다층자료의 속성을 고려하지 않는 전통적인 랜덤포레스트(Breiman, 2001)를 적용해 각각 예측성 과를 비교하고, 상대적으로 예측성과가 좋은 기법을 적용해 주요 예측변수를 확인하였다. 이때, 한국 청소년의 읽기, 수학 및 과학 성취 관련 주요 예측변수는 SHAP(shapley additive explanations)를 통해 나타난 변수 중요도(SHAP value 절대값의 평균)를 기준으로 판별하였다. 본 연구는 각 인지적 성취의 유의측정값(PV) 10개를 모두 활용하였기 때문에 모든 과정을 10회 반복 분석하였으며, 10회 반복 분석에서 매번 변수 중요도가 높게 도출된 변수를 주요 예측변수라 하였다.

한국 청소년의 읽기, 수학 및 과학 성취에 영향을 미치는 주요 예측변수를 확인한 후, 도출된 주요

예측변수가 세 영역에 미치는 영향을 살펴보기 위해 마찬가지로 자료의 위계적 구조를 반영할 수 있는 전통적인 방법인 다층모형 분석을 실시하였다. 다층모형은 예측변수의 투입 여부에 따라 기초모형과 연구모형으로 구분되며, 기초모형은 예측변수를 투입하지 않은 모형으로 종속변수의 학생 및 학교 분산을 구분하게 된다. 다층모형의 기초모형 식은 다음과 같다(강상진, 2016).

$$\text{학생수준 } Y_{ij} = \beta_{0j} + r_{ij}, \quad r_{ij} \sim N(0, \sigma^2) \quad (2)$$

$$\text{학교수준 } \beta_{0j} = \gamma_{00} + u_{0j}, \quad u_{0j} \sim N(0, \tau_{00}) \quad (3)$$

위에서 학생 개인(i)과 학교(j)에 대하여 Y_{ij} 는 종속변수(읽기, 수학 및 과학 성취), β_{0j} 는 학교별 평균 종속변수, r_{ij} 는 개인 간 분산을, γ_{00} 는 전체 종속변수 평균, u_{0j} 는 학교 간 분산을 의미한다. 기초모형에 혼합효과 랜덤포레스트를 통해 도출된 주요 예측변수를 투입한 연구모형은 다음과 같다.

$$Y_{ij} = \beta_{0j} + \sum_{i=1}^p \beta_{pj}(X_p)_{ij} + r_{ij}, \quad r_{ij} \sim N(0, \sigma^2) \quad (4)$$

$$\begin{cases} \beta_{0j} = \gamma_{00} + \sum_{j=1}^q \gamma_{0q}(Z_q)_j + u_{0j}, & u_{0j} \sim N(0, \tau_{00}) \\ \beta_{pj} = \gamma_{p0} \end{cases} \quad (5)$$

연구모형에서 X_p 는 학생수준 예측변수, Z_q 는 학교수준 예측변수, β_{0j} 는 학교별 평균 종속변수, γ_{00} 는 전체 종속변수 평균, γ_{0q} 는 학교수준 주요 예측변수의 회귀계수를 의미한다. β_{pj} 는 학교별 학생수준 예측변수의 회귀계수, γ_{p0} 는 전체 학생수준 주요 예측변수 회귀계수를 의미하며, 본 연구에서는 모형의 간명성을 위해 학교 간 차이가 없다고 가정하는 무선절편(random intercept) 모형으로 분석을 실시하였다(Bryk & Raudenbush, 1992). 혼합효과 랜덤포레스트 분석과 마찬가지로 다층모형도 10개의 유의측정값을 활용해 분석을 반복한 후 통합된 결과를 도출하였다.

혼합효과 랜덤포레스트 분석에는 파이썬(Python)의 MERF(Hajjem et al., 2014) 라이브러리를, 랜덤포레스트 분석에는 파이썬의 sklearn(RandomForestRegressor) 라이브러리를 활용하였다. 다층모형 분석은 Mplus 8.3(Muthén & Muthén, 1998-2020) 프로그램을 사용하였다.

IV. 연구결과

1. 혼합효과 랜덤포레스트와 랜덤포레스트 간 예측성과 비교

혼합효과 랜덤포레스트와 전통적인 랜덤포레스트 간 예측오차는 평균제곱근오차(RMSE)를 통해

확인하였으며, 읽기, 수학 및 과학 성취의 10개 유의측정값을 사용하여 두 기법을 각각 10회 반복 분석한 평균 결과값을 <표 2>에 제시하였다. 그 결과, 읽기, 수학 및 과학 성취에서 모두 혼합효과 랜덤 포레스트가 전통적인 랜덤포레스트보다 평균제곱근오차가 낮은 것으로 나타났다. 이에 본 연구는 상대적으로 예측성도가 우수한 혼합효과 랜덤포레스트를 통해 읽기, 수학 및 과학 성취 관련 주요 예측 변수를 탐색하였다.

<표 2> 혼합효과 랜덤포레스트와 랜덤포레스트 평균제곱근오차 비교

영역	혼합효과 랜덤포레스트			랜덤포레스트		
	평균제곱근오차			평균제곱근오차		
	Mean	Min	Max	Mean	Min	Max
읽기	69.811	68.833	70.610	69.872	68.925	70.800
수학	69.222	67.946	70.514	70.675	69.452	71.771
과학	67.457	65.616	68.283	68.600	66.915	69.621

2. 혼합효과 랜덤포레스트 분석 결과

본 연구는 혼합효과 랜덤포레스트 분석을 통해 청소년의 읽기, 수학 및 과학 성취의 주요 예측변수를 탐색하였다. 주요 예측변수는 투입된 493개 예측변수의 변수 중요도를 기준으로 살펴보았으며, 각 인지적 성취의 10개 유의측정값을 모두 활용하였기 때문에 분석 과정을 10회 반복 실시한 후 10번 모두 변수 중요도가 상위 30위 안에 포함된 변수를 확인하였다. 그 결과, 읽기 성취는 23개, 수학 성취는 15개, 과학 성취는 14개가 각 영역에서 주요 예측변수로 선택되었으며, 선택된 변수를 <표 3>~<표 5>에 제시하였다. 분석에 포함된 다양한 학생, 부모, 교사 및 학교 변수 중 주요 예측변수는 부모관련 변수 1개를 제외하고 대부분 학생관련 변수가 선택되었다. 교사 및 학교관련 변수는 변수 중요도가 상대적으로 낮아 주요 예측변수로 선택되지 않았다.

