

II. 연구 계획

1. 연구의 목적 및 필요성

- 코로나-19 시기의 디지털 기술은 우리 학생들이 원격 또는 하이브리드 방식으로 학습을 지속할 수 있도록 하여 교육 기회를 보장하는 데 기반이 되었음. 이후 디지털 기술과 활용이 국가사회 전반에 내재화되어 감에 따라 디지털 전환기를 살아가는 우리 학생들은 무엇을 배워야 하고 어떻게 학습해야 하는지 그리고 학교를 포함한 교육 시스템은 이를 어떻게 지원해야 하는지에 대한 논의가 활발히 이루어지고 있음. 이는 디지털 기술이 우리 학생들에게 교육 기회를 균등하게 보장하는 차원을 넘어, 의미 있는 교육적 경험을 통해 성과로 이어져야 한다는 사회적 요구와도 맞물리는 지점임.
- 2022 개정 교육과정은 미래 사회가 요구하는 역량 함양이 가능한 교육과정, 디지털·AI 교육환경에 맞는 교수·학습 및 평가 체제 구축을 개정의 중점 사항으로 제시하였음. 그리고 인간상 설정 시 자기주도성, 창의와 혁신, 포용성과 시민성을 고려하였는데(교육부, 2021), 이는 디지털 기술이 학생 중심 맞춤형 학습을 실현하는 데 기반이 되어 학교 교육의 책무성을 증진하는 데 혁신적으로 활용되어야 해야 함을 시사함.
- 교육부 역시, ‘디지털 기반 교육혁신 방안(2023. 2)’ 과 ‘디지털 기반 교육혁신 역량강화 지원방안(2024. 4)’ 등 정책 방안을 마련한 바 있음. 그리고 2025년부터 AI 디지털교과서를 학교에 도입하여 개별화 및 초개별화 학습을 실현하고자 하며 교원의 AI 교수·학습 역량을 강화하는 등 다양한 디지털 기반 교육혁신 정책을 추진 중임. 이러한 정책적 노력은 디지털 기술을 단지 수업의 보조 수단으로 이용하는 것에 그치는 아니라, 학습하는 방식을 바꾸는, 즉, ‘교육의 디지털 전환’ 을 위한 것임.
- 이러한 전환 노력은 디지털 교육의 성과에 대한 체계적인 분석에 터할 때 교육 현장에 효과적으로 안착될 수 있음. 특히, 디지털 교육은 ‘교육의 디

지텔 전환’을 추진한다는 점에서 다양한 기대와 우려가 공존함. 이에 디지털 교육의 실제와 성과를 투입-과정-산출의 전반에 걸쳐 실증 분석하고 그 결과를 사회적으로 공유함으로써 디지털 교육에 대한 이해를 높여 교육 주체들의 전환 의지를 제고하고 디지털 교육이 교육 현장에 효과적으로 안착할 수 있도록 지원할 필요가 매우 높음.

- (디지털 교육에 대한 기대) OECD(2023a)에서는 디지털 교육이 가져올 효과에 대해 5가지로 제시함. ① 학생 개인의 학습 수준과 필요에 맞는 맞춤형 교육 가능, ② 장애학생, 특수학생을 포함한 모든 학생의 교육 접근성을 제고하여 포용성과 형평성 증진, ③ 교사의 효과적인 수업 지원을 통한 교수의 질 향상, ④ 데이터 기반 의사결정과 행정을 통한 업무 효율성 증진, ⑤ 대규모 데이터 분석을 통한 연구 및 혁신 강화임. 즉, 디지털 교육은 교사의 수업과 학생들의 학습을 혁신함으로써 개별화, 초개인화 교육을 실현하는 데 기여할 수 있을 것으로 기대됨.
- (디지털 교육에 대한 우려) 이에 반하여 디지털 교육이 학생들의 문해력 및 사회정서 역량의 발달을 저해하며, 디지털 기기 및 인프라에 대한 접근성 차이나 학생과 교사의 디지털 활용 역량 차이는 오히려 학습 격차를 심화시킬 수 있다는 우려도 존재함(문희원, 정동욱, 2025).
- 이상의 디지털 교육에 대한 기대와 우려는 선행연구에서 일부 검증된 바 있으며, 국가 및 지역 수준의 패널 연구에 디지털 교육 현황 관련 문항이 포함되어 제한적이거나 조사되고 있음. 그러나 ‘교육의 디지털 전환’이라는 시대적 과제의 중요성에 비추어 볼 때 단편적인 접근보다는 디지털 교육 중심으로 설계되어 학생, 교사, 학교의 성장과 발달에 대한 데이터를 수집할 수 있는 조사 체계를 만들어 디지털 교육 정책의 성과를 체계적으로 분석하여 디지털 교육에 대한 기대의 실현과 우려의 개선을 도모해야 할 필요가 그 어느 때보다 높음.
- 관련하여 지난 2024년 수행된 「한국 디지털 교육 중단연구 설계 및 실행방안 연구」(박현정 외, 2025)는 디지털 교육의 성과를 체계적으로 진단하는데 출발점이 되는 기초연구에 해당함. 이에, 이 연구에 기반한 조사가 실제 수행되어 투입-과정-산출의 전 과정에 걸친 디지털 교육의 실제 분석과 효

과성 검증이 체계적으로 이루어진다면, 학생, 학부모, 교원 등 다양한 교육 주체의 디지털 교육에 대한 이해를 돕고 이를 현장에 안착시키는 데 유의미한 시사점을 제공할 수 있을 것임.

○ 이상의 필요성에 터하여 본 연구의 목적을 3가지로 제시하면 다음과 같음.

- 첫째, 디지털 교육 정책의 성과 진단을 위한 조사 설계 방안을 마련하여 횡단 조사(2025년)와 3개년도 종단조사를 신뢰롭고 타당하게 추진할 필요가 있음.
- 디지털 교육의 효과를 과학적으로 분석하고 데이터에 기반한 정책 방향을 설정하기 위해서는 교육 주체의 디지털 역량과 학교 단위의 전환 수준 등을 정기적으로 모니터링할 수 있는 조사 체계가 마련되어야 함.
- 박현정 외(2025)의 연구는 이를 위한 기초연구로서 투입-과정-산출의 연구모형에 근거하여 학교, 교사, 학생, 학부모용 등 4종의 조사 도구를 개발하였고, 62개 초등학교 4,173명의 4학년 학생과 1,665명의 초등학교 교사, 33개 중학교 4,542명의 1학년 학생과 1,022명의 중학교 교사를 표본 크기로 제안한 바 있음.
- 이에 본 연구를 통해 실제 조사를 수행하기 위한 표본설계 및 표집 등이 현행화되어야 함. 특히 본 연구가 2025년 횡단 조사와 함께 3개년도 종단 조사(2025-2027년)를 모두 수행하여 이루어진다는 점에서 패널 선정 및 관리 방안 등을 효율적으로 모색할 필요가 있음. 또한, 조사가 종료된 이후에도 디지털 교육의 성과를 지속적으로 진단해 나갈 수 있도록 중·장기적인 방안을 모색하는 미래지향적인 접근도 필요함.
- 또한, 대규모 조사연구를 통해 수집된 정량적·정성적 데이터는 디지털 교육 정책의 수립과 실행을 뒷받침하는 데 중요한 역할을 하므로, 지속적인 확장과 고도화가 필수적임. 이에 박현정 외(2025)의 연구를 학생의 성장뿐만 아니라 교사의 수업 변화 등 다양한 관점에서 검토하여 디지털 교육의 성과 진단을 위한 평가 틀로 개선할 필요가 있을 것임.
- 이상의 설계 과정을 거쳐 학교 교육 현장과 협력하여 조사 체계를 마련하고, 실제 조사 수행을 통해 신뢰롭고 타당한 자료를 수집하고 관리해

야 나가야 함.

- 둘째, 수집된 자료를 활용하여 이루어질 디지털 교육의 효과 분석은 교육 시스템 전반의 구조적 전환과 연계되어야 함. 특히, 학생의 관점에서는 학습자 주도성, 디지털 리터러시 등 역량을 중심으로 한 효과 분석뿐만 아니라 수학과 영어 교과의 성취도를 주요 성과로 설정하여 단기와 중장기에 걸친 디지털 교육의 효과를 실증적으로 분석할 필요가 있음.
 - 디지털 교육의 실질적인 효과는 개별 수업이나 기술 도입 여부만으로 설명되기 어려우며, 교육과정 및 평가의 운영 방식, 학습 환경, 상호작용 구조 등 전반적인 시스템의 변화와 밀접하게 연결됨. 이는 디지털 교육의 효과를 보다 신중하고 다각적인 관점에서 검토되어야 할 필요가 있음을 보여줌. 따라서 효과 분석은 구조적 전환의 맥락을 함께 고려해야 하며, 기술 중심의 단편적 분석을 넘어 교육 전반의 질적 변화를 포착할 수 있는 분석 틀의 설계가 요구됨.
 - 이상의 분석 틀에 근거하여 ① 투입-과정-산출의 전반에 걸쳐 디지털 교육의 실재를 파악할 수 있는 기초 분석이 충실히 이루어짐과 함께, 디지털 교육이 가져올 학생과 학교(교사)의 성장과 변화를 다각도로 분석하고 이를 증거로 삼아 교육 정책의 수립과 실행의 개선을 가져오는 순환적인 구조를 지원할 필요가 있을 것임.
 - 그리고 ② 디지털 교육 정책 중에서도 2025년에 도입된 AIDT(AI Digital Textbook)에 초점을 맞추어 AIDT의 효과성 분석을 위한 실험연구, 조사자료 분석 등을 엄밀하게 실시하여 AIDT의 교육적 활용과 현장 안착을 데이터 기반으로 지원할 필요가 있음. 이처럼 수학과 영어 교과성취에 미치는 영향을 객관적으로 분석한 결과 및 디지털 리터러시에 대한 분석 결과는 학생들이 디지털 교육을 통해 무엇을 학습하며 어떻게 학습하였는지를 경험적으로 이해하는 데 중요한 자료가 될 수 있음.
- 셋째, 수집된 자료는 국가 수준 뿐만 아니라 학교 단위에서 분석되어 디지털 교육의 효과를 극대화하기 위한 교육 주체들의 디지털 교육에 대한 이해와 역량 강화를 지원할 필요가 있음.
 - 디지털 교육 정책을 실천해 나가는 학교 단위에서 디지털 교육에 대한

진단적 정보를 다양하게 파악할 수 있도록 지원하는 것은 디지털 교육에 대한 우려를 개선하고 기대를 실현하는 데 의미 있는 기제가 됨. 이에 학교 단위에서 구성원들이 디지털 교육의 실재를 이해하고 분석 결과를 형성적으로 피드백하여 디지털 교육의 개선을 모색하는 데 도움이 될 수 있는 진단 자료를 본 연구를 통해 제공할 필요가 있음.

- 또한, 디지털 교육을 설계하고 운영하는 교사는 AI 및 디지털 기술을 비판적이고 창의적으로 활용할 수 있는 능력과 윤리적 태도를 갖추어야 하며, 이를 위한 실천적 역량 기반의 연수와 지원이 요구됨. 역량 강화는 단순한 기능 습득을 넘어서 디지털 환경에 적응하고 변화를 주도할 수 있는 수용성과 전문성을 개발하는 방향으로 이루어져야 함. 이에 본 연구를 통해 연수의 방향과 내용에 있어 유의미한 함의를 제공할 필요가 있음.

2. 연구의 기대효과

- 본 연구를 통해 마련될 디지털 교육 중단조사 체계는 디지털 교육에 초점을 맞추어 투입-과정-성과를 체계적으로 진단할 수 있는 지표로 구성되었다는 점에서 기존 국가와 지역 수준의 중단연구들과 차별성이 있음. 즉, 본 연구를 통해 구축될 3개년도 디지털 교육 중단자료는 학생, 교사, 학부모들이 경험(인식)하는 디지털 교육의 실재와 성과를 면밀히 들여다 볼 수 있는 자료로서 국가 수준에서의 증거 기반 디지털 교육 정책 수립 및 실행을 지원할 수 있다는 점에서 그 가치가 매우 높음.
- 본 연구에서 수집될 중단조사 자료는 국가 수준에서 디지털 교육의 실재를 파악하고 교육적 성과를 심층 분석하기 위한 자료이므로 자료의 질 관리가 매우 중요할 것임. 또한, 디지털 교육의 성과를 주관적 인식에 터하여 측정하는 것과 함께 수학과 영어 성취도를 객관적으로 측정하여 종합적으로 분석한다는 점에서 본 연구를 통해 구축될 조사자료 자체가 매우 중요한 산출물이라 할 수 있음.
- 구축된 디지털 교육 중단연구 조사자료를 활용하여 디지털 교육의 투입-과정-성과의 전반에 걸쳐 학생과 교사의 관점에서 그 실재와 성장을 이해할

수 있는 다양한 실증 분석이 이루어짐으로써 향후 국가와 지역 수준에서 디지털 교육 정책 수립-실행-성과 피드백을 데이터 기반으로 지원할 수 있을 것으로 기대됨. 2025년에 한하나 AIDT를 중심으로 실험 연구를 실시하여 그 효과를 통해 검증한 결과 역시 AIDT 도입과 확산의 과정에서 경험 과학적으로 관련 정책들을 수립하는 데 중요한 기초자료가 될 것임.

- 나아가, 본 연구를 기반으로 하여 디지털 교육의 장기적 성과를 분석할 수 있는 조사 방안이 마련될 수 있다는 점에서 의미가 있음. 예컨대 AIDT 로그데이터 연계나 맞춤형 국가 수준 학업성취도 자율 평가나 시·도 연합 모의고사 등과의 연계 등을 통한 장기적 디지털 교육의 성과 진단 방법 모색함으로써 본 조사가 지속 가능한 방식으로 안착되어 갈 수 있을 것임.

- 이상의 연구 결과는 교육의 디지털 전환을 마주한 교사, 학생과 학부모의 디지털 교육의 실제와 성과에 대한 이해 전반을 높임으로써 향후 우리나라에서 디지털 교육정책을 추진함에 있어 의미 있는 자료로 활용될 것임.

3. 연구의 내용 및 범위

3-1. [과제 1] 디지털 교육 정책 성과 진단

가. 종단연구 대상 선정 및 표본설계

1) 연구 대상 정의

- 본 연구는 전국 초등학교 4학년과 중학교 1학년 학생을 중심으로 복수 코호트를 설계, 3개년 종단연구를 통해 디지털 교육 정책의 성과를 진단하고, 패널의 다년간 변화를 추적·분석하는 것을 목적으로 함. 또한, 디지털 교육 정책의 성과에 대한 체계적이고 종합적인 분석을 위해서는 교사 개인의 수업 혁신 및 교사 디지털 역량의 변화 또한 지속적으로 추적 및 분석할 필요성을 지님. 따라서, 본 연구를 위해 학생뿐만 아니라 교사 또한 패널의 성격을 지니는 조사 설계가 필요함.

- 본 연구의 세부 조사 대상은 다음과 같음.