가. 읽기 성취

분석에 포함된 493개 예측변수 중 읽기 성취 관련 주요 예측변수는 23개가 선택되었으며, 이는 <표 3>과 같다. 구체적으로 각 요인별 선택된 예측변수를 살펴보면, 학생관련 변수 중 학생배경 요인 2개, 학생과정 요인 11개, 진로관련 요인 3개, ICT 요인 6개 및 부모관련 변수 1개가 선택되었다. 학생배경 요인에서는 가정의 'ESCS'와 '가정에서 사용 가능한 ICT'가 주요 예측변수로 도출되었다. 학생과정 요인 중에서는 '읽기 전략-요약', '읽기 전략-이해 및 기억', '디지털 읽기 전략-질과 신뢰성에 대한 평가', '읽기에 대한 능력 지각', '읽기에 대한 어려움 지각', '읽기에 대한 즐거움' 및 '온라인 읽기 활동(온라인으로 정보 탐색)' 등과 같이 읽기와 관련된 변수가 도출되었다. 또한 PISA 검사에 대해 직접적으로 묻고 있는 'PISA 검사 어려움에 대한 인식', 'PISA 검사에 대한 노력', 'PISA 응답에 대한 노력', '총 학습시간'

이 주요 예측변수로 도출되었다. 진로관련 요인 중에는 ‘학생이 기대하는 직업 지위’, ‘미래 직업 결정에 중요한 요소(친한 친구의 미래 계획)’와 ‘직업에 대한 정보’가 선택되어 직업이나 진로와 관련된 변수가 읽기 성취의 주요 예측변수임을 확인하였다. ICT 요인 중에는 ‘ICT 사용에 대한 지각된 자율성’, ‘ICT에 대한 흥미’, ‘학교에서 사용 가능한 ICT’, ‘수업에서 과목과 관련된 ICT 사용’, ‘학교에서 일상적인 ICT 사용’ 및 ‘학교 밖에서의 ICT 사용(학교 활동 목적)’이 주요 예측변수로 도출되었다. 또한 부모관련 변수 중 ‘자녀가 10살 무렵 자료를 읽은 빈도(책)’도 주요 예측변수로 선택되었다. 이 중 ‘읽기 전략-요약’은 10회 반복 실시된 분석에서 모두 가장 주요 예측변수로 도출되었다. 다음으로 ‘PISA 검사 어려움에 대한 인식’, ‘디지털 읽기 전략-질과 신뢰성에 대한 평가’, ‘읽기에 대한 능력 지각’이 10회 반복 실시한 결과에서 모두 2-4위로 나타나, 중요도 지수 상위 1-4위 변수는 매번 동일하게 도출되었다.

〈표 3〉 읽기 성취의 주요 예측변수

구분	요인	변수 수	변수
학생관련	학생배경	2	ESCS
			가정에서 사용 가능한 ICT
			읽기 전략-요약
			읽기 전략-이해 및 기억
	학생과정	11	디지털 읽기 전략-질과 신뢰성에 대한 평가
			읽기에 대한 능력 지각
			읽기에 대한 어려움 지각
			읽기에 대한 즐거움
			온라인 읽기 활동(온라인으로 정보 탐색)
			PISA 검사 어려움에 대한 인식
			PISA 검사에 대한 노력
			PISA 응답에 대한 노력
			총 학습시간
	진로관련	3	학생이 기대하는 직업 지위
			미래 직업 결정에 중요한 요소(친한 친구의 미래 계획)
			직업에 대한 정보
			ICT 사용에 대한 지각된 자율성
	ICT	6	ICT에 대한 흥미
			학교에서 사용 가능한 ICT
			수업에서 과목과 관련된 ICT 사용
			학교에서 일상적인 ICT 사용
			학교 밖에서의 ICT 사용(학교 활동 목적)
부모관련		1	자녀가 10살 무렵 자료를 읽은 빈도(책)

나. 수학 성취

수학 성취의 주요 예측변수로는 15개의 변수가 선택되었으며, 자세한 변수를 〈표 4〉에 제시하였다.

학생관련 변수 중 학생배경 요인 2개, 학생과정 요인 11개, 진로관련 요인 1개, ICT 요인 1개가 선택되었다. 학생배경 요인에서는 ‘ESCS’와 ‘연간 가계 소득’이 주요 예측변수로 도출되었다. 학생과정 요인 중에는 ‘읽기 전략-요약’, ‘읽기 전략-이해 및 기억’, ‘디지털 읽기 전략-질과 신뢰성에 대한 평가’, ‘읽기에 대한 능력 지각’, ‘읽기에 대한 즐거움’, ‘PISA 검사 어려움에 대한 인식’, ‘PISA 검사에 대한 노력’, ‘PISA 응답에 대한 노력’, ‘총 학습시간’, ‘정규 수업 외에 받는 추가적인 교육(수학 과목)’과 ‘방과 후 공부 시간’이 주요 예측변수로 도출되었다. 진로관련 요인 중에는 ‘최종 학력에 대한 기대(대학교 및 대학원)’이 도출되었다. 또한 ICT 요인 중에는 ‘ICT 사용에 대한 지각된 자율성’이 주요 예측변수로 도출되었다. 이 중 ‘읽기 전략-요약’은 10회 반복 실시된 분석에서 모두 가장 주요 예측변수로 나타났다.

〈표 4〉 수학 성취의 주요 예측변수

구분	요인	변수 수	변수
학생관련	학생배경	2	ESCS
			연간 가계 소득
			읽기 전략-요약
			읽기 전략-이해 및 기억
			디지털 읽기 전략-질과 신뢰성에 대한 평가
			읽기에 대한 능력 지각
	학생과정	11	읽기에 대한 즐거움
			PISA 검사 어려움에 대한 인식
			PISA 검사에 대한 노력
			PISA 응답에 대한 노력
			총 학습시간
			정규 수업 외에 받는 추가적인 교육(수학 과목)
	진로관련	1	방과 후 공부 시간
			최종 학력에 대한 기대(대학교 및 대학원)
	ICT	1	ICT 사용에 대한 지각된 자율성

다. 과학 성취

493개 예측변수 중 과학 성취 관련 주요 예측변수로는 14개의 변수가 선택되었으며, 이는 〈표 5〉와 같다. 과학 성취의 주요 예측변수는 모두 학생관련 변수가 선택되었으며, 학생관련 변수 중 학생과정 요인 8개, 진로관련 요인 2개, 사회·환경적 요인 1개, ICT 요인 3개가 선택되었다. 학생과정 요인에서는 ‘읽기 전략-요약’, ‘읽기 전략-이해 및 기억’, ‘디지털 읽기 전략-질과 신뢰성에 대한 평가’, ‘읽기에 대한 능력 지각’, ‘읽기에 대한 즐거움’, ‘PISA 검사 어려움에 대한 인식’, ‘PISA 검사에 대한 노력’, ‘PISA 응답에 대한 노력’이 주요 예측변수로 도출되었다. 진로관련 요인 중에서 ‘미래 직업 결정에 중요한 요소(친한 친구의 미래 계획)’과 ‘학생이 기대하는 직업 지위’가 도출되었고, 사회·환경적 요인으로는 ‘학교 소속감’이 주요 예측변수로 나타났다. ICT 요인 중에는 ‘ICT 사용에 대한 지각된 자율성’, ‘수업에서 과목과 관련된 ICT 사용’, ‘학교에서 일상적인 ICT 사용’이 도출되었다. 이 중 과학 성취에서도 마찬가지로 ‘읽기 전략-요약’이 10회 반복 실시된 분석에서 모두 가장 주요 예측변수로 도출되었

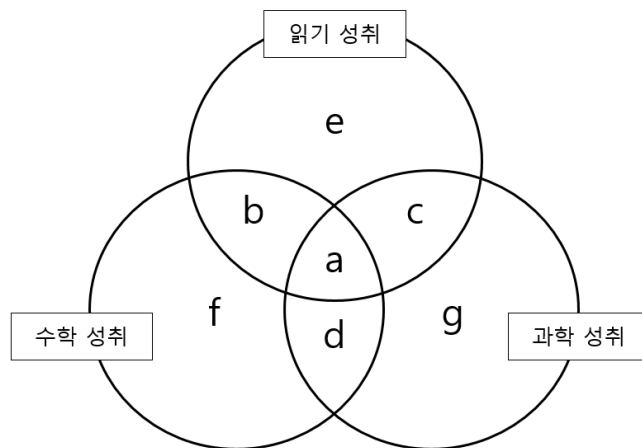
다. 다음으로 'PISA 검사 어려움에 대한 인식', '디지털 읽기 전략-질과 신뢰성에 대한 평가'가 10회 반복 실시한 분석결과에서 모두 각각 2위, 3위로 나타나, 중요도 지수 상위 1-3위 변수는 매번 동일하게 도출되었다.

〈표 5〉 과학 성취의 주요 예측변수

구분	요인	변수 수	변수
학생관련	학생과정	8	읽기 전략-요약
			읽기 전략-이해 및 기억
			디지털 읽기 전략-질과 신뢰성에 대한 평가
			읽기에 대한 능력 지각
			읽기에 대한 즐거움
			PISA 검사 어려움에 대한 인식
			PISA 검사에 대한 노력
			PISA 응답에 대한 노력
	진로관련	2	미래 직업 결정에 중요한 요소(친한 친구의 미래 계획)
	사회·환경적	1	학생이 기대하는 직업 지위
ICT	ICT	3	학교 소속감
			ICT 사용에 대한 지각된 자율성
			수업에서 과목과 관련된 ICT 사용
			학교에서 일상적인 ICT 사용

라. 읽기, 수학 및 과학 성취의 주요 예측변수 비교

앞서 읽기, 수학 및 과학 성취의 주요 예측변수를 각각 도출한 결과를 확인하였다. 세 영역의 결과를 비교하기 위해 도출된 주요 예측변수를 [그림 1]과 같이 구분하였다.



[그림 1] 세 영역의 주요 예측변수 비교 도식화

[그림 1]에 따라 읽기, 수학 및 과학 성취의 주요 예측변수를 구분하면 <표 6>과 같다. 세 영역에서 공통적으로 도출된 주요 예측변인은 총 9개로 나타났다([그림 1]의 a부분). 9개의 변수 중 학생과정 요인 8개('읽기 전략-요약', '읽기 전략-이해 및 기억', '디지털 읽기 전략-질과 신뢰성에 대한 평가', '읽기에 대한 능력 지각', '읽기에 대한 즐거움', 'PISA 검사 어려움에 대한 인식', 'PISA 검사에 대한 노력', 'PISA 응답에 대한 노력')와 ICT 요인 1개('ICT 사용에 대한 지각된 자율성')가 공통적으로 나타났다. 읽기와 수학 성취에 공통적으로 나타난 주요 예측변수([그림 1]의 b부분)는 'ESCS'와 '총 학습시간'이었다. 읽기와 과학 성취에 공통적으로 나타난 변수([그림 1]의 c부분)는 '학생이 기대하는 직업 지위', '미래 직업 결정에 중요한 요소(친한 친구의 미래 계획)', '수업에서 과목과 관련된 ICT 사용', '학교에서 일상적인 ICT 사용'으로 나타났다. 수학과 과학 성취 간에는 공통적인 주요 예측변수가 선택되지 않았다([그림 1]의 d부분).

반면 특정 영역에서만 도출된 주요 예측변수도 나타났다. 읽기 성취에서만 주요 예측변수로 도출된 변수를 살펴보면 학생관련 변수 중 '가정에서 사용 가능한 ICT', '읽기에 대한 어려움 지각', '온라인 읽기 활동(온라인 정보 탐색)', '직업에 대한 정보', 'ICT에 대한 흥미', '학교에서 사용 가능한 ICT', '학교 밖에서의 ICT 사용'과 부모관련 변수 중 '자녀가 10살 무렵 자료를 읽은 빈도(책)'가 선택되었다([그림 1]의 e). 또한 수학 성취에서만 '연간 가계 소득', '최종 학력에 대한 기대(대학교 및 대학원)', '정규 수업 외에 받는 추가적인 교육(수학 과목)' 및 '방과 후 공부 시간'이 주요 예측변수로 나타났다([그림 1]의 f). 학생관련 변수 중 사회·환경적 요인인 '학교 소속감'은 과학 성취에만 주요 예측변수로 도출되었다([그림 1]의 g).