<표 1> 연구 조사 대상 및 정의

조사 대상	조사 대상 정의
학생	표집 학교의 초4, 중1 학생
학부모	표집 학생의 보호자 1인
교사	표집 학교의 수업 담당 교사 전체 (교장·감, 상담, 사서, 보건, 영양교사 제외)
학교	표집 학교의 대표 교사 1인

2) 표본 설계(안)

- 디지털 교육 정책의 투입이 상당 부분 학교 단위로 이루어진다는 점과 본 조사의 경우 학생 패널과 교사 패널이 상호 유기적으로 연계되어야 한다는 특성을 지니고 있다는 점을 고려하여 본 연구에서는 학교 단위 표집을 실시하고자 함.
- 표본의 대표성 확보를 위해 층화표집을 실시하고자 하며, 층화 요인으로는 시도, 지역규모, 학교규모를 고려할 수 있음. 시도는 디지털 교육 정책 반영의 차이를 고려하여 가장 기본적으로 고려할 수 있는 요인이며, 디지털 교육 여건의 격차를 반영할 수 있는 지역규모(대도시/중소도시/읍·면·특수)와 학교규모(대규모/중규모/소규모) 또한 층화 요인으로 함께 고려할 수 있음.
- 전국 학교(또는 학생) 수 분포를 기준으로 각 층화에 따른 비례 할당을 통해 표본 학교 수를 우선 배분한 후, 각 층 내에서 단순무작위추출(Simple Random Sampling)을 통해 1차적으로 학교 표집을 실시함. 표집된 학교에 대해서는 해당 학교의 학생 전수 또는 학급 단위 무작위추출을 통해 일부 학급을 조사 대상으로 선정할 예정임.
- 모집단¹⁾
- 전국 초등학교 6,183개교, 초4 기준 학생 422,333명
 - 전국 중학교 3,272개교, 중1 기준 학생 455,387명
- ※ 본교만 포함, 폐교는 제외

1) 출처: 2024년 유·초·중등 교육통계 학교별 주요 현황 자료 (2024.4.1.기준)

- **(학교 수 기준 표본설계 방안, 1안)** 모집단 대표성을 확보하기 위한 기본 방안으로 전국 학교 수의 약 1%를 표본으로 추출하고, 해당 학교의 학생 및 교사 전수를 조사하는 방식을 고려할 수 있음. 이 경우, 전국 초등학교 약 62개교, 중학교 약 33개교로 총 95개교를 표본으로 설정할 수 있음.
- (장점) 이러한 설계의 주요 장점은 표집된 학교의 모든 수업 담당교사를 포함함으로써 학생 패널과 더불어 교사 패널 구성이 가능하다는 점임. 또한, 학교 단위에서 디지털 교육 정책의 성과를 종합적으로 분석할 수 있고, 학생-교사 간 패널 연계를 통한 종단적 분석이 가능하다는 것에 큰 강점이 있음.
- (한계점1) 다만 학교 수 기준 1% 표집 시 전체 학교 수가 95개교로 제한되기 때문에 17개 시도를 층화로 활용할 경우 일부 시도의 표본 배정에 어려움이 있음(상대적으로 학교 수가 적은 울산/세종/제주의 경우 표본에 포함되기 어려움). 이로 인해 시도별 대표성 확보에는 제약이 따르며, 시도를 층화 요인으로 설정하기보다는 지역규모 또는 학교규모를 층화 요인으로 활용하는 대안적 접근이 필요함.
- (한계점2) 표집 학교의 규모에 따라 전체 패널 수 및 지역별 분포에 상당한 편차가 발생할 수 있음. 특히, 학급 수 편차가 큰 중학교의 경우 학급 수가 많은 대규모 학교 표집 시 조사 대상이 과도하게 증가할 수 있으며, 이로 인해 학교와 교사의 조사 협조도 저하가 우려됨. 또한, 학교 조사의 특성상 학교별 상황에 따라 조사 협조율(예: 협력 교사의 협조도, 학부모 동의율 등)이 크게 달라지는 경향이 있다는 점을 고려하면, 전체 표본 학교 수가 적을 경우 특정 학교의 협조도에 따라 표본 구성 전반에 큰 영향을 미칠 수 있음.
- **(학생 수 기준 표본설계 방안, 2안)** 본 표본설계는 한 학교당 조사 학급 수를 2개로 고정하고 전국 학생 수 분포를 기준으로 표본학교를 할당함으로써 전체 조사 학교 수를 늘리는 방안임. 이는 기존의 국내 교육 종단연구나 학교 기반 조사에서 자주 활용되는 방식으로 학교 수 비율에 기반한 표본 설계보다 학생 모집단을 더욱 정밀하게 반영할 수 있다는 점에서 장점을 지님.
- 본 표본 설계에서는 층화요인(예: 시도, 지역규모 등)에 따라 전국 학생 모집

단 분포를 파악한 후 모집단 비율에 따라 층화별 표본 학생 수를 우선 확정함. 이후, 학급당 평균 학생 수를 고려하여 층화별 필요 표본 학생 수를 조사하기 위한 표본 학교 수를 산출함.

- 예를 들어, 초등학교 학급당 평균 학생 수가 20명이고, 학교당 2개 학급을 조사할 경우, 1개 초등학교당 평균 40명의 학생 조사가 가능하므로, 100명을 조사하기 위해 대략 3개교를 표본으로 배정하는 방식임. 이 경우 전국 초등학교 약 105개교, 중학교 약 92개교로 총 197개교에 대한 조사가 필요함.
- (장점) 이 경우 모집단 대표성을 보다 정밀하게 확보할 수 있으며, 조사 효율성 측면에서도 강점을 지님.
- (한계점) 학교 수가 약 2배 이상 증가함에 따라 해당 학교의 수업 담당 교사 전체를 대상으로 패널로 구성하는 것은 현실적으로 어려움이 있음. 앞서 언급했듯이, 본 연구의 경우 연구모형에 따라 교사 또한 패널의 성격을 지닐 필요성이 있으나, 이와 같은 방식에서는 교사 조사를 해당 학년 수업 담당 교사로 한정하고, 횡단조사 방식으로 제한될 수밖에 없다는 제약이 존재함.
- 학교 수를 기준으로 한 표본 설계 방안과 학생 수를 기준으로 한 표본설계 방안의 장단점은 <표 2>와 같이 요약할 수 있음.

<표 2> 표본설계 방안 비교 요약

구분	1안: 학교 수 기준	2안: 학생 수 기준
표본 규모	약 95개교 (초: 62개교, 중: 33개교)	약 197개교 (초: 105개교, 중: 92개교)
학생 표집	해당 학교 전체	학교당 2개 학급
교사 표집	전체 수업 교사 (패널 설계 가능)	해당 학년 수업 교사 (횡단조사 중심)
장점	교사 패널 구성 가능 학생-교사 연계 분석 가능	모집단 대표성 확보 학교 부담감 경감
단점	시도별 층화 어려움 조사 협조도 저하 우려	교사 패널 확보 어려움

- 다만, 본 연구의 목적 및 연구모형, 모집단 대표성과 조사 효율성을 종합적으로 고려했을 때, 학교 수 기준 비례 할당 표집을 유지하되, 학교당 조사

학급 수를 최대 3개로 제한하고, 전체 학교 수를 기존 1안 대비 학교급별로 약 20~40% 과표집하는 방안(전체 학교 수 기준 120개~130개교)을 제안하고자 함.

- 이 경우 학교당 조사 학급 수를 최대 3개로 제한함으로써 과도한 조사 부담 및 학교별 표본 구성의 편차가 커지는 문제를 완화할 수 있음. 또한, 시도 단위 층화 적용이 가능하다는 점에서 대표성 확보를 강화하면서도, 전체 교사 조사를 통해 교사 패널 구성을 확보할 수 있다는 장점이 있음.

3) 연구 참여 대상 관리 방안 제시

- 3개년 동안 연속 운영 가능한 통합 홈페이지를 제작하여 조사 안내 및 패널 소통 창구로 활용함으로써 조사에 대한 신뢰성 제고함.
 - 조사 취지, 내용 및 일정, 주요 질의응답 등 세부적인 안내 사항 탑재.
- 학교 담당자용 조사 안내문 및 매뉴얼 제공을 통해 협조 효율성 및 신뢰도를 제고함.
 - 매뉴얼 제작: 학교 내 본 조사 대표 협력교사와 학교 및 교사 설문 응답자를 위해 조사 전체 협력 방법 및 조사 진행 매뉴얼을 배포하여, 조사 협조 및 입력 과정에서의 혼란을 줄이고자 함.
 - 안내문 제작: 학교 관계자 외에 학생 및 학부모를 위한 조사 안내문을 별도로 제작하여, 조사의 목적과 응답 방법에 대해 설명하고자 함.
- 조사 기간 동안 협력 교원의 전 업무 과정을 실시간으로 지원 가능한 통합 콜센터를 운영, 학교 현장에서의 애로사항을 즉시 해결함으로써 조사 완결성을 제고함.
 - 본 조사 응답 방법 안내, 학교 관계자, 교사, 학생 및 학부모의 조사 관련 문의사항 응대, 입력 오류 등 발생 시 원격 지원 포함(학교 단위).
 - 콜센터 직원은 학교 및 학부모 조사 경험이 있는 전화 면접원으로 우선 선발.
 - 콜센터 번호는 수신자 부담 080 번호로 일원화하며, 연결된 전화는 콜센터 직원에게 순차적으로 배치 (단일 번호라도 통화 중 처리되지 않음).
 - 콜센터 직원 대상 집체 교육을 실시하여 조사 특성에 대해 충분히 이해하

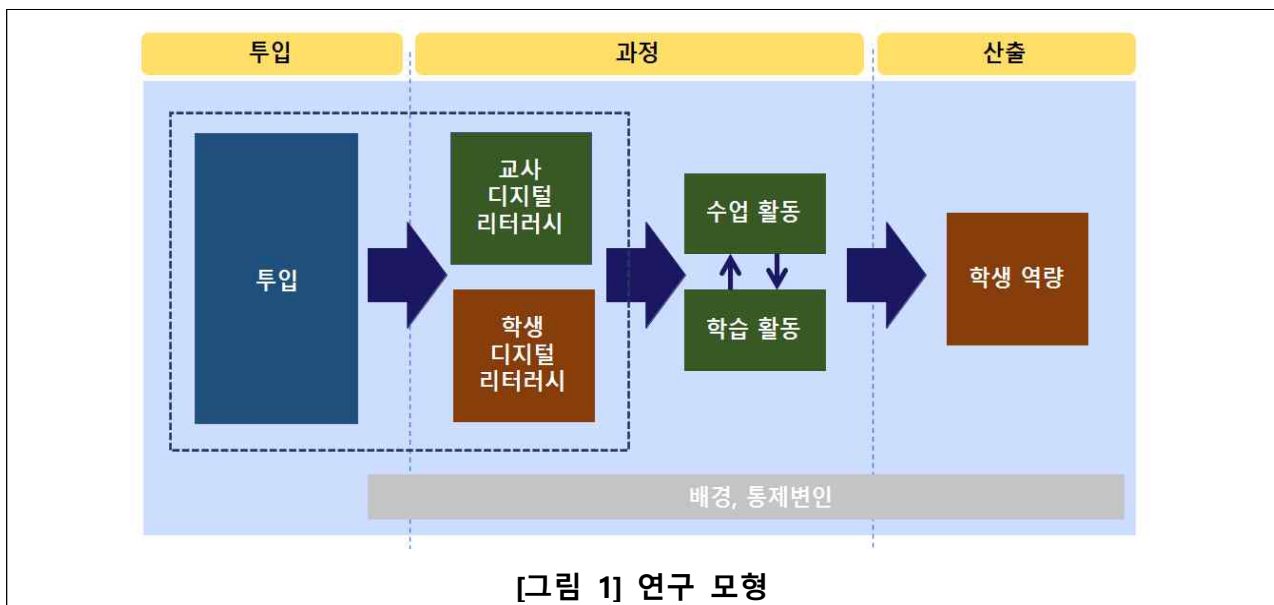
도록 하며, 조사 응답 체계 및 응대 요령, 원격 지원 방법 등을 숙지.

- 패널 대상 경품 추첨 이벤트 진행을 통한 조사 참여 유도.
- 조사 완료 후 학교에서 흥미를 가질 만한 주요 결과를 담은 조사 소식지를 제작하여 배포함으로써 조사에 대한 흥미를 높이고, 연속적인 참여를 유도하고자 함.

나. 디지털교육 성과 진단을 위한 평가틀 및 조사설계 방안 제시

1) ‘투입 - 과정 - 산출’ 모형 기반의 기 개발된* 연구 모형을 참조하여 디지털 교육 성과 진단을 위한 평가 틀 제시

- 디지털 교육의 성과를 학생의 성장이라는 측면에서 제시한 기존 평가틀에 대한 검토 및 수정 보완.



※ 한국 디지털 교육 종단 연구 모형 설계 및 실행방안(KERIS RR 2025-01) 연구 참조
(KERIS 홈페이지 - 지식정보 - KERIS 보고서 - 연구보고서 참조)

- 이외에 디지털 교육의 정책적 투입이 교사의 디지털 리터러시 향상에 미치는 성과나 교사의 수업(수업 시간에서의 교사와 학생의 교수-학습 활동 양태) 변화에 미치는 성과 등에 대한 평가 틀 검토.

2) 장기적 디지털 교육 성과 진단을 위한 조사 설계 방안 제시

- [과제 1]에서 (현재 예정된 3년 조사 이후에도) 디지털 교육의 장기적 성과를 분석할 수 있는 조사방안 제안.
 - AIDT 로그데이터 연계나 맞춤형 국가수준 학업성취도 자율평가나 시도연합 모의고사 등과의 연계 등을 통한 장기적 디지털 교육의 성과 진단 방법 모색.
- [과제 2](효과성 연구)에 참여하는 실험 학교를 [과제 1]에서 과대표집 하거나 [과제 2]의 일반 학교를 [과제 1]의 표집학교에서 모집하도록 하여 AIDT 활용의 장기적 성과를 분석할 수 있는 조사방안 제안.

다. 성취도 검사도구 개발

1) 컴퓨터 기반 성취도 검사 도구 개발

<표 3> 성취도 검사 도구 (총 12종) 개발 내용

구분	내용
대상 학년	초등학교 3,4학년, 중학교 1학년
과목	영어, 수학 (※ AIDT 사용 중인 과목 중심)
문항 수	초등학교: 20문항, 중학교: 25문항
검사 종류	중간검사, 사후검사
검사 시기	중간검사: 7월, 사후검사: 11월
검사 범위	중간검사: 해당 학년 1학기 범위 사후검사: 해당 학년 전체 범위
공통 문항	검사 동등화를 위해 사후검사 문항 중 초등학교: 4문항, 중학교: 5문항 공통
출제 및 검토 위원	국가 수준 표준화 검사 출제 경험이 있는 공동연구원이 팀장 역할 수행 출제/검토진: 출제 경험 있는 3인 출제위원 + 1인 검토위원으로 구성

- (학업성취도 검사 개발) 초등학교 3학년과 4학년 및 중학교 1학년 세 개 학년에 대하여 영어와 수학 두 개 교과에 평가 도구를 1학기 말, 2학기 말 기준으로 2종씩 개발하고자 함(총 12개 종의 성취도 검사 도구 개발).
- 검사 도구 개발을 위해 2022 교육과정 내용을 분석하여 평가 틀 구성
- 예비조사 설계 및 분석: 초3, 초4 각각 150명 정도, 중1 200명 정도 조사
- 컴퓨터 기반 검사(Computer Based Test, CBT)로 구축
- (검사 대상 및 실시 과목) 초등학교 3학년과 4학년 및 중학교 1학년이며, 과목은 AIDT를 사용하고 있는 영어와 수학임.
- (문항 수) 학교급에 따른 차이를 고려하여 초등학교는 20문항, 중학교는 25문항으로 설정함.
- (검사 종류 및 시기) 중간검사와 사후검사 2종을 개발하며, 중간검사는 1학기 말인 7월에, 사후검사는 학년 말인 11월에 실시하고자 함.
- (검사 범위) 중간검사의 경우 해당 학년 1학기, 사후검사는 해당 학년 1학기 및 2학기 전체로 설정함.
- (검사 동등화) 중간검사와 사후검사 간 검사 동등화(test equating)를 수행하기 위하여 영어와 수학 모두 초등학교 3·4학년은 사후검사 문항 중 4문항, 중학교 1학년은 사후검사 문항 중 5문항을 공통 문항으로 설정함.
- (연구진 구성) 영어와 수학 모두 국가 수준 표준화 검사의 출제 기획위원 경험을 지닌 공동 연구원이 팀장 역할을 수행하며, 출제 및 검토진은 마찬가지로 국가 수준 표준화 검사의 출제 및 검토 경험이 있는 3인의 출제위원과 1인의 검토위원으로 구성하고자 함.

○ 다음과 같은 절차에 따라 검사 개발.

- 교육과정 및 성취기준을 분석하여 검사 개발 청사진 작성(<표 4> 참조).

<표 4> 초등학교 4학년 수학 성취기준과 출제 계획(예시)

영역	성취기준	학년-학기	문항수	출제자
수와 연산	[1] 다섯 자리 이상의 수 [4수01-01] 10000 이상의 큰 수에 대한 자릿값과 위치적 기수법을 이해하고, 수를 읽고 쓸 수 있다. [4수01-02] 다섯 자리 이상의 수의 범위에서 수의 계열을 이해하고 수의 크기를 비교할 수 있다.	4-1	3	B
	[3] 곱셈 [4수01-05] 곱하는 수가 한 자리 수 또는 두 자리 수인 곱셈의 계산 원리를 이해하고 그 계산을 할 수 있다. [4수01-06] 곱하는 수가 한 자리 수 또는 두 자리 수인 곱셈에서 계산 결과를 어림할 수 있다.	4-1	3	C
	[4] 나눗셈 [4수01-09] 나누는 수가 두 자리 수인 나눗셈의 계산 원리를 이해하고 그 계산을 할 수 있다.	4-1	3	C
도형	[2] 평면도형의 이동 [4수02-04] 구체물이나 평면도형의 밀기, 뒤집기, 돌리기 활동을 통하여 그 변화를 이해한다. [4수02-05] 평면도형의 이동을 이용하여 규칙적인 무늬를 꾸밀 수 있다.	4-1	3	B
하락				

- 문항 초안 개발: 최종 문항 수의 약 1.5배 분량으로 문항 초안을 구성.
- 예비 검사 실시: 학년당 150명을 대상으로 예비 검사를 실시.
- 문항 분석: <표 5> ~ <표 7>과 같이 고전검사이론 및 문항반응이론을 적용하여 문항의 특성을 분석 (검사 수준: 타당도·신뢰도; 문항 수준: 난이도·변별도·답지 반응).