<표 6> 주요 예측변수 비교

그림 부분	구분	요인	변수	M	SD	Min	Max
a	학생 관련	학생과정	읽기 전략-요약*	-0.194	1.100	-1.720	1.360
			읽기 전략-이해 및 기억*	-0.172	1.081	-1.640	1.500
			디지털 읽기 전략-질과 신뢰성에 대한 평가*	-0.297	0.993	-1.410	1.330
			읽기에 대한 능력 지각*	-0.189	0.951	-2.440	1.884
			읽기에 대한 즐거움*	0.238	0.911	-2.711	2.657
			PISA 검사 어려움에 대한 인식*	0.075	1.048	-1.272	3.006
			PISA 검사에 대한 노력	8.250	1.920	1.000	10.000
			PISA 응답에 대한 노력	9.080	1.629	1.000	10.000
		ICT	ICT 사용에 대한 지각된 자율성*	-0.208	0.956	-2.514	2.026
b	학생 관련	학생배경	ESCS*	0.085	0.772	-3.051	3.959
		학생과정	총 학습시간(분)	1834	428	275	3000
c	학생 관련	진로관련	학생이 기대하는 직업 지위	62.015	18.761	11.010	88.960
			미래 직업 결정에 중요한 요소(친한 친구의 미래 계획)	2.000	0.936	1.000	4.000
		ICT	수업에서 과목과 관련된 ICT 사용*	0.077	1.121	-1.219	2.439

그림 부분	구분	요인	변수	M	SD	Min	Max
e	학생 관련	학생배경	학교에서 일상적인 ICT 사용*	-0.722	0.996	-1.716	3.304
			가정에서 사용 가능한 ICT	7.650	2.244	0.000	11.000
		학생과정	읽기에 대한 어려움 지각*	0.201	0.973	-1.888	2.775
			온라인 읽기 활동(온라인으로 정보 탐색)	3.980	1.137	1.000	5.000
		진로관련	직업에 대한 정보*	0.029	0.916	-1.917	2.250
			ICT에 대한 흥미*	-0.112	0.971	-2.934	2.619
	ICT	학교에서 사용 가능한 ICT	6.350	2.590	0.000	10.000	
		학교 밖에서의 ICT 사용(학교 활동 목적)*	-0.032	0.913	-2.301	3.307	
	부모 관련	-	자녀가 10살 무렵 자료를 읽은 빈도(책)	3.150	0.847	1.000	4.000
		f	학생배경	연간 가계 소득	4.350	1.710	1.000
학생과정	정규 수업 외에 받는 추가적인 교육(수학)+		0.400	0.490	0.000	1.000	
	방과 후 공부 시간		3.370	2.183	0.000	9.000	
진로관련	최종 학력에 대한 기대(대학교 및 대학원)+		0.790	0.409	0.000	1.000	
g	학생 관련	사회· 환경적	학교 소속감*	0.284	1.051	-3.237	2.756

* : 평균이 0, 표준편차가 1인 표준화 지수로 변환한 변인

+ : 더미 변수(예=1, 아니오=0)

마. 주요 예측변수의 영향력 검증

다음으로 혼합효과 랜덤포레스트를 통해 도출된 주요 예측변수가 한국 청소년의 읽기, 수학 및 과학 성취에 미치는 영향을 검증하기 위해 추가적으로 다층모형 분석을 실시하였으며, 결과는 <표 7>과 같다³⁾. 세 영역의 결과를 비교하기 위해 공통적으로 도출된 주요 예측변수뿐만 아니라 각 영역에서 차별적으로 도출된 예측변수도 동일하게 투입하였다. 우선 예측변수를 투입하지 않은 기초모형에서 ICC (급내상관계수)를 확인한 결과, 읽기 성취는 .260, 수학 성취는 .306, 과학 성취는 .292로, 전체 분산 중 약 26.0~30.6%가 학교수준 분산에 의해 설명되는 것으로 나타났다.

<표 7>을 보면, 학생배경 변수인 'ESCS'는 읽기, 수학 및 과학 성취에 모두 정적인 영향을 미치는 것으로 나타났다. '연간 가계 소득'의 경우에는 영역에 따라 상이한 결과가 확인되었는데, 수학 성취에는 정적인 영향을 미쳤으나 읽기와 과학 성취에는 유의한 영향을 미치지 않는 것으로 나타났다. 이와 달리 '가정에서 사용 가능한 ICT'는 읽기 및 과학 성취에 부적인 영향을 미치는 것으로 나타났으며, 수학 성취에는 유의한 영향을 미치지 않았다.

3) 학생과정 변수 중 수학 성취에만 주요 예측변수로 도출된 '정규 수업 외에 받는 추가적인 교육(수학)'의 경우 특정 영역과 관련된 변수이므로 다층모형 분석에서 제외하였음. 다층모형 분석에 앞서 주요 예측변수 간 다중공선성 문제를 VIF(분산팽창계수)를 확인한 결과, 다중공선성 기준(VIF 10 미만)에 충족하는 것으로 나타났음(VIF 1.020~1.844).

학생과정 변수인 ‘읽기 전략-요약’, ‘읽기 전략-이해 및 기억’, ‘디지털 읽기 전략-질과 신뢰성에 대한 평가’, ‘읽기에 대한 능력 지각’, ‘PISA 검사에 대한 노력’, ‘PISA 응답에 대한 노력’ 및 ‘온라인 읽기 활동(온라인으로 정보 탐색)’은 세 영역에 모두 정적인 영향을 미치는 것으로 나타났다. 반면, ‘PISA 검사 어려움에 대한 인식’과 ‘총 학습시간’은 세 영역에 모두 부적인 영향을 미쳤다. 영역에 따라 다른 결과도 나타났는데, ‘읽기에 대한 즐거움’은 읽기 성취에 긍정적인 영향을 미쳤으나 수학과 과학 성취에는 유의한 영향을 미치지 않았으며, ‘방과 후 공부 시간’은 수학 및 과학 성취에 긍정적인 영향을 미쳤으나, 읽기 성취에는 유의한 영향을 미치지 않았다.