<표 5> 초등학교 4학년 수학 예비검사 문항특성 분석 결과(예시)

문항번호	변별도	난이도	문항제거시 α	α
1	.31	.76	0.688	0.698
2	.44	.70	0.684	
3	.26	.34	0.693	
4	.37	.51	0.685	
5	.43	.67	0.686	
<하락>				

<표 6> 초등학교 4학년 수학 예비검사 답지 반응 분포(예시)

문항번호	정답	답지 응답 비율(%)					
		①	②	③	④	⑤	기타
1	①	76.0	3.3	7.7	1.5	11.0	0.6
2	④	2.7	7.1	9.5	69.7	10.1	0.9
3	③	6.8	37.1	33.5	14.8	7.7	0.0
4	⑤	22.3	13.1	5.0	8.0	51.0	0.6
5	②	15.7	67.4	7.4	5.0	3.3	1.2
<하락>							

*정답에 응답한 비율에 음영, 정답보다 높은 응답 비율을 갖는 항목에 진하게 표시함

<표 7> 초등학교 4학년 수학 예비검사 문항모수 추정치(예시)

문항번호		문항모수 및 단계모수 추정치					
		α_i	β_i	τ_1	τ_2	τ_3	τ_4
이분 문항	1	.563	-1.393	-	-	-	-
	2	.742	-.835	-	-	-	-
	3	.347	1.242	-	-	-	-
	4	.590	-.045	-	-	-	-
	5	.684	-.765	-	-	-	-
<하락>							

- 문항 수정 및 확정: 분석 결과를 바탕으로 문항을 수정·보완하고 최종 문항을 확정.
- 시스템 탑재: 확정된 최종 문항을 CBT 시스템에 탑재.
- 부속자료 개발: 문항카드([그림 2] 참조), 이원목적분류표(<표 8> 참조), 정답 기준표 등 각종 부속자료 및 검사 매뉴얼 작성.

< ○○과 문항 카드 예시 >

교과 및 학년	교과:	학년:	문항번호
영역	대영역:	(중영역: , 소영역:)	
내용			
성취기준			
평가기준			
문항			
답지 (정답)	① ② ③ ④ ⑤		
문항특성	난이도: 상□, 중□, 하□ 변별도: 상□, 중□, 하□		
소요시간			
배점			

[그림 2] 문항카드 예시

<표 8> 초등학교 4학년 수학 이원목적분류표(예시)

번호	내용영역	교과역량(하위요소)		정답	예상정답률 (%)	소요시간 (분)
1	수와 연산	(1)	추론(관찰과 추측)	④	70	1
2	수와 연산	(1)	정보처리(정보해석 및 활용)	④	60	2
3	수와 연산	(1)	추론(논리적 절차 수행)	③	70	2
4	수와 연산	(1)	추론(논리적 절차 수행)	⑤	70	3
5	수와 연산	(1)	창의·융합(수학 내적 연결)	⑤	70	2
6	수와 연산	(1)	문제해결(문제 이해 및 전략 탐색)	④	80	2
7	수와 연산	(1)	창의성(융통성)	②	80	1
8	수와 연산	(1)	추론(수학적 사실 분석)	③	50	3
9	도형	(1)	의사소통(자신의 생각 표현)	②	70	1
10	도형	(1)	문제해결(계획 실행 및 반성)	③	60	2

<하락>

2) 검사 동등화 및 수준 설정

○ 중간검사와 사후검사는 학생의 성장과 변화를 정확하게 측정할 수 있도록 검사 동등화를 수행할 필요가 있음.

- 중간검사와 사후검사의 점수는 서로 다른 척도에 위치하기 때문에 원점수 만으로는 점수 차이에 대해 적절한 해석이 불가능함.
- 객관적이고 타당한 비교를 위해 검사형들 간 난이도 차이를 통계적으로 조정하여 동일한 척도로 위치하게 하는 과정, 즉 검사 동등화를 수행할 필요가 있음(Kolen & Brennan, 2014).
- 문항반응이론을 적용하여 검사를 동등화하기 위해서는 각 검사형에 대한 문항모수와 능력모수를 추정한 후 동일한 척도에 위치시키는 척도 연계화(scale linking) 과정이 필요함.
- 척도 연계화 과정을 거쳐 동일 척도에 놓인 점수는 상호 호환적으로 사용 가능함.

○ 검사 동등화 설계 및 방법

- 검사 동등화를 위해서는 동등화 설계와 동등화 방법을 결정해야 함. 이 연구에서는 동등화 설계로는 공통 문항 설계, 동등화 방법으로는 척도 연계 방법을 적용하고자 함.
 - 동등화 설계는 공통 피험자 설계, 무작위 집단 설계, 공통 문항 설계 등이 가능함. 이 연구에서는 동일한 피험자가 반복측정되며, 피험자의 능력이 변화한다는 점, 중간검사와 사후검사 간 시점이 다르다는 점 등을 고려하여 공통 문항 설계 사용할 예정임.
 - 동등화 방법으로는 문항반응이론 기반의 척도 연계(scale linking) 방법 적용 예정임. 문항반응이론 기반의 척도 연계를 위해 문항 모수의 추정과 척도화가 필요하며, 동시 추정(concurrent [simultaneous] calibration), 문항 모수 별도 추정 및 척도 변환(scale transformation with separate item parameter estimation), 문항모수 고정 추정(fixed item parameter calibration) 등의 방법을 사용할 수 있음. 중간검사와 사후검사 간 시점이 다르다는 점을 고려하여 고정 추정 방법을 기본으로 하되 필요 시 척도 변환 방법을 병행하고자 함.

○ 수준 설정

- 사전검사로 실시되는 맞춤형 자율평가와 연계를 위해 맞춤형 자율평가와 동일한 수준의 준거를 사용하여 등급 산출 필요함.
- 수준설정 방법은 무엇을 고려하는가에 따라 크게 다섯 가지의 범주로 분류되지만 가장 널리 사용되는 방법은 북마크 방법(Mitzel, Lewis, Patz, & Green, 2001)과 수정된 앵고프 방법(Angoff, 1971)이 있음.
 - 수정된 앵고프 방법: Angoff(1971)가 처음 제안한 방법으로 해당 분야의 전문가(평정자)가 문항을 분석한 뒤 ‘도달’에 속하는 사람들 중에서 경계선에 있는 최소능력 보유자의 문항정답률을 예상하여 분할점수를 산출하는 방법.
 - 북마크 방법: 1996년 NAEP(National Assessment of Educational Progress)에서 성취수준 설정을 위한 방법으로 처음 소개된 후 미국의 28개 주 이상에서 주 단위 성취도 평가의 성취수준 설정 방법으로 상용되고 있으며, 수정된 앵고프 방법에 문항반응이론 모형을 접목시킨 방식으로, 평가 전문가가 문항반응이론에 의해 문항난이도를 추정하고 난이도에 따라 문항을 배열하여 문항집(문항책자)을 만든 후 교사, 혹은 교과내용 전문가에게 제공하는 것에서 시작하여 평정자들은 소그룹, 중그룹 등에서 토의한 후 쉬운 문항부터 검토하여 해당 수준의 최소능력자가 응답하지 못할 것이라고 판단되기 시작하는 문항에 북마크(bookmark)로 표시하는 방법.

○ (수준 설정) 기초미달, 기초, 보통, 우수학력의 4가지 수준으로 수준 설정.

- 등급 수 결정 및 수행수준기술 작성: 맞춤형 자율평가와의 연계를 위해 각 등급의 수준을 정의하는 수행수준기술(performance level description) 차용.
- 수준설정 방법 결정: 평가도구 내용 분석에 의한 방법으로 문항반응이론 분석을 통해 산출한 순서화된 문항집을 이용하는 ‘북마크 방법’과 고전 검사이론 기반의 ‘수정된 앵고프 방법’의 장·단점을 비교하고 검사 및 문항의 특성을 고려하여 수준설정 방법 결정.
- 수준설정위원회 구성 및 운영: 12명의 수준설정위원회(패널) 구성.

- 각 수준설정 위원은 최소능력보유 피험자(minimally competent examinee)의 수행 결과를 예측하고, 이들의 수렴된 예측 결과를 종합하여 등급 분할점수 결정.
 - 등급 산출 및 검증: 평가 결과에 분할점수를 적용하여 등급 산출 후 타당성 검증.
- 수준 설정의 절차는 [그림 3]과 같음.



라. 학생, 교사, 학부모, 학교 대상 조사 시행

1) 통합 홈페이지 시스템 구현

- 조사 참여율 제고를 위해 학교 내 협력 교사 및 학생/학부모의 조사 협조 편의를 제고할 필요가 있음. 이를 위해 조사 응답, 조사 자료 배포, 조사 현황 확인을 단일 경로를 통해 가능하도록 통합 홈페이지 시스템을 구현함.
- 단, [과제 1]과 [과제 2]의 조사가 동일한 홈페이지를 이용할 경우, 학교와 응답자의 혼란 및 불편을 야기할 수 있는 관계로 [과제 1]과 [과제 2]는 각각 별도의 홈페이지를 생성, 이원화하여 운영함.

○ 조사 관리 및 현황 확인

- 조사 정보 보안을 위해 학교별 로그인 ID/PW 제공.
- 학교 정보 등록 완료 시, 등록된 번호로 로그인 정보 문자 자동 발송.
- 통합 홈페이지 시스템 내에서 실시간 응답 현황을 확인 가능하도록 구축.

- 필요 시 재확인이 가능하도록, 홈페이지에 조사 안내문, Q&A, 매뉴얼 등 관련 자료 탑재.

○ 조사 대상 학급 선정 및 동의서 수합 시스템

- (일부 학급만 조사 시) 학교별 조사 대상 학급을 무작위로 선정하기 위한 학급 선정 프로그램 제공하여 대상 학급 세트를 무작위로 선정할 수 있도록 3순위 내외로 제시.
- 조사 대상 학급의 학년, 반, 번호 순으로 패넬조사 참여 동의서 링크가 포함된 가정통신문 자동 생성 기능 제공.
- 조사 대상 학급의 가정에서 온라인 동의서 응답을 완료하면, 홈페이지의 협력 교사 로그인 화면에서 해당 정보를 연동하여 확인 가능.

○ 조사 대상자별 설문 응답 방법

- 학교조사: 교내 응답 가능자 또는 대상자를 선정하여 학교별 대표 1인이 학교 온라인 설문에 응답(학교별 단일 설문 링크 제공).
- 교사조사: 교내 협력교사가 홈페이지에 조사 대상 교원 수를 입력하면, 해당 수 만큼 개별 설문 링크가 형성된 조사 안내문을 자동으로 생성 및 출력하도록 구성.
- 학생조사: 동의서 응답 완료 학생에 한하여, 학교로부터 조사 협조를 받아 학업성취도(영/수), 온라인 설문 순으로 교내 컴퓨터실에서 집체 조사를 진행(사전 설문 1개 시수, 사후 설문+성취도 3개 시수로 총 4개 시수 필요).
 - ※ 사전 성취도 검사는 맞춤형 학업성취도 자율평가(KICE) 활용 예정
 - ※ 사전 설문조사는 설문조사도구 타당화 일정에 따라 조율
- 학부모조사: 학생 개별ID와 매칭되는 학부모 개별ID 자동 생성 및 학부모 조사 개별 링크가 포함된 가정통신문 자동 생성 및 출력 가능. 학부모는 가정에서 학교로부터 가정통신문을 전달받아 조사 링크 또는 QR코드를 활용하여 설문 조사 참여.

2) 온라인 설문 개발

- [그림 4]와 같이 6단계에 걸친 설문 검토 및 보완을 통해 최적화된 온라인

설문 개발.



○ 응답 정확성 확보 필요성

- 자기기입식으로 진행되지만, 정확한 응답을 할 수 있도록 질문 순서나 응답 에러 시 제공되는 메시지, 전체 응답 환경 등이 최대한 일반 대면조사와 유사하게 구성되어야 함.

○ 대응 방안

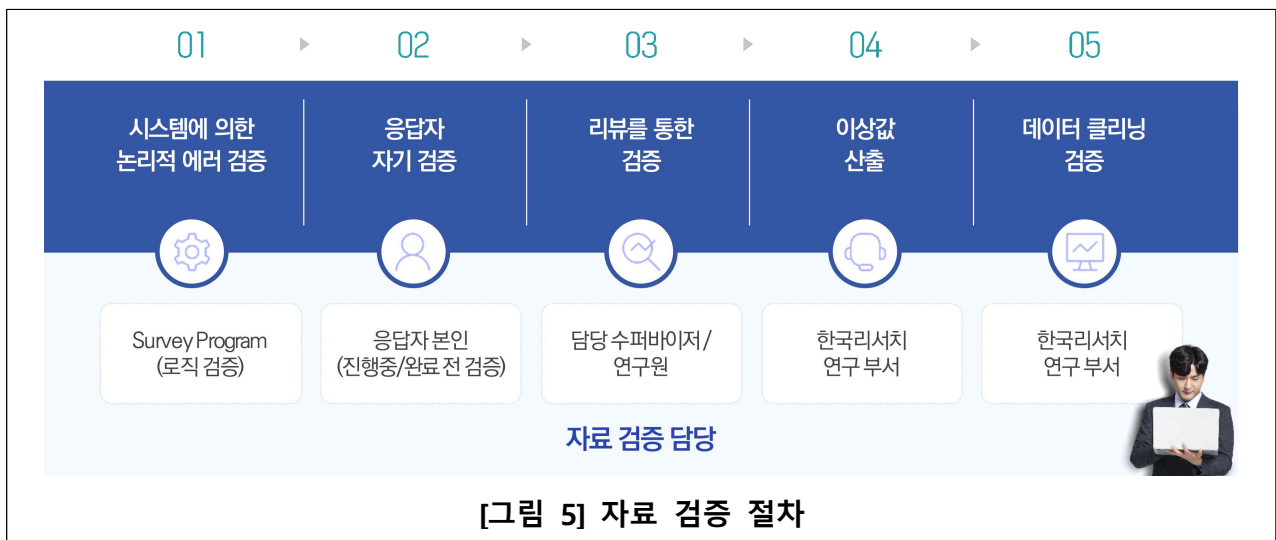
- 학교 유형, 학교급, 학년 등 일반 정보는 사전 DB화하여 입력 에러 최소화.
- 보조자료를 구현하여 미확인 시, 다음 문항으로 이동할 수 없도록 설정.
- 응답하지 않고 다음으로 넘어갈 시, 경고창 노출.
- 오류로 추정되는 응답 값 입력 시, 확인창 노출.
- 가독성 및 보조자료와 설문 내용과의 연결성을 위해 보조자료 고정 제시.

○ 참여 경로별 맞춤형 화면 제공

- 본사의 온라인 조사 시스템은 기기별 화면에도 최적화되어 있어, 응답자 편의에 따라 PC와 Mobile로 조사 참여가 가능하며, 참여 경로 및 맞춤형 안내 및 화면 제공 가능함.

3) 조사 자료 검증 및 처리

- [그림 5]와 같이 5단계 자료 검증 시스템을 통하여 수집된 자료에 대한 신뢰도를 제고함.

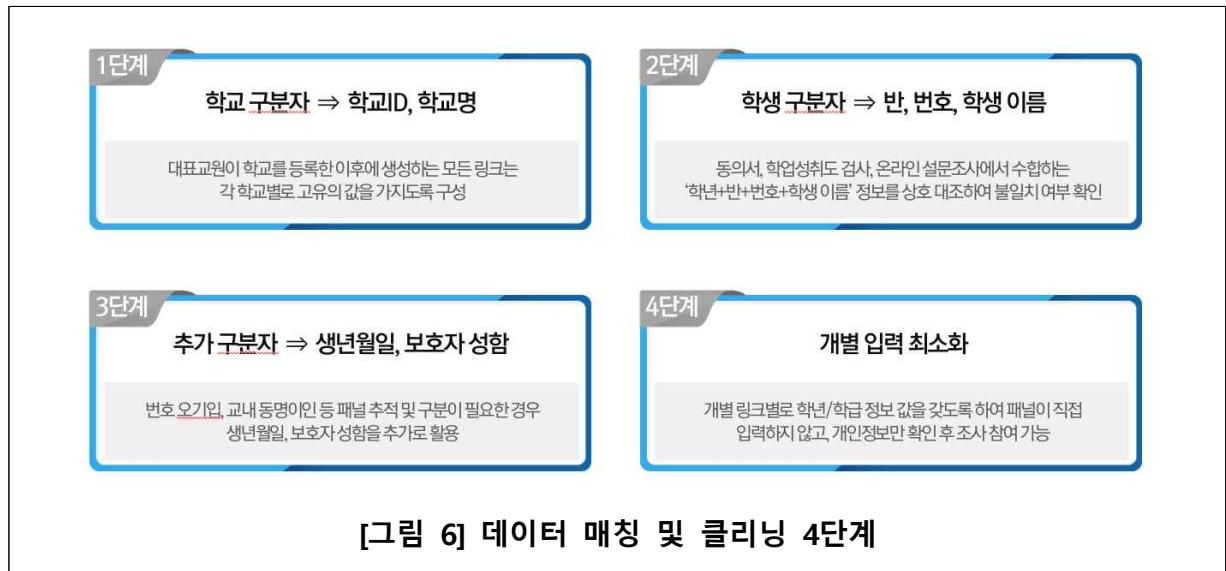


○ 이상값 처리

- 이상값은 ① 빈도표 산출, ② 이상값 확인, ③ 데이터상 해당 샘플 확인, ④ 해당 샘플의 특이사항 변수 내용 확인, ⑤ 해당 값 대체 여부 협의의 5가지 절차를 거침.
- 하드 체크: 논리적으로 확인할 수 있는 클리닝 항목은 데이터클리닝 프로그램을 이용하여 에러 내용을 확인하고 수정함.
- 소프트 체크: 각각의 경우에 따라 변수의 값이 달라져 논리적으로는 확인할 수 없는 클리닝 항목에 대하여 SPSS 프로그램의 Syntax 기능을 이용하여 확인하고 수정함.

○ 데이터 매칭 및 클리닝 리스트 활용

- 데이터 매칭 및 클리닝 리스트 초안 제작.
- 응답 매칭: 학교ID, 교사ID, 학생ID, 학부모ID를 각각 고유 생성하되, 학년, 반, 번호 및 그 외 인적정보를 활용하여 당해 및 차년도 차수와 연계함.



- 기초정보 클리닝: 학교-학생-학부모 간 응답을 비교하여 일치 여부 확인.
- 문항 정보 클리닝: 문항 내 에러, 문항 간 에러, 오픈 문항에 따라 순차적으로 진행.

클리닝 리스트 작성

일련	문항	유형	클리닝 내용	수정 및 클리닝 방향	케이스	제조사 대상여부	제조사 완료 여부	비고
6	문3, 문3-1	하나의 변수로 리코드 필요	문3과 문3-1을 하나의 변수(RQ3)로 리코드 해야 함	리코드				
7	문4, 문4-1	기초, 신규자별 표지	신규 자별인데 문4에 응답이 있는 경우	여러 확인	0			
8	문4, 문4-1	하나의 변수로 리코드 필요	1. 문4와 문4-1을 하나의 변수(RQ4)로 리코드 해야 함 2. 문4-1에서 "단"을 가릴수 있는지 확인해야 함	리코드				
9	문4-1	응답값	1. 문4-1에서 중 "단"이 20 이상일 경우 2. 문4-1에서 중 "단"과 "가"의 합이 나보다 많은 경우	1. q4_1 >= 20 경우 중리닝 방향 결정 필요 2. q1_2 < q4_1 경우 오류	1) 9 2) 재체크 필요			
10	문5, 문26	문항간 표지	문5에서 ①, ② 응답을 했는데, 문26, 4의 응답(이전스독값)이 1 이상 있는 경우	문26, 4(이전스독값)=0으로 클리닝	474			
11	문5, 문5-1	문항간 표지	문5에 ②, ③인데, 문5-1에 응답이 있는 경우	여러 확인 문5-1 b8저지	0			
12	문5-1	문항 표지	문5-1에서 4가지 항목 모두 비수렴인 경우 → 응답 ②, ③, ④, ⑤인 경우	여러 확인	0			
13	문6-문7-3	문항간 표지	문6 응답이 ①, 문7 살고 있다인데 문6-1-문7-3까지 응답이 하나라도 있는 경우	여러 확인 문6-1-문7-3 b8저지	0			
14	문6-1, 문6-2	문항간 표지	문6-1에서 나온 사람 수가 문6-2의 응답과 맞지 않은 경우	여러 확인 → 문6-1의 집계 응답으로 문6-2 클리닝	0			
15	문6-2	응답값	문6-2에서 응답이 20명 이상인 경우	클리닝 여부 결정(Max값 설정)	2			

[그림 7] 클리닝 리스트 작성 예시

○ 데이터 생성 및 관리

- 데이터 클리닝 작업 완료 후, 프로그램을 통한 최종 데이터 생성.
- 복잡한 데이터를 필요한 목적에 맞는 구조로 생성하는 것이 필요하며, 원하는 목적에 맞는 데이터를 Union 프로그램을 이용하여 생성.

- 최종 데이터의 변수 설명, 보기 설명 등의 값을 서버에서 직접 수정하고 관리.
- 데이터의 변수 설명, 보기 설명 등을 수정한 후 업데이트된 코드북 생성.

마. 조사결과 분석을 통한 성과진단 및 분석

○ 수집된 데이터 활용 기초통계 분석 및 비교 분석.

- 학생, 교사, 학부모, 학교 설문조사 결과와 학생 성취도검사 점수에 대한 기초통계 분석 실시.
- 학생 성취도와 함께 사전-사후로 조사되는 내용에 대해서는 사전, 사후에 대한 기술통계 분석과 더불어 두 시점간 변화, 차이에 대한 기술통계 분석 및 비교 분석 실시.
- 또한 학생 특성과 학교 특성, 디지털교육 정책 투입에 따른 집단별 비교분석 실시.

○ 학생의 성장과 변화를 분석하고 영향 요인 도출.

- 연구 모형에 따라, 투입-과정의 변인이 산출 변인(학업성취도 및 자기주도성, 협력적 소통 역량 등)에 미치는 영향에 대한 분석 실시.
- 데이터 구조를 반영하여 다층모형으로 분석하되, 산출 변인을 종속변인으로 하고, 투입 변인과 과정 변인을 순차적으로 모형에 투입하는 위계적 분석 실시(위계적 다층모형분석, 다층 구조방정식모형 등).
- 학생들의 사전 성취도 점수와 사전에 조사된 역량을 고려하여 학생의 성장과 변화에 영향을 미치는 요인을 분석할 수 있도록, 두 시점간 변화나 차이 점수를 산출변인으로 사용하는 방안 고려.

○ 디지털 교육의 성과 진단 및 변화 분석을 통한 정책연구 제안.

- 학업성취도 및 설문조사 결과 분석을 통해 디지털 교육 정책에 대한 정책연구 주제 및 연구방안 제시.

바. 차년도 설문도구 개발 및 타당화

- 1차년도 조사의 기초통계 분석 및 비교 분석을 통하여 차년도 조사에서 수정·보완되어야 할 설문조사 문항 확인 및 수정.
- 차년도 설문도구의 타당화 계획 수립.

사. 성과 확산을 위한 방안 제시

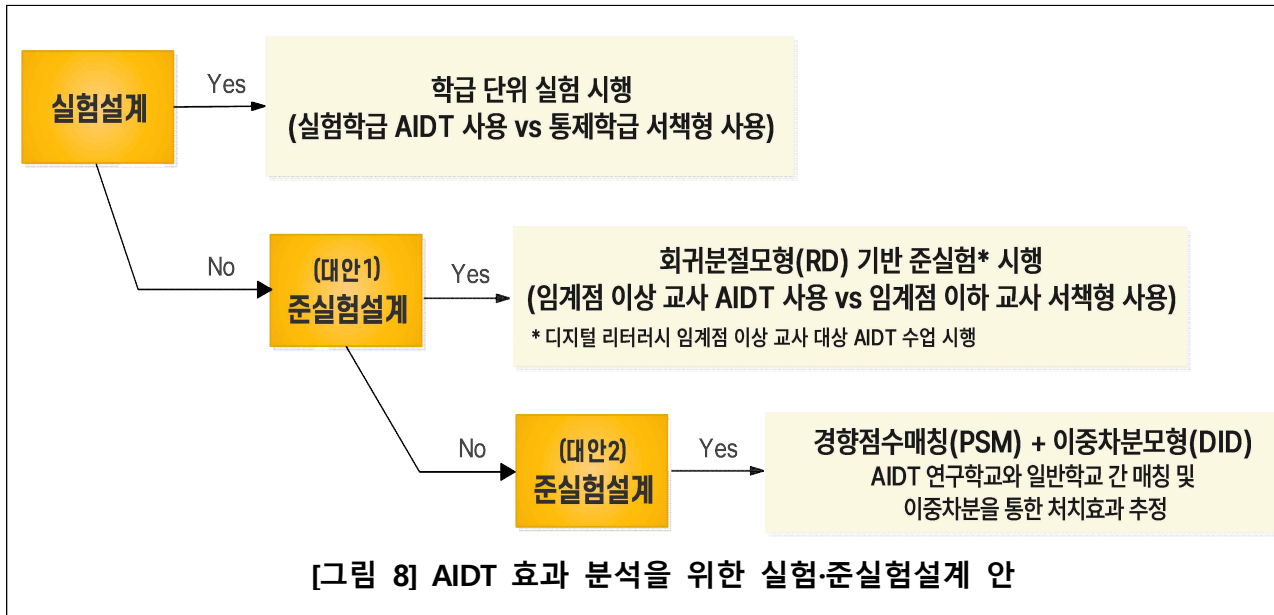
- (데이터셋 구성) 데이터 공개와 관련된 법령 검토를 토대로 공개가능한 데이터 식별 및 데이터셋 구축하고, 데이터 설명 자료 작성.
 - 공개 가능한 데이터셋이 구축되고 나면, 조사 설계 및 표본 추출 절차를 포함하여 학생-교사-학부모-학교 간 데이터 연계 정보, 사전-사후 조사 간 연계 정보 등에 대한 내용을 포함하는 공개용 데이터 관련 사용자 지침서를 코드북과 함께 작성.
- (데이터셋 공개) 홈페이지 공개나 학술대회 등 구축된 데이터의 외부 공개 절차 및 제공 방안 제안.
 - 한국교육중단연구나 한국 아동·청소년패널, 서울교육중단연구 등 기존 국내 교육 중단연구의 데이터셋 공개 현황 및 기관 내 공개 절차를 파악하고, 기관 담당자와의 협의회를 토대로 관련 이슈를 정리하여, 홈페이지 공개 및 학술대회 개최를 위한 적절한 데이터 공개 방안 제안.
 - 또한 구축된 데이터셋을 활용한 정책연구 주제 및 방법론을 제안함으로써, 조사 결과를 활용한 2차 분석 연구의 활성화를 도모.

3-2. [과제 2] 2025년 디지털교육 효과 분석

가. 2025 디지털교육 효과 분석을 위한 실험 연구 설계 방안 제시

- 본 연구는 2025년 AI 디지털교과서(AIDT, AI Digital Textbook)를 사용한 집단과 미사용한 집단을 설정하여 실험을 설계하고, 이를 통해 AI 디지털교과서를 포함한 디지털 교육의 효과를 분석하는 것을 목적으로 함.

- 이를 위해 현장의 상황과 실행 가능성을 종합적으로 고려하여 [그림 8]과 같은 세 가지 설계 방안을 마련하였으며, 디지털 교육의 인과 효과를 추정하기에 가장 적합한 방안을 타진하고자 함.



1) 실험 설계안: 무작위 대조군 실험(Randomized Controlled Trial; RCT)

가) AI 디지털교과서 사용 효과 분석을 위한 실험연구 설계

- 실험 단위: AI 디지털교과서 사용 여부에 따른 학급 단위 실험 설계

- 무작위 대조군 실험(RCT)을 위해서는 실험/통제 집단의 ‘무선 할당(random assignment)’이 원칙이나, AI 디지털교과서 연구학교가 지정*되어 있는 상황에서 ‘학교 단위’의 무작위 대조군 실험 설계는 한계가 존재함.

※ 연구학교 지정 시 연구학교 희망 여부, 재직교원 동의율, 당해 학교의 연구수행 능력 및 연구계획 등을 종합적으로 고려함에 따라(「연구학교에 관한 규칙」 제4조), 연구학교 선정은 무선 할당에 해당하지 않음

- 또한, 「연구학교에 관한 규칙」 제6조에 따라 연구학교의 장이 연구수행에 필요한 교원을 연구학교에 우선 배치할 수 있어, 연구학교와 일반학교 간 교사의 디지털 교육 역량 및 태도 등에 차이가 존재할 수 있음.
- 즉, 연구학교와 일반학교 간에 이질적 특성이 존재하는 상황에서 단순히 학교 유형에 따라 실험/통제 집단을 구분하는 경우 선택편향(selection bias)

의 문제가 발생할 수 있음. 이에, 본 연구에서는 동일 학교 내 ‘학급 단위’의 무작위 대조군 실험을 설계하고자 함.

○ 모집단 및 표집 방법

- 본 연구의 모집단은 전국 AI 디지털교과서 연구학교와 [과제 1]에서 표집된 일반학교* 내 AI 디지털교과서를 도입한 학년(초등 3~4학년, 중학교 1학년)으로 설정함. 실험학교는 AI 디지털교과서 연구학교와 패널데이터에 포함된 일반학교를 대상으로 실험 참여 공모를 통해 최종 선정함.

※ 패널데이터에서 표집된 일반학교를 대상으로 실험 참여 학교를 선정함으로써 처치 효과의 종단적 추적 가능

- 본 연구에서는 무선화 유도 설계(randomized encouragement design) 방식을 활용하여, 연구학교와 일반학교 모두 자발적으로 실험에 참여하도록 실험 참여 공모과정을 설계함. 연구학교는 적극적인 참여를 권장하고, 일반학교는 인센티브 제공을 통해 참여를 유도하여 충분한 표본을 확보하고자 함.

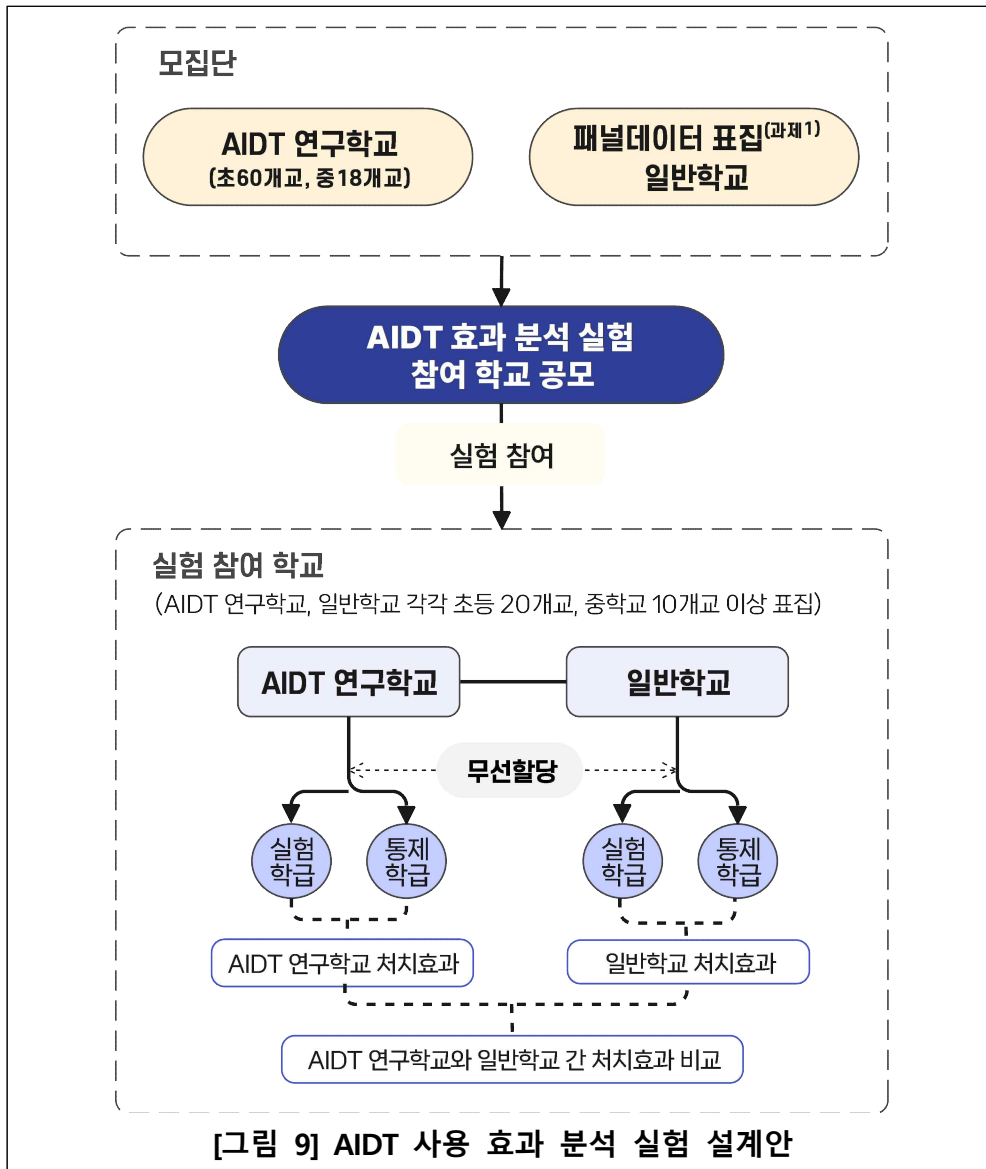
※ 실험 참여 인센티브: 학교별 추가 예산 지급, 교사 대상 AIDT 전문적 학습 공동체 참여 기회 제공, 학생·학부모 대상 학생별 종합 학습 분석 결과 제공 예정

※ AIDT 연구학교와 일반학교 각각 초등 20개교, 중학교 10개교 이상 표집 예정

- 실험 참여 공모에 지원한 학교를 대상으로 학교 규모, 교원 및 학급 수 등을 고려하여 실험 대상 학교를 확정함. 이때 학급 단위의 실험 설계를 위해 실험 학년에 2개 이상의 학급이 존재하는 학교로 제한함.