진로관련 변수인 ‘학생이 기대하는 직업 지위’와 ‘최종 학력에 대한 기대(대학교 및 대학원)’는 세 영역에 모두 정적인 영향을 미치는 것으로 나타났다. 이는 한국 청소년이 미래의 직업 지위와 최종 학력에 대해 높게 기대할수록 인지적 성취가 높아짐을 의미한다. 반면, ‘미래 직업 결정에 중요한 요소(친한 친구의 미래 계획)’는 세 영역에 모두 부적인 영향을 미치는 것으로 나타났다. 즉, 미래 직업을 결정하는 데 친한 친구의 미래 계획을 중요하게 생각하는 것은 읽기, 수학 및 과학 성취에 부정적인 영향을 미치는 것으로 나타났다.

ICT 변수 중 ‘ICT 사용에 대한 지각된 자율성’은 세 영역에 모두 정적인 영향을 미치는 것으로 나타난데 반해, ‘학교 밖에서의 ICT 사용(학교 활동 목적)’은 세 영역에 모두 부적인 영향을 미치는 것으로 나타났다. 이 외 변수는 영역에 따라 상이한 결과가 나타났는데, ‘ICT에 대한 흥미’의 경우에는 읽기 성취에 긍정적인 영향을 미치는 것으로 나타났으나 그 외 수학 및 과학 성취에는 유의한 영향을 미치지 않았다. ‘학교에서 일상적인 ICT 사용’은 읽기 및 과학 성취에 부적인 영향을 미치는 것으로 나타난 반면, ‘수업에서 과목과 관련된 ICT 사용’의 경우에는 읽기 및 과학 성취에 정적인 영향을 미치는 것으로 나타났다. 그러나 ‘학교에서 일상적인 ICT 사용’과 ‘수업에서 과목과 관련된 ICT 사용’은 수학 성취에서는 유의한 영향을 미치지 않는 것으로 나타났다. ‘학교에서 사용 가능한 ICT’는 어느 영역에도 유의한 영향을 미치지 못했다.

사회·환경적 변수인 ‘학교 소속감’은 읽기, 수학 및 과학 성취에 모두 부정적인 영향을 미치는 것으로 나타났다. 이와 달리 부모관련 변수인 ‘자녀가 10살 무렵 책을 읽은 빈도’는 읽기, 수학 및 과학 성취에 모두 긍정적인 영향을 미치는 것으로 나타났다.

〈표 7〉 읽기, 수학 및 과학 성취 다층모형 분석 결과

구분	변수	읽기 성취		수학 성취		과학 성취	
		회귀계수	표준오차	회귀계수	표준오차	회귀계수	표준오차
학생배경	절편	392.926***	12.234	383.434***	14.185	408.475***	11.337
	ESCS	5.110**	1.571	6.752***	1.728	3.981**	1.507
	연간 가계 소득	1.149	0.636	4.400***	0.653	1.030	0.658
	가정에서 사용 가능한 ICT	-2.015***	0.544	-0.765	0.616	-1.110*	0.534
	읽기 전략-요약	14.079***	1.071	12.722***	1.200	13.467***	1.164
	읽기 전략-이해 및 기억	9.657***	1.085	9.746***	1.271	8.862***	1.204
	디지털 읽기 전략-질과 신뢰성에 대한 평가	13.878***	1.127	11.206***	1.247	13.842***	1.188
	읽기에 대한 능력 지각	8.992***	1.343	8.334***	1.578	8.210***	1.613
	읽기에 대한 즐거움	4.719***	1.293	-1.885	1.357	2.089	1.344
	PISA 검사 어려움에 대한 인식	-12.470***	1.136	-10.319***	1.234	-13.636***	1.152
학생과정	PISA 검사에 대한 노력	3.972***	0.628	4.204***	0.680	3.819***	0.683
	PISA 응답에 대한 노력	5.473***	0.708	5.676***	0.747	5.038***	0.707
	읽기에 대한 어려움 지각	0.250	1.217	0.973	1.361	-1.124	1.167
	온라인 읽기 활동(온라인으로 정보 탐색)	5.952***	0.793	2.747**	0.975	3.433***	0.941
	총 학습시간	-0.010***	0.003	-0.008**	0.003	-0.008**	0.002
	방과 후 공부 시간	-0.136	0.459	2.171***	0.554	1.080*	0.497
	학생이 기대하는 작업 지위	0.413***	0.055	0.409***	0.057	0.358***	0.050
	미래 직업 결정에 중요한 요소(친한 친구의 미래 계획)	-9.591***	1.100	-6.808***	1.257	-7.239***	1.187
	직업에 대한 정보	0.001	1.165	0.362	1.531	-1.056	1.330
	최종 학력에 대한 기대(대학교 및 대학원)	18.986***	2.656	23.751***	2.887	14.218***	2.400
ICT	ICT 사용에 대한 자각된 자율성	3.214**	1.121	5.149***	1.278	6.789***	1.090
	ICT에 대한 흥미	4.555***	1.051	0.161	1.226	0.826	1.141
	학교에서 일상적인 ICT 사용	-6.897***	1.170	-1.514	1.349	-4.078***	1.122
	학교 밖에서의 ICT 사용 (학교 활동 목적)	-3.568**	1.313	-5.626***	1.517	-4.530**	1.451
	수업에서 과목과 관련된 ICT 사용	4.448***	0.902	2.342	1.382	6.243***	1.357
	학교에서 사용 가능한 ICT	0.623	0.439	0.138	0.586	0.073	0.513
	학교 소속감	-4.610***	0.927	-5.297***	1.079	-7.357***	1.050
	자녀가 10살 무렵 책을 읽은 빈도	8.711***	1.323	6.317***	1.713	8.454***	1.205
	분산(임의효과)						
	학생수준 분산	4523.049***	118.047	4616.617***	114.008	4102.785***	104.878
	학교수준 분산	524.206***	90.996	776.271***	136.675	668.962***	109.546

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

V. 논의 및 결론

본 연구는 학생이 학교에 내재된 다층자료인 PISA 2018에 혼합효과 랜덤포레스트와 전통적인 랜덤포레스트를 적용해 한국 청소년의 읽기, 수학 및 과학 성취 예측력을 비교하였으며, 상대적으로 예측성도가 우수하게 나타난 혼합효과 랜덤포레스트를 통해 한국 청소년의 읽기, 수학 및 과학 성취에 영향을 미치는 주요 예측변수를 도출하였다. 또한 다층모형 분석을 통해 도출된 주요 예측변수가 읽기, 수학 및 과학 성취에 어떠한 영향을 미치는지 검증하였다. 본 연구에서 나타난 주요 결과는 다음과 같다.