- 또한 교과 전담 교사가 배치된 중학교와 초등 영어 교과의 경우, 학급 간의 실험 확산(diffusion of treatment)을 방지하기 위해 동일 교사가 실험 학급과 통제 학급을 동시에 담당하지 않도록 하며, 이를 위해 교과 교사가 배치된 학교는 학년당 2개 이상 학급, 2명 이상의 교과 교사가 배치된 학교로 제한함.

※ 담임교사가 영어 수업을 담당하는 초등학교는 학년당 2개 이상 학급이 존재할 경우 실험 참여 가능



○ 처치 특성 및 절차

- 최종 확정된 실험학교의 초등 3·4학년, 중학교 1학년 전체 학급*을 실험 학급과 통제 학급으로 무작위 배정함. 교과 교사가 배치된 초등학교 영어, 중학교의 경우 교사 단위로 AI 디지털교과서 활용 여부를 무작위 배정함으로써 학급 단위 실험을 시행할 수 있음.

※ 가급적 해당 학년 내 전체 학급이 실험에 참여하도록 하며, 비순응 교사 및 학급 발생 시 해당 학급 제외 후 무작위 배정 (세부 사항은 KERIS와 협의 후 조정)

- 실험 학급에는 AI 디지털교과서를 활용하도록 하고, 통제 학급에는 동일 출판사의 서책형 교과서를 활용하도록 함으로써 각 학년의 교육과정은 동일하게 유지되

교수·학습 방식*을 차별화함.

※ (AIDT) AIDT를 통한 개별 학생 맞춤형 학습
(서책형교과서) 교사 중심의 학습 내용 구성

<표 9> AIDT 효과 분석 무작위 대조군 실험(RCT) 절차

구분	대상	연구 기간				
		25년 3~5월	1차 실험 (25년 6~7월)	25년 7월	2차 실험 (25년 9~11월)	25년 11~12월
연구학교	실험학급	사전검사 (성취도*, 설문조사)	AIDT 사용	중간검사 (성취도, 설문조사)	AIDT 사용	사후검사 (성취도, 설문조사)
	통제학급		서책형교과서 사용		서책형교과서 사용	
일반학교	실험학급		AIDT 사용		AIDT 사용	
	통제학급		서책형교과서 사용		서책형교과서 사용	

* 사전 성취도 검사는 맞춤형 학업성취도 자율평가(KICE) 자료 활용

○ 처치 효과 분석

- 실험/통제 학급을 대상으로 사전·중간·사후검사를 실시하며, 이를 통해
① AI 디지털교과서 활용 학급과 미활용 학급 간의 평균 처치 효과, ② AI
디지털교과서 연구학교와 일반학교 간의 평균 처치 효과* 비교 분석을 실
시함.

※ AIDT 연구학교와 일반학교 간의 평균 처치 효과 분석 시 매칭 방법(쌍대클러스터
매칭, 경향점수매칭 등)을 활용하여 사전 특성(학교규모, 지역 등)의 동질성 확보

- 이때, 실험 학급으로 배정되었으나 AIDT를 미사용하거나 특정 기간만 활용
한 경우, 통제 학급으로 배정되었으나 AIDT를 사용한 경우 등 실험적 처치
에 대한 비순응(noncompliance) 혹은 중도 이탈(dropout) 사례가 발생할 수
있으며, 이를 식별하기 위해 AIDT 로그기록을 연계하여 활용하고자 함.
- 인과효과 분석의 엄밀성을 제고하기 위해 최초 무작위 할당된 처치 상태를
기준으로 분석하는 ITT(Intent-to-Treat) 분석을 실시하고, 나아가 무선 할
당된 처치(실험 학급 배정 여부)를 도구변수로 활용하여 실제 처치를 받은
대상의 효과(Treatment-on-the-Treated, TOT)를 추정하고자 함(Bishop &
Thompson, 2023).

○ 예상 쟁점 및 대응 방안

- (학급 배정 관련) 실험/통제 학급을 단위 학교에서 자체 배정할 경우, 무작위 배정이 아닌 교사의 선호 및 특징(교직 경력, 디지털 역량 등)에 따라 배정할 가능성이 존재함. 특히, 초등학교 고경력 교원이 3~4학년 담임을 선호하는 경향이 있는 반면(김미화, 정동욱, 2013), 상대적으로 디지털 역량 수준은 낮은 점을 고려할 때, 실험 학급 배정 시 개인의 선호가 반영되지 않도록 주의할 필요 있음.
 - 따라서 실험 진행 시 연구진이 무선 할당을 시행한 후 학교에서 이를 이행할 수 있도록 교육청 및 KERIS의 협조가 요구됨.
- (중학교 대상 실험) 중학교 1학년의 경우 매 학기 정기시험이 시행됨에 따라, 학급 간 공통된 교과서를 사용하지 않을 경우 민원 제기 가능성이 우려됨.
 - 이에, 중학교의 경우 학교 단위의 준실험설계를 고려하거나, 자유학기제를 활용한 단기 효과 분석을 실시할 필요 있음.
- (윤리적 쟁점 관련) AIDT 연구학교는 승진가산점 대상 학교로(「교육공무원 승진규정」 제41조), 교사 간 AIDT 활용 여부에 차이가 있는 경우 무임승차 혹은 역차별의 문제가 제기될 수 있음. 이러한 문제를 해결하기 위해 선행연구에서는 전체 교사를 실험집단으로 하되, 처치 강도를 차별화하여 실험을 실시하기도 함(Jackson & Makarin, 2018). 이 외에도 연구 종료 후에 집단별로 해당 프로그램을 미사용/사용할 기회를 제공하거나, 동일 집단 내 두 가지 조건(예: AIDT, 서책형)을 시간차를 두고 교차 적용하는 방안 등을 적용한 바 있음(Hegedus et al., 2015).
- (교사 자율성 침해 우려) 단위 학교에서 AI 디지털교과서를 채택하였더라도 AIDT 활용 정도는 교사 재량에 맡기고 있음. 엄격한 실험을 위해 실험 학급에 AIDT 활용을 강제할 경우 교사의 저항 및 중도 이탈이 우려됨. 이에, 연구 초기 단계에서 교사를 이해관계자에 참여시켜 연구 목표 공유 및 AIDT 활용을 유연화하는 등 협업적인 실험 설계가 필요함(Cohen et al., 2003; OECD, 2024)

[참고] 무작위 대조군 실험(RCT)과 교사자율성 침해

1. RCT 실험 조건의 충실도(Fidelity)가 교사의 상황 및 맥락에 대한 적응성을 제한할 수 있음

- 수업은 학생, 교사, 내용 간의 역동적인 상호작용으로 이루어지며 교실의 구체적 맥락에 따라 유연하게 변화할 수 있음. 따라서 수업에서의 개입을 충실도의 기준만으로 제약하는 것은 한계가 있으며, 교사의 자율성과 맥락적 적응성을 함께 고려할 필요가 있음(Cohen et al., 2003).

2. 교사의 신념과 외부의 통제 간의 충돌

- Datnow & Castellano(2000)는 교사가 외부 프로그램이 자신의 교육적 신념과 충돌할 경우 거부감을 느끼게 되며, 그 결과 프로그램의 형식적인 집행자로 전락하게 된다고 지적함. 또한, 교사의 자율성에 대한 존중이 부족할 경우 교사의 동기가 저하되고, 프로그램의 효과성 또한 감소할 수 있음을 강조함.
- Reeve & Cheon(2021)은 교사의 자율성 지지 환경은 교사의 전문성 발휘에 중요하며, 강제적 절차의 구안은 심리적 저항을 유발할 수 있음을 주장함.

3. 프로그램 충실도와 맥락적 자율성 간 균형에 관한 고찰 필요

- OECD(2024)는 실험연구의 집행에 있어 세부적 충실도 대신 핵심 원칙만을 규정하고, 교사에게 일정 수준의 자율성을 부여하는 틈새 자율성(niche autonomy) 모델을 제시함. 예를 들어, 디지털 도구의 사용 자체는 의무화하되(핵심 원칙), 적용 맥락이나 보조자료의 선택과 같은 구체적인 사항은 교사의 자율적 판단에 맡기는 방식의 적응적 충실도(adaptive fidelity)를 도입할 것을 권고함.
- 즉, 실험연구 시행 시 핵심 사항만 규정하고 기타 사항에 대해서는 교사에게 자율성을 허용하는 접근을 통해 기존의 한계를 극복할 필요가 있음을 강조함.

나) AIDT 전문적 학습 공동체 효과 분석을 위한 실험연구 설계

○ 연구 필요성 및 목적

- AI 디지털교과서를 사용하는 집단 내에서도 이를 활용하는 교사의 역량 및 활용 방식에 따라 처치 효과가 다르게 나타날 수 있다는 점(Jeong et al., 2024)을 고려하여, 교사 대상 AIDT 전문적 학습 공동체의 효과에 관한 추가 실험을 제안하고자 함.

○ 실험 단위: AIDT 전문적 학습 공동체 참여 여부에 따른 교사 단위 실험 설계

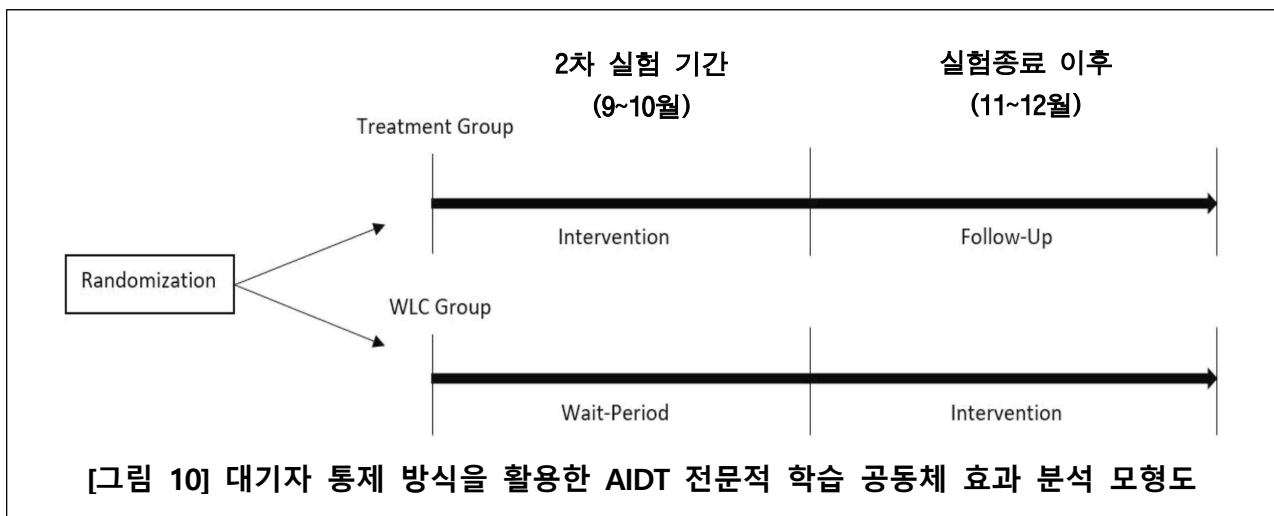
- AIDT 전문적 학습 공동체 참여 교사와 미참여 교사를 무선 할당하여 AIDT 전문적 학습 공동체 참여 효과를 관찰함.

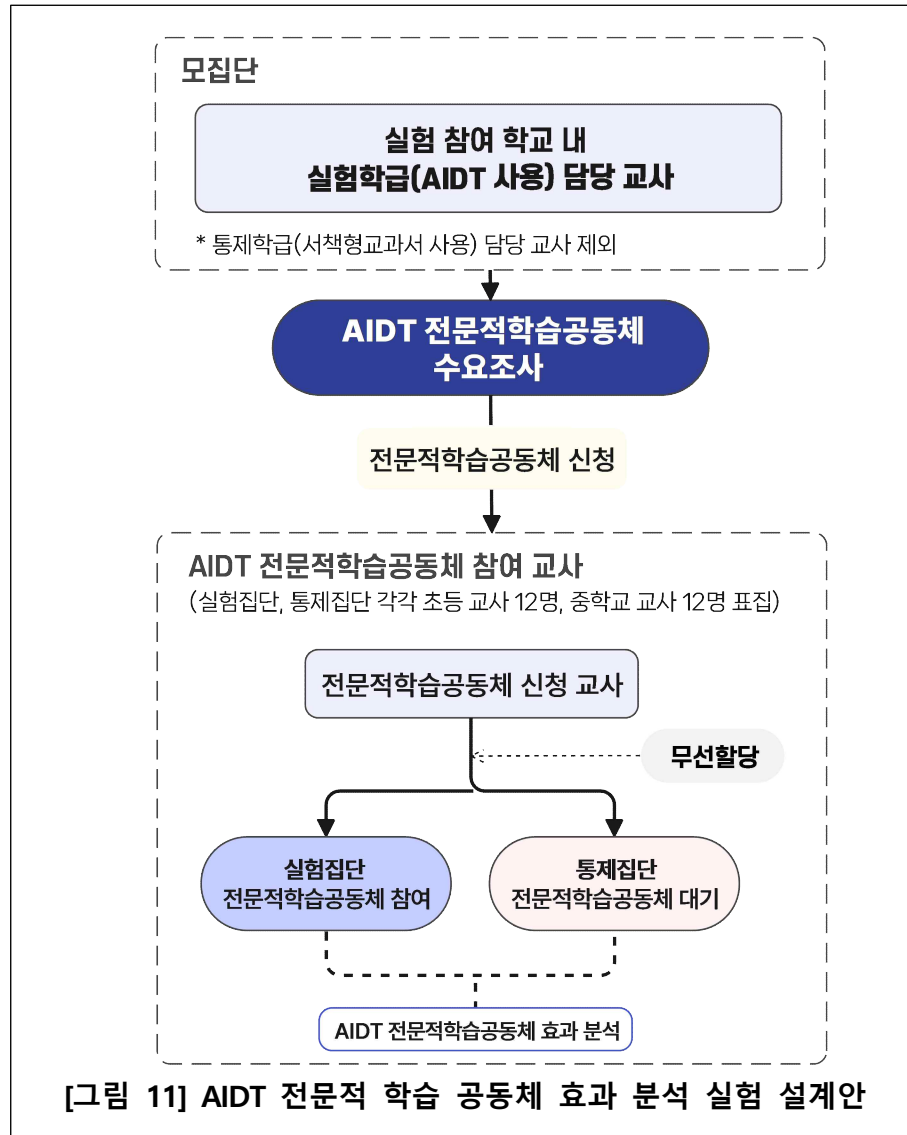
○ 모집단 및 표집 방법

- 본 연구의 모집단은 ‘AI 디지털교과서 사용 효과 분석을 위한 실험’에

참여하는 연구학교 및 일반학교 내 ‘AIDT 활용 교사’로 설정함. 해당 교사를 대상으로 AIDT 전문적 학습 공동체 수요조사를 실시하여 초등교사 12명과 중등교사 12명을 모집하고, 중등교사는 영어와 수학 교사를 6명씩 모집함.

- ※ 교사가 자발적으로 실험에 참여하도록 하는 무선택 유도 설계 방식(randomized encouragement design) 활용
- 실험집단은 2학기 실험기간 중에 AIDT 전문적 학습 공동체 활동에 참여하고, 통제집단은 대기자 명단에 배치하여 실험종료 후에 전문적 학습 공동체 활동에 참여하는 대기자 통제 방식(Wait-list Control Design; WLC)을 활용함(Bishop & Thompson, 2023). 이를 통해 전문적 학습 공동체 참여에 따른 AIDT 활용의 차별적 효과를 분석하고자 함([그림 10], [그림 11] 참조).



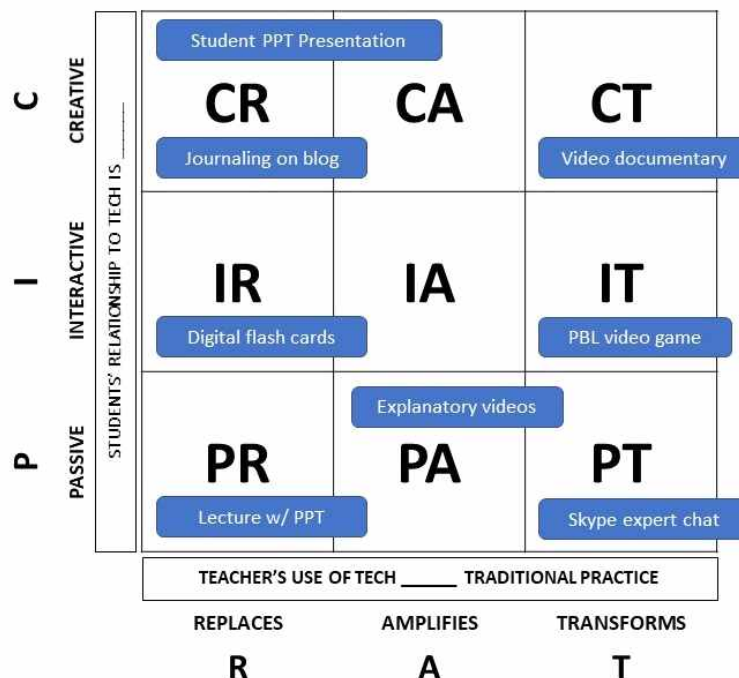


○ 처치 특성 및 절차

- 실험집단 교사 12명을 대상으로 9-10월에 AIDT 전문적 학습 공동체 활동을 실시함. 사전(8-9월)과 사후(10-11월)에 연구에 참여하는 모든 교사(24명)의 AIDT 활용 수업을 비디오로 촬영하고 면담을 실시함. 비디오 촬영은 교사의 수업 활동에 초점을 두고 이루어지며 교실 측면과 후면에 두 대의 캠코더를 설치하고 연구자가 직접 촬영함. 면담은 AIDT의 효과, 수업설계, 학생의 반응, AIDT 활용 수업의 개선점, 향후 수업 계획 등을 포함함.
- AIDT 전문적 학습 공동체 활동은 수학과 영어 교과목으로 구분하여 2팀으로 운영되며 1주일에 1회씩 총 6회에 걸쳐 이루어짐. 회당 소요 시간은 1시간이며 원거리에 있는 교사들이 손쉽게 참여할 수 있도록 실시간 화상회

의를 통해 운영됨.

- AIDT 전문적 학습 공동체에서 교사들은 AIDT 활용 수업 영상을 함께 분석하고 공통으로 직면하는 문제점에 대해 논의함. 연구자가 사전에 촬영한 수업 영상 중 일부를 편집하여 교사들에게 제시하고 수업에 참여한 교사가 AIDT를 어떻게 활용했는지 설명함. 교사들이 각자 영상에서 관찰한 내용을 공유하고 PICRAT 모형(Kimmons 외, 2020)에 따라 영상을 분석함. 교육공학 및 교과교육 전문가의 피드백과 함께 수업 개선을 위한 방안을 논의함.
- PICRAT 모형에 따르면, 학생과 테크놀로지 간 관계를 Passive(소극적), Interactive(상호작용적), Creative(창의적)의 세 범주로 구분하고, 교사의 테크놀로지 활용을 Replacement(대체), Amplification(증강), Transformation(전환)의 세 범주로 구분할 수 있음. 두 개의 축을 서로 결합하여 총 9개의 범주로 테크놀로지 활용 수업을 분류함.



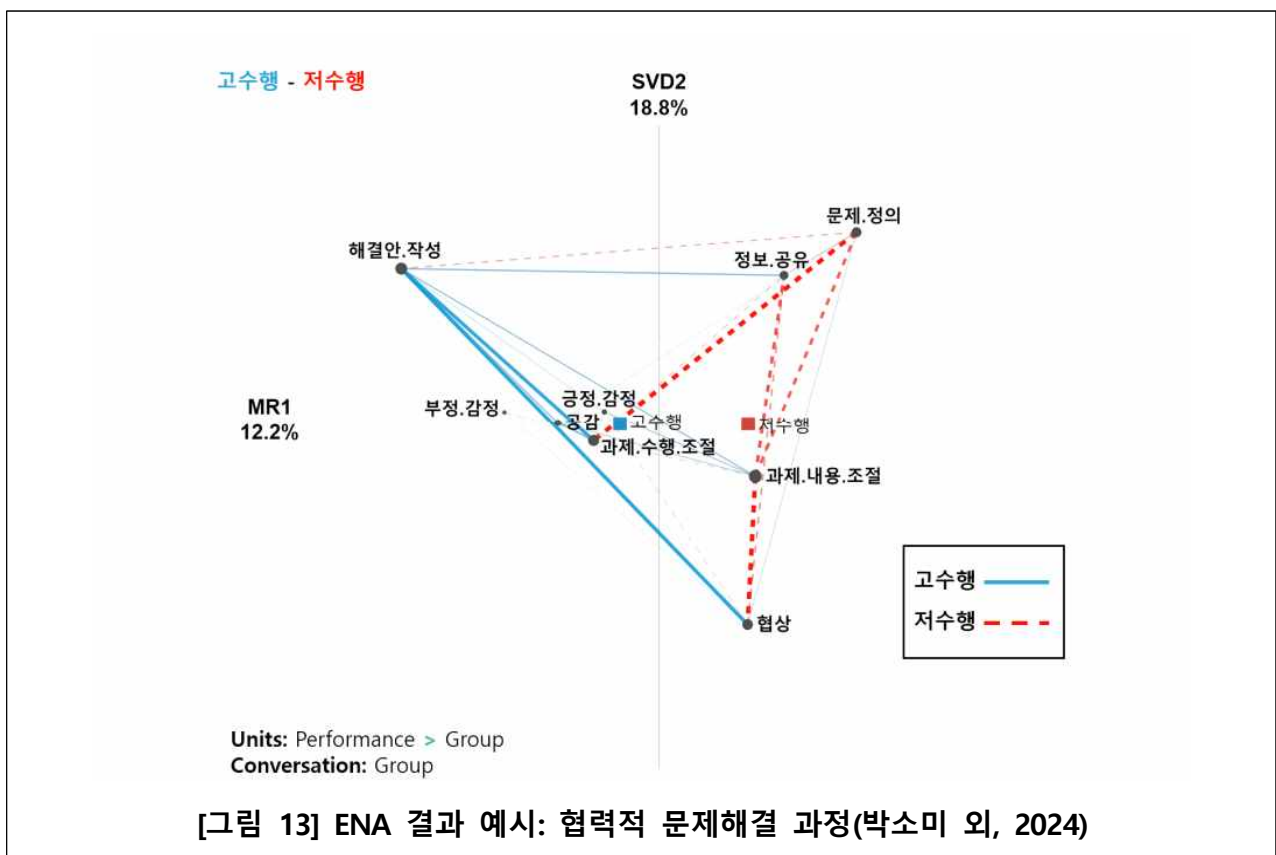
[그림 12] 테크놀로지 활용 수업 분석: PICRAT (Kimmons 외, 2020)

○ 처치 효과 분석

- 영상자료를 분석하기 위해 PICRAT 모형에 따라 구체적인 활동 지표를 도출하고 두 명의 연구자가 ATLAS.ti 소프트웨어를 이용하여 독립적으로 분

석함. 그 결과를 비교하여 평정자 간 신뢰도를 계산하고 불일치하는 내용에 대해 논의를 통해 결정함.

- 실험집단과 통제집단 간 활동 패턴의 차이를 분석하기 위해 ENA(Epistemic Network Analysis)를 실시함. ENA를 통해 교수학습 활동 간 연결성을 분석하고 실험집단과 통제집단의 차이뿐만 아니라 각 집단에 있어 사전검사와 사후검사의 차이를 [그림 13]과 같이 시각화하여 비교함(Shaffer, Collier, & Ruis, 2016).



- 교사와의 면담 내용에 대한 주제 분석을 통해 AIDT 전문적 학습 공동체 활동이 수업 개선에 어떤 영향을 미치는지와 학생의 역량 향상에 어떤 기여를 하는지 질적으로 분석함.
- ‘AI 디지털교과서 사용 효과 분석을 위한 실험’ 연구에서 수집된 학업성취도 검사와 설문조사 결과를 사전검사(7월 시행)와 사후검사(11-12월 시행) 자료로 활용하여 교사의 AIDT 전문적 학습 공동체 참여가 학생에게 미치는 영향을 통계적으로 분석함.

2) 준실험 설계안

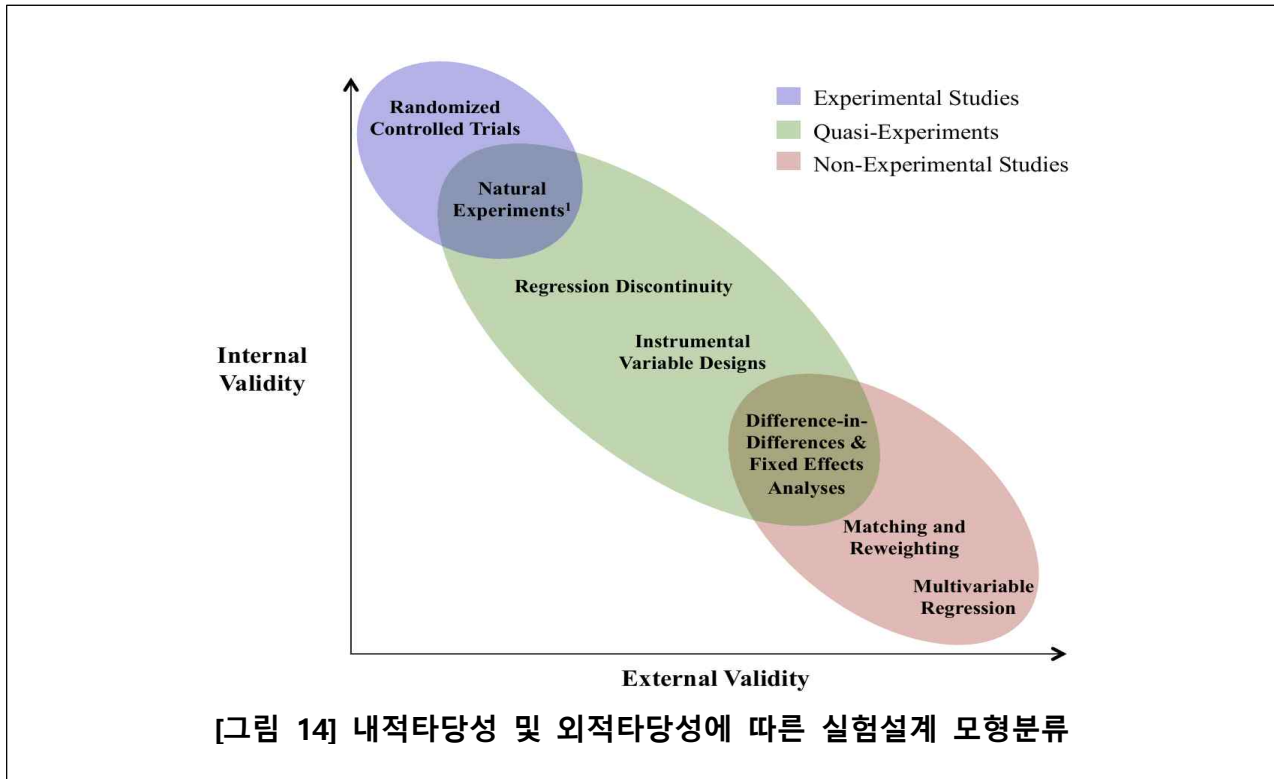
- (실험설계의 어려움) 앞서 제시한 실험 설계 방식은 무작위 추출(random selection)과 무작위 배정(random assignment)을 통해 다른 변수의 영향을 최소화한 상태에서 AIDT를 활용한 디지털교육이 교육성과에 미치는 인과적 효과를 가장 명확하게 추정할 수 있다는 점에서 의의가 있으나, 현실적 측면에서 다음과 같은 한계점이 노정됨(Ong-Dean et al., 2011; Styles & Torgerson, 2018).
- 우선, 일선 학교 현장에서 무작위 배정(random assignment)에 어려움이 있음. 학교 상황을 고려해 볼 때, AI 디지털 교과서(AIDT) 도입 여부를 학교나 학급에 무작위로 배정하기는 현실적으로 쉽지 않음. 실험 연구 기간에 실험집단에는 AIDT를 제공하는 교육활동을 제공하는 반면 통제집단에서는 서책형 교과서를 활용한 전통적 방식의 교육활동이 이루어지게 될 것임. 이 과정에서 실험집단으로 배정된 학생과 학부모 중 AIDT를 활용한 교육활동에 참여하길 희망하지 않을 경우 원활한 연구 진행이 어려울 수 있음. 반대로 AIDT를 활용한 수업에 참여하길 희망하는 학생과 학부모가 통제집단에 배정될 경우 조사 참여에 대한 저항이 존재할 수 있음. 특히, 조사 대상에 포함된 중학교 1학년의 경우 학교의 정기시험(중간고사, 기말고사 등) 결과가 상급학교 진학 과정에서 평가 항목으로 포함될 여지가 크다는 점에서 무작위 배정에 어려움을 야기함.
- 다음으로, 연구참여자의 희망 의사를 충분히 고려하지 않은 상황에서 무작위 배정을 이유로 실험집단이나 통제집단 참여를 강제할 경우 윤리적 문제(Ethical Issues)를 초래할 수 있음. 가령, 학부모나 학생이 AIDT 사용이 학업성취도 향상에 더욱 유리하다는 인식을 갖게 될 경우 통제집단으로 배정된 학생과 학부모의 불만과 민원이 발생할 수 있음. 이런 상황에서 실험설계에 의한 연구 참여를 강제하여 실험집단 학생에게만 AIDT를 제공하고 통제집단 학생에게는 AIDT를 사용하지 못하게 하는 것은 윤리적 문제를 불러 일으킴.
- 아울러, 이런 상황에서 표본 이탈(sample attrition) 문제 역시 발생할 수 있음. 무작위 배정에 따라 실험집단과 통제집단으로 할당된 학생과 학부모가

연구 상황에 불만이 발생할 경우 중단조사 과정에서 이탈하는 상황이 발생할 수 있음. 이렇게 실험 과정에서 학생이 중도에 이탈할 경우 실험집단과 통제집단 구성에 변화가 발생하며, 표본 이탈이 발생한 자료를 분석하고 해석하는 과정에서 편향이 발생할 수 있음.

- 또한, 실험설계는 연구자의 개입을 통해 수업 방식과 AIDT 활용 방식 등을 정교하게 통제해야 한다는 점에서 높은 비용과 관리 부담이 존재함. 연구진이 무작위 배정, 처치 통제, 데이터 수집 및 분석 등과 같은 전체적인 과정을 관리해야 하기 때문에 연구자가 부담해야 하는 시간과 비용이 큼. 특히, 학교와 교사들이 연구에 적극적으로 협력하지 않거나 AIDT를 활용한 실험설계에 학교의 거부감이 클 경우 실험 자체가 원활하게 이루어지기 어려울 수 있다는 한계가 존재함.
- (준실험연구 설계 병행의 필요성) 이런 맥락에서 무작위 배정 없이 처치의 인과적 효과를 추정하는 준실험연구 설계(Quasi-experimental design)를 통해 대안적 접근을 시도할 필요가 있음.
- 준실험연구 설계를 활용할 경우, 다음과 같은 이점이 존재함(Bärnighausen et al., 2017). 우선, 현실적 제약으로 인해 무작위 배정(Random Assignment)이 어려울 경우 준실험연구 설계는 대안이 될 수 있음. 실제 학교 현장에서는 학생과 학부모의 요구와 기대 수준이 다르기 때문에 연구자가 임의로 AIDT 사용 여부를 결정하기 어려움. 준실험설계는 무작위배정을 고려하지 않지만 자연발생적 처치 구조를 인정하고 이를 활용해서 정책 처치의 효과를 분석할 수 있다는 점에서 의의가 있음.
- 다음으로, 준실험연구 설계는 실험설계에서 발생하는 윤리적 문제를 회피하는 데 도움이 됨. AIDT를 활용한 교육활동에 참여한 학생들이 교육성과와 혜택을 얻을 수 있는 상황에서 통제집단 학생에게 참여할 기회를 배제하는 것은 윤리적 문제가 될 수 있음. 이에 반해 준실험설계는 자연스럽게 AIDT를 도입한 집단과 그렇지 않은 집단을 비교 대상으로 삼아 분석하기 때문에, 학생과 학부모, 교사의 반발을 최소화할 수 있다는 이점이 있음.
- 또한, 준실험연구 설계는 비용과 관리 측면에서도 실험설계의 현실적 대안이 될 수 있음. 실험설계 방식으로 전국 단위로 수백 개의 AIDT 선도학교

및 연구학교를 대상으로 AIDT의 교육적 효과를 분석하려면 막대한 비용과 시간이 소요됨. 이에 반해, 준실험설계는 종단조사를 통해 수집된 데이터를 활용하거나 자연스러운 AIDT 도입 상황을 분석 대상으로 삼을 수 있다는 점에서 상대적으로 적은 비용과 시간으로 AIDT의 효과를 분석할 수 있다는 이점이 있음. 특히, 준실험설계는 회귀불절모형(Regression Discontinuity, RD), 도구변수 추정법(Instrumental Variable Regression, IVR), 이중차분법(Difference-in-Differences, DID), 경향점수매칭(Propensity Score Matching, PSM) 등 다양한 분석 기법을 활용하여 무작위 배정이 없더라도 최대한 인과추론의 엄밀성을 확보할 수 있음.