첫째, ‘읽기 전략-요약’, ‘읽기 전략-이해 및 기억’ 및 ‘디지털 읽기 전략-질과 신뢰성에 대한 평가’는 메타인지와 관련된 변수로, 청소년의 읽기, 수학 및 과학 성취에서 공통적인 주요 예측변수로 나타났다. 이는 전통적인 랜덤포레스트를 활용해 청소년의 읽기 성취 관련 주요 예측변수를 탐색한 선행연구(손윤희 외, 2020; 정혜원 외, 2021)와 유사한 결과이다. 주목할 점은 ‘읽기 전략-요약’, ‘읽기 전략-이해 및 기억’ 및 ‘디지털 읽기 전략-질과 신뢰성에 대한 평가’는 읽기 성취뿐만 아니라 수학 성취와 과학 성취에서도 상위 예측변수로 포함되었다는 것이다. 읽기 전략은 학생이 텍스트를 이해하는 데 도움이 되는 정서적 및 행동적 활동이므로(최숙기, 박기범, 2012), 요약 전략, 이해 및 기억 전략의 유용성을 올바르게 이해하여 개인에게 효과적인 읽기 전략을 학습하고 활용하는 방법을 훈련하는 것이 필요하다.

둘째, ‘읽기에 대한 능력 지각’과 ‘읽기에 대한 즐거움’은 읽기 성취 외에 수학 및 과학 성취에서도 주요 예측변수로 도출되었다. 이는 읽기에 대한 능력을 지각하고, 읽기에 대한 즐거움을 가지는 것이 범교과적으로 인지적 성취에 중요한 요인임을 의미한다. 선행연구(이향미, 정혜원, 2019)에서도 유사하게 초등학교 시기의 독서 활동이 국어뿐만 아니라 수학 학업성취도의 종단적 변화에도 긍정적인 영향을 미쳤다고 보고된 바 있다. 따라서 청소년이 스스로 읽기에 대한 능력을 지각하고 읽기에 대한 유능감을 높여줌으로써 읽기에 대한 내재적 동기를 유발하는 것이 필요하며, 이를 위해서는 개별 수준에 맞는 읽기 활동이나 과제를 부여하여 성공적인 읽기 경험을 지속적으로 할 수 있도록 체계적인 지원이 필요하다. 또한 읽기에 대한 즐거움을 향상시키기 위해 아동기에 가정에서부터 독서에 대한 관심을 높이고 관심 있는 책을 많이 접할 수 있는 기회를 제공해줄 수 있어야 할 것이다. 이는 부모가 응답한 ‘자녀가 10살 무렵 책을 읽은 빈도’가 읽기, 수학 및 과학 성취에 모두 정적인 영향을 미쳤다는 결과와 일맥상통한 시사점을 제시한다. 즉, 그간 범교과 학습에서 독서의 중요성이 언급되어 온 상황에서 교과 학습을 위한 독서의 영향력을 실증적으로 확인한 것이다. 따라서 개별학교나 시도교육청에서는 국어 과목에만 한정하지 않고 범교과 학습을 위해 독서를 연계하는 효과적인 독서 교육이 지원될 필요가 있다.

이처럼 읽기관련 변수는 읽기 영역은 물론이고 다른 영역의 성취에도 영향을 미칠 수 있으므로, 여러 이유로 읽기에 대해 어려움이 있는 학생에 대해 가정, 학교 및 지역사회가 연계하여 적극적이고 체계적인 지원 방안을 모색하는 것이 필요하다. 그러나 PISA의 수학 및 과학 영역 평가에는 문장제 구조,

텍스트 자료 분석이 필요한 문항 등이 포함되어 있어 읽기관련 변수가 수학 및 과학 성취에도 정적인 영향을 미칠 수 있다는 점을 간과해서는 안 될 것이다.

셋째, 대표적인 학생배경 변수인 'ESCS'는 읽기와 수학 성취의 주요 예측변수로 도출되었으며, 특히 수학 성취에서는 'ESCS'외에도 '연간 가계 소득'과 사교육 경험과 관련된 '정규 수업 외 추가적으로 받는 교육(수학)'이 주요 예측변수로 선택되었는데, 이는 모두 가정의 경제적 수준과 관련 있는 변수이다. 다층모형 분석에서도 '연간 가계 소득'과 '방과 후 공부시간'이 수학 성취에 긍정적인 영향을 미치는 것으로 나타났다. 수학의 경우 '수포자'라는 말이 통용될 정도로 수학 학업에 어려움을 겪는 경우가 많고, 다른 과목에 비해 정규교육 외 사교육 등에 의존하는 경우가 많다(박현정, 2010). 따라서 청소년의 수학 성취 수준 향상을 위해서는 가정의 경제적 수준이나 부모의 지원 차이를 극복할 수 있도록 수학 교육과 지원이 필요한 학생에 대해 학교나 지역사회 차원의 다각적이고 체계적인 지원이 요구된다. 가정 환경이 취약한 집단 중 읽기 성취가 높은 학생들을 대상으로 분석한 정혜경과 조지민(2013)의 연구에서 방과 후 학교에 참여하는 것이 읽기 성취에 긍정적인 영향을 미친다고 보고된 것을 고려할 때, 지원 방안으로는 효과적이고 내실화된 방과 후 학교 운영을 적극 활용해 볼 수 있다.