- 이와 함께, 준실험설계는 실험설계보다 상대적으로 외적 타당성(External Validity)이 높다는 점에서도 활용의 의의가 있음. 준실험설계는 실제 교육 현장에서 자연스럽게 이루어진 AIDT 활용 현황을 활용해서 분석을 진행함. 실험설계와 같이 인위적으로 통제된 실험 환경이 아닌, AIDT를 활용한 상황을 그대로 반영하기 때문에, 연구 결과를 일반 학교 현장에 일반화하기 용이함. 이를 통해 AIDT 정책 확산이나 개선 방향을 제시하는 데 실질적으로 기여할 수 있음.
- 실험설계 모형은 내적 타당성(Internal Validity)과 외적 타당성(External Validity)의 수준에 따라 [그림 14]와 같이 구분됨(Geldsetzer & Fawzi, 2017). 준실험설계 중에서도 회귀불절모형(RD)은 우수한 효과 추정치를 산출하여 실험 설계에 근접한 준실험설계로 분류되며, 이중차분법(DID)은 준실험설계에 속하지만, 내적 타당성 확보를 위해 실험집단과 통제집단 간 사전 특성의 유사성을 확보하는 것이 중요함(Shadish, Cook, & Campbell, 2002). 이로 인해 DID 활용 시 경향점수를 통해 실험집단과 통제집단을 유사하게 매칭한 후 분석하는 방식이 권장되기도 함(Smith & Todd, 2005).
- 본 연구에서는 실험설계의 대안으로 회귀불절모형(RD) 기반 준실험설계와 경향점수매칭 및 이중차분법(PSM-DID)을 활용한 준실험설계 방안을 제시하고자 함.

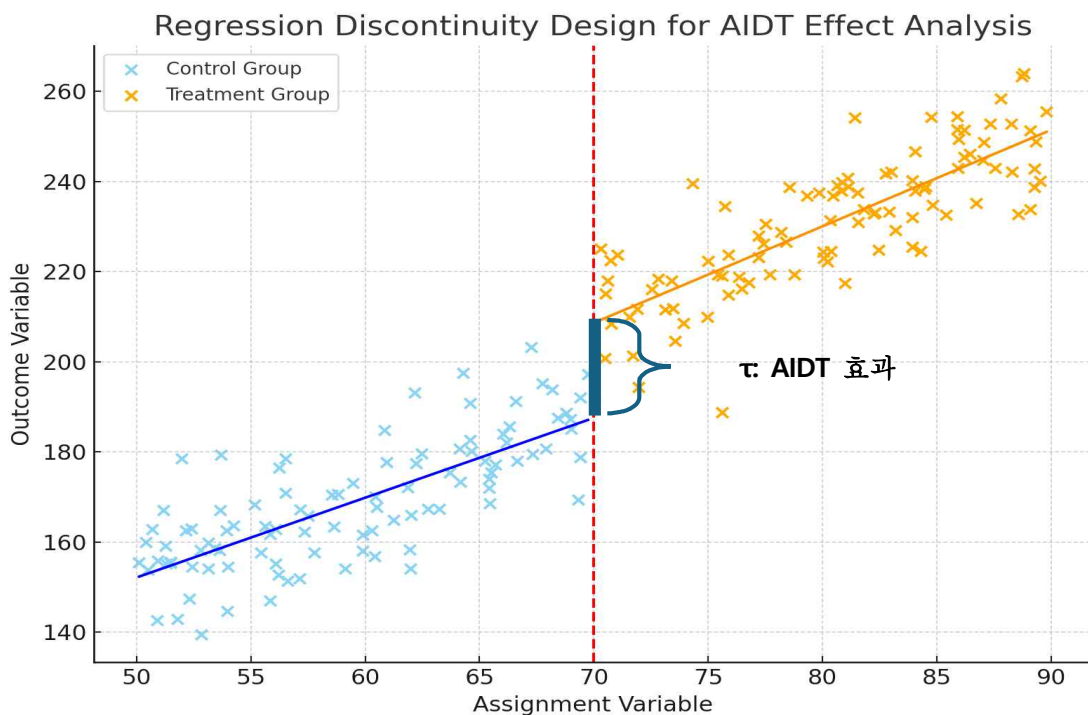


○ (준실험설계 모형) 회귀분절모형(Regression Discontinuity, RD) 기반 준실험설계

- (개념) AIDT 도입 및 사용과 관련해서 무작위 배정이 이루어지지 않을 경우 대안으로 생각해 볼 수 있는 준실험설계 방식은 회귀분절모형을 활용한 접근 방식임. 현실에서 AIDT 도입은 교육청과 학교의 디지털 인프라 수준이나 학생의 디지털 역량 수준에 관한 평가를 토대로 결정될 가능성이 큼. 즉, 디지털 인프라가 좋고 학생들의 디지털 역량 수준이 높은 학교가 AIDT를 우선적으로 도입할 가능성이 높은 편임. 이런 상황에서 AIDT 도입 학교(실험집단)와 미도입 학교(통제집단)를 비교할 경우 원래 학교의 여건과 수준 차이로 인해 AIDT 효과 분석 결과가 왜곡될 가능성이 있음.
- 회귀분절모형은 처치 여부가 특정한 임계점(cutoff)을 기준으로 결정될 때, 해당 임계점 근처에 있는 실험집단과 통제집단의 결과값을 비교하여 처치 효과를 추정하는 방법임(Imbens & Lemieux, 2008). 달리 말해, AIDT를 도입한 집단과 그렇지 않은 집단 중에 임계점 근처에 있는 관측치의 결과값을 비교하여 AIDT 도입에 따른 효과를 분석하는 접근 방식임. 회귀분절모형은 무작위 배정 없이도 임계점 부근의 관측치만을 대상으로 임계값 주변

의 국소비교(local comparison)를 진행한다는 점에서 인과추론이 가능함(Hahn et al., 2001).

- 회귀분절모형은 아래의 그림과 같이 교사의 AIDT 활용 능력을 기준으로 AIDT 디지털 교과서 도입 여부가 결정되는 상황을 기반으로 회귀분절모형을 적용하여 AIDT의 교육적 효과를 추정하는 구조를 나타낸 것임. 가로축은 교사의 AIDT 활용 능력(assignment variable or running variable)을, 세로축은 학생의 학업성취도 또는 자기주도학습역량과 같은 결과변수를 의미함. AIDT 활용 능력 임계값(cutoff)을 기준으로 AIDT를 도입한 집단(실험집단)과 미도입한 집단(통제집단)이 구분되며, 각 집단에 대해 추정한 회귀선이 제시되어 있음. 여기서 임계값을 기준으로 두 집단의 결과변수 평균이 불연속적으로 변화하는 크기(τ)가 AIDT 도입으로 인한 효과를 의미함.



[그림 15] 회귀분절모형(RD)을 활용한 AIDT 효과 분석 모형도

- 이러한 회귀분절모형의 추정 결과가 인과추론에 근거한 결과라는 점을 보장하기 위해서는 몇 가지 가정이 필요함(Imbens & Lemieux, 2008). 구체적으로, 임계값 근처의 관측치가 본질적으로 비슷하다는 임계값 근처의 무작

위성 가정, 임계값 부근에서 처치 여부를 제외한 모든 특성이 유사하기 때문에 임계값에서 처치 이외에 모든 변수들은 연속성(continuity)을 가진다는 가정, 임계점은 인위적으로 조작할 수 없다는 가정 등을 전제하는데, 이런 가정이 충족되어야 AIDT 도입의 인과적 효과 추정이 가능함.

- (조사 설계) AIDT 효과 분석을 위해 회귀분절모형(RD)을 활용한 준실험설계는 <표 10>과 같은 절차로 구성됨.

<표 10> 회귀분절모형(RD)을 활용한 준실험설계(안)

설계 단계	설계 내용
기준 변수 설정	<ul style="list-style-type: none"> ▪ AIDT 도입 여부를 결정하는 기준변수(assignment variable) 설정 ▪ 교사(학급) 수준에서 ① AIDT 활용 능력 ② 디지털 교육 준비도 ③ 디지털 기기 활용 수업 빈도 등을 기준변수로 활용 가능
사전 정보 수집	<ul style="list-style-type: none"> ▪ AIDT 시행 이전에 조사에 참여하는 교사(학급)을 대상으로 해당 학교 또는 학급의 사전 특성, 기준 변수, 결과변수 등에 관한 정보 수집 ▪ 이 때, 기준 변수는 처치 여부를 결정하는 기초 자료로 활용
임계값 정의	<ul style="list-style-type: none"> ▪ 수집된 사전 정보를 활용하여 기준변수의 분포를 검토한 후 AIDT 도입 여부를 결정하는 기준변수의 임계값 설정 ▪ 예) 교사의 AIDT 활용 능력 50점, 디지털 교육 준비도 80점 이상 도입, 디지털 기기 활용 수업 비율 60% 이상 도입 등
AIDT 시행	<ul style="list-style-type: none"> ▪ 임계점을 넘은 교사(학급)을 대상으로 AIDT 수업 진행(처치 집단) ▪ 임계점 미만인 학교 또는 학급은 기존 방식대로 수업 진행 (통제집단) ▪ 두 집단의 학업 성과, 자기주도성, 협력역량 등 결과변수에 관한 정보 수집
모형 설정	<ul style="list-style-type: none"> ▪ 임계값 주변(예: ± 5점, ± 10점) 범위에 포함된 관측치를 중심으로 분석 대상 선정 ▪ Sharp RD vs. Fuzzy RD, 로컬 선형회귀 vs. 로컬 다항회귀, 최적 범위(optimal bandwidth) 등을 고려하여 최종 분석 모형 설정
정책효과 추정	<ul style="list-style-type: none"> ▪ RD 모형을 활용하여 AIDT 효과 분석 진행 ▪ 임계점 부근에서의 사전 특성의 연속성 검정 수행 ▪ 강건성 분석(Robustness test) 수행

- 우선, 기준변수 설정 단계에서는 AIDT 도입 여부를 결정하는 기준변수(assignment variable)를 설정함. 이 기준변수와 관련해서 다양한 접근 방식이 가능함. 예컨대, 교육청 및 학교 수준에서 디지털 기기 보유 현

황, 1인당 디지털 기기 수, 디지털 교육 준비도 등을 기준변수로 활용할 수 있음. 또한 교사 혹은 학교 수준에서는 디지털 기기 활용 수업 빈도와 디지털 교육 준비도와 같은 항목도 설정할 수 있음. 기준변수에 따라 AIDT 도입 여부 및 우선순위가 결정되어 RD 설계가 가능해지기 때문에 명확한 기준변수 설정이 요구됨.

- 다음으로, 임계값 설정 단계에서는 AIDT 도입 여부를 결정하는 기준변수의 임계값을 설정해야 함. 예를 들어, 학생 1인당 디지털 기기 보유수가 0.75대 이상이거나 디지털 교육 준비도가 80점 이상인 경우 AIDT를 도입하거나 디지털 기기 활용 수업 비율이 60% 이상인 학급을 대상으로 AIDT를 도입하는 방안을 고려해 볼 수 있음.
- 사전 정보 수집 단계에서는 AIDT 시행 이전에 조사에 참여할 학교 또는 학급을 선정하고, 해당 학교나 학급의 사전 특성, 기준변수, 결과변수에 대한 정보를 수집함. 학교나 학급의 사전 특성에 관한 정보를 수집하는 것은 임계점 부근에서 사전 특성의 연속성을 확인하기 위함이며, 기준변수의 경우 처치 집단과 통제 집단을 구분하는 기준이 된다는 점에서 자료 확보가 필요함.
- AIDT 시행 단계는 앞서 설정한 기준변수를 기준으로 실험집단과 통제 집단을 구분하고 실험집단을 대상으로 AIDT 수업을 실제로 진행하는 단계임. 임계값을 충족한 교사 또는 학급(실험집단)을 대상으로 AIDT 수업을 적용하고, 임계값 미만인 교사 또는 학급(통제집단)은 기존 수업방식을 유지함. 일정 기간 동안 AIDT 수업을 진행한 이후 두 집단을 대상으로 학업성취도, 자기주도성, 협력역량 등에 관한 정보를 수집함.
- 모형 설정 단계는 임계값 주변(예: ± 5 점, ± 10 점) 범위에 해당하는 관측치만 분석 대상으로 선정함. 이 과정에서 회귀분절모형의 분석모형을 구체화하는 노력이 요구되는데, Sharp RD와 Fuzzy RD, 로컬 선형 회귀(local linear regression), 로컬 다항회귀(local polynomial regression), 최적 범위(optimal bandwidth) 설정 등에 관한 사항을 고려해서 분석모형을 설정함.

- 정책효과 추정 단계에서는 회귀분절모형을 이용하여 AIDT 도입의 효과를 실질적으로 분석함. 회귀분절모형을 활용해서 AIDT 효과를 추정하고, 임계값 부근에서 사전 특성의 연속성을 검증하여 모형의 타당성을 검토함. 아울러, 강건성 분석(robustness test)을 수행하여 분석결과와 타당성과 신뢰성을 제고함.
- (효과 분석 방법) 회귀분절모형(RD)을 기반으로 준실험설계를 적용할 경우 크게 Sharp 회귀분절모형과 Fuzzy 회귀분절모형 등 두 가지 분석방법을 적용해서 AIDT의 교육적 효과를 분석할 수 있음(Lee & Lemieux, 2010).
 - 첫째, Sharp 회귀분절모형은 특정 기준(cutoff)을 기준으로 처치 여부가 결정되는 상황에서 기준점 주변의 관측치들을 비교하여 정책이나 프로그램의 효과를 추정하는 방법임(Lee & Lemieux, 2010). 이를 AIDT 효과 분석에 적용해 볼 수 있음. 예컨대, 각 학교를 대상으로 ① 디지털 기기 보유 현황 측면에서 학생 1인당 디지털기기 수, ② 디지털 교육 준비도 등을 활용하여 “디지털 준비 지수”라는 복합지수(100점 만점)를 산출하고 해당 지수가 70점 이상인 학교만을 대상으로 AIDT 활용 수업을 실시하는 상황을 고려할 수 있음.
 - Sharp 회귀분절모형은 다음과 같은 수식을 활용해서 AIDT의 효과를 추정함.

$$Y_i = \beta_0 + \beta_1 R_i + \beta_2 D_i + \beta_3 (R_i \times D_i) + \epsilon_i,$$
 여기서 Y_i 는 결과변수를, D_i 는 AIDT 도입 여부($D_i = 1 \{R_i > c\}$, 즉, 디지털 준비 지수(R_i) 70점 이상 (c)일 경우 1, 아닐 경우 0)를, R_i 는 기준변수인 디지털 준비 지수 점수를, $R_i \times D_i$ 는 AIDT 도입 여부와 디지털 준비 지수의 상호작용항을 각각 의미함. 이 중 이 연구의 관심 회귀계수는 β_2 로 임계점에서 실험집단과 통제집단의 결과변수 차이가 바로 AIDT의 효과라고 볼 수 있음.
 - Sharp 회귀분절모형의 핵심은 디지털 준비 지수 70점 부근, 예를 들면 69점에서 71점 사이에 해당하는 학교들을 비교하는 데 있음. 이는 69점과 71점 사이의 학교는 디지털 준비 지수는 매우 비슷하지만 AIDT

의 도입 여부만 다르다는 점에서 차이가 있음. 만약 71점 학교 학생들의 성취도가 69점 학교 학생들보다 유의미하게 높다면 이는 AIDT 도입이라는 처치 때문이라고 해석함. 이처럼 기준점 바로 근처의 관측치들만 비교하여 처치 이외의 다른 요인들은 통제하고 처치의 순수한 효과만을 추정할 수 있다는 점이 Sharp 회귀분절모형의 장점임.