넷째, 흥미롭게도 진로 또는 직업과 관련된 변수가 인지적 성취에서 주요 예측변수로 도출되었다. 구체적으로 '학생이 기대하는 직업 지위'와 '미래 직업 결정에 중요한 요소(친한 친구의 미래 계획)'는 읽기와 과학 성취에서 주요 예측변수로 도출되었으며, '직업에 대한 정보'는 읽기 성취, '최종 학력에 대한 기대(대학교 및 대학원)'는 수학 성취에 대한 주요 예측변수로 도출되었다. PISA 2018에 표집된 한국의 만 15세 청소년들은 자신의 진로나 직업에 대한 포부를 높이는 것이 중요한 시기이므로, 청소년 시기에 자신의 강점과 관심을 파악하고 진로나 직업에 대한 탐색을 통해 학업에 대한 열의와 성과를 높일 수 있는 방안을 제고하는 것이 필요하다. 한국에서는 혁신학교나 자유학기제 등 혁신적인 교육과정 변화를 통해 학생들의 강점과 진로를 탐색하기 위한 다양한 교육을 시도하고 있다. 이처럼 체계적이고 효과적인 진로활동 및 교육을 통해 청소년들이 성숙한 진로태도를 형성하고, 자신의 진로나 직업과 관련된 포부를 설정하도록 하여 이것이 학업에 대한 열의나 성취로 이어질 수 있도록 해야 할 것이다. 특히 다층모형 분석 결과에서 미래 직업 결정에 친한 친구의 미래 계획이 보다 중요하다고 생각할수록 읽기, 수학 및 과학 성취 수준이 낮은 것으로 나타났다는 점은 매우 흥미로운 결과이다. 청소년이 자신의 미래 직업이나 진로를 결정하는 데 있어 주체적인 고민과 선택이 아닌 타인에 의해 의존적으로 선택한다는 것이 인지적 성취에까지 부정적인 영향을 미칠 수 있음을 의미한다. 선행연구(박현주, 김봉환, 2006; 이명희, 김아영, 2008)에서도 청소년이 주체적으로 진로 등을 선택하기보다는 부모, 교사, 친구 등 타인에 의존하거나 영향을 받는 경향이 있다고 지적했다. 청소년기의 학업과 진로를 긴밀하게 연관된 개념으로 이해하고, 청소년이 진로나 직업을 탐색하고 선택할 때 보다 더 성숙한 태도와 주체적인 진로 결정이 요구된다.

다섯째, ICT 변수는 읽기, 수학 및 과학 성취에서 주요 예측변수로 나타났다. 'ICT 사용에 대한 지각된 자율성'은 세 영역에서 공통적으로 도출되었고, '수업에서 과목과 관련된 ICT 사용'과 '학교에서 일상적인 ICT 사용'은 읽기와 과학 성취 영역에서 주요 예측변수로 도출되었다. 이 외에도 읽기 성취의 경우 'ICT에 대한 흥미', '가정에서 사용 가능한 ICT', '학교에서 사용 가능한 ICT', 및 '학교 밖에서의 ICT 사용(학교 활동 목적)' 등 총 6개 ICT 변수가 주요 예측변수로 도출되었다. 그러나 다층모형 분석

에서 학교에서 일상적인 ICT 사용'은 읽기와 과학 성취와 부적인 관계가 있는 것으로 나타났다. 따라서 학업적인 부분에서 ICT를 활용하는 것은 다양한 학습 환경이 학습상황에서 전제될 때 학습의 효과성을 높일 수 있으므로(김세리, 남창우, 장선영, 2012), 보다 효과적인 ICT 활용 방안을 모색하는 것이 요구된다. 교수-학습 상황에서 ICT를 활용할 시 교사의 역량이나 학교의 상황에만 의존하기보다는 학교와 교육청에서는 전문가와 연계하여 교수자가 효율적으로 ICT를 사용할 수 있도록 가이드를 제공해 주는 것이 필요할 것으로 사료된다.

여섯째, 사회·환경적 변수인 '학교 소속감'은 청소년의 읽기, 수학 및 과학 성취에 부적인 영향을 미치는 것으로 나타났다. 본 연구와 유사하게 PISA 2012 자료를 활용한 선행연구(송미영, 임현정, 임해미, 박혜영, 구자옥, 2015)에서 남학생인 경우 학교 소속감이 읽기, 수학 및 과학 성취에 모두 부적인 영향을 미친다고 하였다. 그러나 연구대상과 분석 자료는 상이하지만, 초등학교를 대상으로 한 설경옥과 정승원(2013)의 연구에서는 학교 소속감과 학업성취도 간에 긍정적 관계가 있는 것으로 나타났다. 따라서 학교 소속감과 읽기, 수학 및 과학 성취와의 관계를 더 면밀히 검증할 필요가 있으며, 학교 소속감과 인지적 성취의 부정적 관계를 해석하기 위해 매개변수와 같이 다른 변수의 존재를 확인하는 등의 추가적인 연구가 필요해 보인다.

본 연구는 대표적인 다층구조인 PISA 2018 자료를 활용한 실제자료 분석을 통해 다층자료 구조를 반영하지 못하는 전통적인 랜덤포레스트 기법에 비해 혼합효과 랜덤포레스트 기법의 예측성도가 상대적으로 우수하다는 것을 확인하였다. 따라서 머신러닝 기법을 다층자료 구조를 가진 데이터를 활용할 시 이를 반영할 수 있는 혼합효과 랜덤포레스트를 적용을 고려해볼 필요가 있음을 제안하는 바이다. 또한 사회과학 분야에서는 예측력만큼 설명과 해석이 중요한데, 혼합효과 랜덤포레스트 기법도 여타 트리기반의 머신러닝 기법의 단점으로 지적되는 것과 같이 변수 중요도만으로는 주요 예측변수 영향력을 해석하는 데 한계가 있다. 이에 본 연구에서는 다층모형 분석을 추가적으로 실시하였으며, 머신러닝 기법과 다른 전통적인 모형을 보완하여 활용할 수 있는 하나의 방안을 제시하였다. 그러나 본 연구는 혼합효과 랜덤포레스트에 기반해 청소년의 읽기, 수학 및 과학 성취의 주요 예측변수를 탐색하여 다층모형 분석에 포함하였기 때문에 학생과 부모관련 변수만 투입되었다. 후속 연구에서는 본 연구에서 도출된 변수와 함께 다양한 교사 및 학교관련 변수의 영향력을 함께 검증하여 더욱 풍부한 결과를 도출할 수 있길 기대한다.

PISA 자료
분석에서
이런식으로
나타나는
결과
↓
어떻게
변화할지
것이야?

참고문헌

- 강상진(2016). **다층모형**. 서울: 학지사.
- 구자옥, 한정아, 김성숙(2015). PISA 2012 상위국 남녀학생의 과학 성취에 미치는 교육맥락변인 영향력 분석. **교육평가연구**, 28(5), 1381-1400.
- 김세리, 남창우, 장선영(2012). 학생의 EBS 및 컴퓨터 활용과 교사의 ICT 활용이 학습자의 학업성취도에 미치는 영향. **교육공학연구**, 28(4), 849-877.
- 김수혜(2018). 청소년의 사회심리적 자원과 학교풍토가 수학 소양성취에 미치는 영향: PISA 2015 한국 자료를 바탕으로. **아시아교육연구**, 19(3), 651-677.
- 김혜숙(2012). ICT 활용이 학업성취도에 미치는 영향: PISA 2009 한국 자료를 중심으로. **아시아교육연구**, 13(1), 1-22.
- 박현정(2010). 학생들의 사교육 참여와 수학 성취도 및 수학 수업이해도간 관계에 대한 종단적 분석. **교육평가연구**, 23(4), 887-907.
- 박현주, 김봉환(2006). 고등학생의 심리적 독립과 진로결정수준 및 진로준비행동의 관계. **한국심리학회지: 상담 및 심리치료**, 18(2), 419-439.
- 설경옥, 정승원(2013). 초등학생의 개인 및 부모환경요인과 학업성취와의 관계에서 학교소속감의 매개효과. **한국심리학회지: 학교**, 10(1), 41-58.
- 손윤희, 박현정, 박민호(2020). 랜덤 포레스트를 활용한 읽기소양 수준에 따른 집단 결정요인 분석: PISA 2018 자료를 중심으로. **아시아교육연구**, 21(1), 191-215.
- 송미영, 임현정, 임해미, 박혜영, 구자옥(2015). PISA 2012 결과에 나타난 성별에 따른 학업성취 영향요인 비교. **교육과학연구**, 46(4), 99-122.
- 윤정일, 이범수(2006). PISA 2003 학업성취도 결정요인 국제비교연구-한국, 일본, 미국, 독일, 핀란드를 중심으로. **교육행정학연구**, 24(2), 523-547.
- 이명희, 김아영(2008). 자기결정성이론에 근거한 한국형 기본 심리 욕구 척도 개발 및 타당화. **한국심리학회지: 사회 및 성격**, 22(4), 157-174.
- 이인화, 구남욱(2019). PISA 2015 읽기 영역 결과에 나타난 한국 학생들의 성취 특성 및 교육맥락변인의 영향력 분석 연구. **독서연구**, 50(-), 113-144.
- 이은하, 최은정(2015). 핵심역량과 국어과 교육과정의 연계 가능성 탐색-PISA 2009 읽기 영역을 중심으로. **아시아교육연구**, 16(2), 201-230.
- 이정민, 정혜원(2019). 다층구조방정식을 적용한 PISA 2015 과학성취와 과학 교수학습, 학업적 자아효능감의 관계 분석. **교과교육학연구**, 23(2), 96-107.

이향미, 정혜원(2019). 청소년의 독서활동, 자기효능감, 학업성취도의 종단적 관계. **교육학연구**, 57(3), 233-257.

정혜경, 조지민(2013). PISA 2009 읽기와 사회경제적 배경 변인에 기반한 역경 극복 학생 특성 분석. **교육과정평가연구**, 16(1), 161-184.

정혜원, 박소영, 김정인, 김아름(2021). 청소년의 읽기 소양과 삶의 만족도의 영향변인 탐색: PISA 2018 한국 핀란드 국제비교. **교육과정평가연구**, 24(1), 123-152.

조성민, 구남욱, 김현정, 이소연, 이인화(2019). **OECD 국제 학업성취도 평가 연구: PISA 2018 결과 보고서**. 진천: 한국교육과정평가원.

최숙기, 박기범(2012). 읽기 성취에 대한 남학생과 여학생 간 성별 차이 분석 및 교육적 방안 탐색 - PISA 2009 읽기 관련 변인을 중심으로 -. **교육과정평가연구**, 15(1), 239-266.

한국교육과정평가원(2019). **PISA 2018 결과 발표 별첨자료**. 진천: 한국교육과정평가원.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.



Hajjem, A., Bellavance, F., & Larocque, D. (2014). Mixed-effects random forest for clustered data. *Journal of Statistical Computation and Simulation*, 84(6), 1313-1328.

Muthén, L. K., & Muthén, B. O. (1998-2020). *Mplus User's Guide*(8th ed.). Los Angeles, CA: Muthén & Muthén.

OECD(2017). *PISA 2015 technical report*. OECD Publishing.

OECD(2020). *PISA 2018 technical report. Chapter 9. Scaling PISA data*. OECD Publishing.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park: Sage Publications, Inc.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), 1-67.

· 논문접수 : 2022.01.05. / 수정본접수 : 2022.01.28. / 게재승인 : 2022.02.09.

ABSTRACT

**Exploring Key Predictors Affecting Adolescents' Reading,
Mathematics, and Science Achievement**

Hyewon Chung

Professor, Chungnam National University

Soyoung Park

Ph.D. candidate, Chungnam National University

The purpose of this study was to explore the key predictors that explain the reading, mathematics, and science achievement of Korean adolescents. To do this, the mixed-effect random forest (MERF) (Hajjem, Bellavance, & Larocque, 2012) considering multilevel data structure and multilevel modeling were applied to PISA 2018 data. The main results are as follows. First, MERF approach revealed that reading strategies, the perception of competence in reading, enjoyment of reading, index of PISA tests, and students' perceived ICT autonomy were commonly selected across three the reading, mathematics, and science achievement. Interestingly, additional instruction in mathematics, learning time after school, and annual household income were explored as key predictors in mathematics achievement only. Second, the multilevel analysis revealed that student-, and parent-related factors were statistically related to Korean adolescents' reading, mathematics, and science achievement. Based on the findings, implications for improving Korean adolescents' reading, mathematics, and science achievement were discussed.

Key Words: *Reading Achievement, Mathematics Achievement, Science Achievement, PISA 2018, Mixed-Effect Random Forest (MERF), Multilevel Modeling*