- 둘째, Fuzzy 회귀분절모형은 처치 여부가 임계값(cutoff)을 기준으로 “완전히” 구분되지 않고 임계점을 기준으로 처치 받을 확률만 달라지는 경우에 적용하는 방법임(Lee & Lemieux, 2010). 쉽게 말해 기준변수의 임계점을 넘었을 때 100% 처치를 받지 않고 처치 받을 확률이 높아진다는 점에서 sharp 회귀분절모형과 차이가 있음. AIDT 사례로 적용해보면, 각 학교마다 디지털 준비 지수를 산출하고 이 점수가 70점을 넘을 경우 AIDT를 도입하는 상황은 동일하나, 70점을 넘은 학교 중 일부 학교는 AIDT를 도입하지 않거나, 70점 미만 학교 중에서 일부 학교는 자체적으로 AIDT를 도입하는 상황이 발생할 수 있음. 이렇게 디지털 준비 지수 70점이라는 임계값이 도입 확률을 급격히 변화시킨 하지만 완벽하게 처치 여부를 결정하지 않는 상황에서 적용 가능한 방법이 Fuzzy 회귀분절모형임.
- 이럴 경우 다음과 같은 2단계 최소자승법(Two-Stage Least Squares, 2SLS)을 적용함. 1단계 모형은 $D_i = \gamma_0 + \gamma_1 1(R_i \geq c) + g(X_i) + \nu_i$, 여기서 D_i 는 AIDT 도입 여부를, R_i 는 기준변수인 디지털 준비 지수를, $1(R_i \geq c)$ 는 디지털 준비 지수가 임계값($c = 70$)을 넘었는지 여부를 나타내는 더미변수를, $g(X_i)$ 는 기준변수인 디지털 준비 지수에 따른 AIDT 도입 여부의 변화를, ν_i 는 오차항을 의미함. 2단계 모형은 1단계 모형에서 추정한 \hat{D}_i 값을 활용해서 결과모형을 추정하는 모형임.

$$Y_i = \beta_0 + \beta_1 R_i + \beta_2 \hat{D}_i + \beta_3 (R_i \times \hat{D}_i) + \epsilon_i,$$

여기서 \hat{D}_i 는 1단계 모형에서 예측한 처치 확률을 의미함. 이렇게 해서 추정한 AIDT의 효과 추정치(β_2)는 임계값 주변에서 결과변수 Y_i 의 불연속 변화량을 임계값 주변에서 처치확률 D_i 의 불연속 변화량으로 나

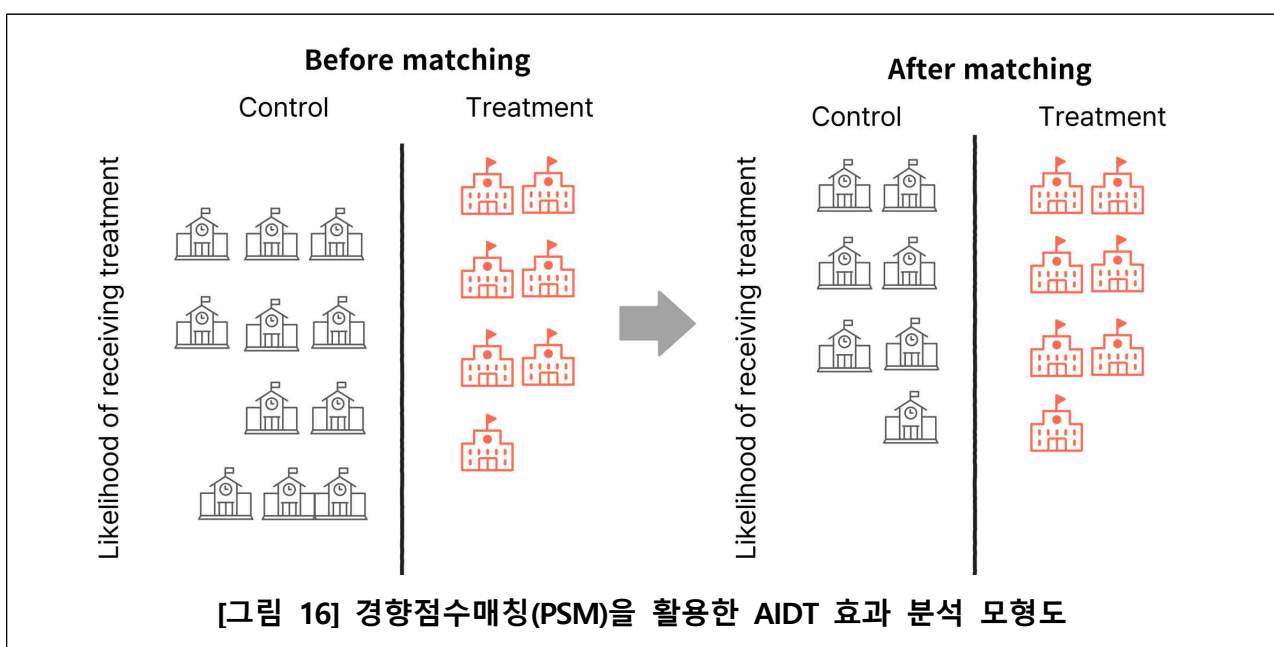
눈 값을 의미함.

- (필요 변수) 회귀분절모형을 적용하여 AIDT 효과를 분석하기 위해서는 몇 가지 핵심 변수를 조사 및 수집해야 함.
 - 첫째, 기준변수(assignment variable)에 관한 정보 수집이 필요함. 이 변수는 AIDT 도입 여부를 결정하는 연속형 변수로, 교육청 및 학교 수준에서 ① 디지털 기기 보유 현황 측면에서 학생 1인당 디지털기기 수, ② 디지털 교육 준비도, 교사 혹은 학급 수준에서 ① 디지털 기기 활용 수업 빈도, ② 디지털 교육 준비도 등을 고려할 수 있음. 이 기준변수는 AIDT 처치 여부를 나누는 기준이 되는 것뿐만 아니라, 분석 과정에서도 결과변수에 대한 기본적인 경향성을 통제하기 위한 변수로 활용된다는 점에서 중요한 변수라고 볼 수 있음.
 - 두 번째로 필요한 변수는 임계값(cutoff point)임. 이는 기준변수 상에서 AIDT 처치 여부가 바뀌는 특정 점수 또는 구간을 의미함. 예를 들어, 디지털 준비 지수 70점을 임계값으로 설정했다면 이 값 이상에서는 AIDT를 도입하고 그렇지 않은 경우 도입하지 않는 방식으로 조사 설계가 이루어져야 함.
 - 세 번째로, 처치 변수(treatment variable)에 관한 정보가 필요함. 이 연구에서는 실제로 학교나 학급이 AIDT를 실제로 도입했는지를 나타내는 변수를 의미함. 회귀분절모형은 임계값을 넘은 경우 처치가 이루어졌는지를 확인해야 하는데, 이에 따라서 실제 처치 여부를 0과 1로 기록한 명확한 변수 관리가 필수적임.
 - 네 번째로, AIDT의 성과를 보여주는 결과변수를 명확하게 확인해야 함. 결과변수는 AIDT의 교육적 효과를 평가하기 위해 필요한 변수로, 이 연구는 학업성취도 점수(수학, 영어), 자기주도학습역량, 협력적 소통역량, 문제해결능력 등 학생 역량에 관한 변수를 활용함. 이때, 결과변수는 AIDT 시행 이후 학생들의 변화를 포착할 수 있어야 하며, 가능하다면 처치 이전(pre-treatment) 데이터도 함께 확보하여 분석의 타당성을 높일 필요가 있음.

○ (준실험설계 모형) 경향점수매칭(Propensity Score matching, PSM) 기반 준실험적 분석

- (개념) 현실에서는 AIDT 도입 및 사용이 무작위로 배정되지 않을 가능성이 큼. 가령, 학교나 학생의 재량적 판단 하에 AIDT 사용 여부를 선택할 수 있으며, 학교의 ICT 인프라 여건 및 상황에 따라 AIDT 도입 여부가 달라질 수 있음. 이럴 경우 실험집단과 통제집단은 사전 특성이 다르기 때문에 단순히 실험집단과 통제집단의 사전 사후 비교를 통해 AIDT 효과를 분석하는 것은 정책 효과를 왜곡할 위험이 있음.
- 경향점수매칭은 관찰 가능한 특성을 통해 실험집단에 할당될 조건부 확률(경향점수)을 산출하고 경향점수를 활용해서 실험집단과 통제집단을 유사하게 구성한 후 두 집단의 결과변수를 비교함으로써 정책의 효과를 분석하는 방법임(Caliendo & Kopeinig, 2008). 이런 접근방식은 사전 특성이 비슷한 실험집단과 통제집단을 선별하여 비교함으로써 무작위 배정(Random Assignment)이 어려운 상황에서도 인과추론을 가능하게 한다는 점에서 의미가 있음. 또한, 경향점수매칭은 관찰 가능한 특성을 활용해서 선택편향(selection bias)을 통제할 수 있다는 이점이 있음.
- 경향점수매칭을 적용하는 데 있어서 몇 가지 중요한 가정에 관한 검토가 필수적임(Caliendo & Kopeinig, 2008). 우선, 처치 여부는 관찰 가능한 변수로 설명이 가능하며 비관측 요인의 영향은 존재하지 않는다는 강한 독립성(strong ignorability)을 가정함. 즉, AIDT 사용 여부는 종단연구로 수집된 관찰 가능한 변수로 온전히 설명되며, 관찰하지 못한 다른 요인들에 의해 영향을 받지 않는다는 가정임. 다음으로, 실험집단과 통제집단이 동일한 범위의 경향점수를 가져야 한다는 공통 지원 영역(common support)을 가정함. 이에 따르면, AIDT 사용 집단과 미사용 집단에서 유사한 특성을 가지는 관측치를 매칭할 때 동일한 범위의 경향점수를 가지는 관측치를 중심으로 매칭이 이루어져야 함.
- 아래 그림은 경향점수매칭을 통해 AIDT 효과를 추정하는 접근 방식을 모형도로 나타낸 것임. 그림의 왼쪽은 경향점수매칭이 이루어지기 전이며 이 경우 실험집단(AIDT 사용 집단)과 통제집단(AIDT 미사용 집단) 사이에는

처치를 받을 확률(경향점수) 분포에서 불균형을 보임. 이 상태에서는 단순 비교를 통해 AIDT 효과를 추정할 경우 편향이 발생할 수 있음. 이를 해결하기 위해 경향점수매칭을 적용하면 실험집단과 통제집단 중 경향점수가 유사한 관측치만을 선별하여 짝을 짓게 됨. 그 결과 그림의 오른쪽과 같이 매칭 후에 실험집단과 통제집단 모두 경향점수 분포가 유사해짐. 이러한 매칭 과정을 통해 두 집단은 관측된 공변량에 대해 통계적으로 유사한 특성을 가지게 되며, 이후 처치 효과를 분석할 때 보다 신뢰할 수 있는 인과 추론이 가능해짐.



- (조사 설계) AIDT 효과 분석을 위한 경향점수매칭(PSM) 기반 준실험설계는 다음과 같은 흐름으로 구성됨. 먼저, 데이터 수집 단계에서는 AIDT 처치 여부를 확인하고, AIDT 처치가 이루어지기 전의 사전 특성 정보(예: 성별, 학년, 사회경제적 지위, 기초 학업성취도, 학교 특성 등)를 수집함. 또한, 결과변수로는 학업성취도, 자기주도역량, 협력역량 등을 활용하기 위해 관련한 정보를 수집해야 함.
- 다음으로 경향점수 추정 단계에서는 로지스틱 회귀(logistic regression)나 프로빗 회귀(probit regression) 등을 통해 AIDT 사용에 대한 조건부 확률(Propensity Score)을 계산함. 이때 예측변수로는 반드시 처치 도입 이전의 학생 및 학교 특성에 관한 정보만을 사용함.

- 이어서, 매칭 단계에서는 실험집단과 통제집단 각각에서 동일한 범위 내의 경향점수를 가지는 관측치를 중심으로 매칭을 실시함. 매칭 방법으로는 1:1 매칭(Nearest Neighbor Matching), 칼리퍼 매칭(Caliper Matching), 커널 매칭(Kernel Matching) 등을 활용 가능함. 매칭 이후에는 사전 특성 차이를 비교하여, 실험집단과 통제집단 간 특성 동질화 수준을 점검함으로써 매칭의 적절성과 품질을 검증함.

<표 11> 경향점수매칭(PSM)을 활용한 준실험설계(안)

설계 단계	설계 내용
데이터 수집	<ul style="list-style-type: none"> ▪ AIDT 처치 여부 ▪ 사전 특성(성별, 학년, 사회경제적지위, 기초 학업성취도, 학교 특성 등) ▪ 결과변수(성취도, 자기주도역량, 협력역량 등)
경향점수 추정	<ul style="list-style-type: none"> ▪ 로지스틱 회귀, 프로빗 회귀 등을 통해 AIDT 처치의 조건부 확률(Propensity Score) 계산 ▪ 예측변수는 반드시 처치 이전 정보만 사용
매칭	<ul style="list-style-type: none"> ▪ 실험집단과 통제집단에서 동일한 범위의 경향점수를 가지는 관측치를 중심으로 매칭 ▪ 1:1 매칭(Nearest Neighbor Matching), 칼리퍼 매칭(Caliper Matching), 커널 매칭(Kernel Matching) 등 활용 ▪ 매칭 전후 사전 특성 차이 비교로 실험집단과 통제집단 간 동질화 정도 점검
정책효과 추정	<ul style="list-style-type: none"> ▪ 매칭된 표본을 사용하여 AIDT 도입 전후 결과 비교 ▪ AIDT 연구학교와 일반학교에서의 AIDT 효과 비교 분석 가능 ▪ 단순 평균 비교, 회귀분석, 이중차분법(DID) 연계 분석 등 가능 ▪ 강건성 분석(Robustness test) 수행

- 매칭 이후, 정책효과 추정 단계에서는 매칭된 표본을 사용하여 AIDT 도입 전후 결과를 비교함. 또한, AIDT 연구학교나 일반학교에서 AIDT 효과가 어떻게 달라지는지를 분석하여 AIDT 효과의 이질성(Heterogeneous effect)을 분석할 수 있음. 특히, 경향점수를 활용해서 매칭한 집단을 대상으로 결과 변수의 평균을 비교할 뿐만 아니라, 회귀분석이나 이중차분법(Difference-in-Differences, DID)과 연계하여 보다 심도 있는 AIDT 효과 분석이 가능함. 추가적으로 강건성 분석(Robustness Test)을 수행하여 결과의 신뢰성을 확보할 예정임.

- (효과 분석 방법) 경향점수매칭(PSM)을 기반으로 준실험설계를 적용할 경우 경향점수 산출 이후 PSM과 DID의 결합(Smith & Todd, 2005), 역확률가중치(Inverse Probability Weighting)를 활용한 회귀모형(Hirano & Imbens, 2001), 이중강건추정법(doubly robust estimation)(Funk et al., 2011) 등 몇 가지 분석 방법을 통해 AIDT의 교육적 효과를 분석할 수 있음.
- 첫째, PSM과 DID의 결합하는 방식은 PSM을 통해 실험집단과 통제집단별로 비슷한 사전 특성을 가진 관측치를 선별하여 비교가능성을 높인 후 DID를 적용하여 AIDT 도입 전후 시간 변화량을 비교하여 AIDT의 인과 효과를 추정함(Smith & Todd, 2005). 사전 특성을 중심으로 실험집단과 통제집단의 유사성을 높여 선택 편의(selection bias)를 줄이고 시간 변화로 인한 영향력도 통제할 수 있다는 이점이 있음. 이를 위해 $e(X) = P(T=1|X)$ 와 같은 수식을 통해 처치 받을 확률(경향점수)을 추정하고 동일한 범위의 경향점수를 가진 관측치를 매칭한 후, 매칭된 관측치를 대상으로 다음과 같은 DID 모형을 추정함.
- $$Y_{it} = \alpha + \beta_1 Treat_i + \beta_2 Post_t + \beta_3 (Treat_i \times Post_t) + X_{it}\gamma + e_{it},$$
 여기서 Y_{it} 는 학생의 학업성취도, 자기주도성, 협력적 소통 역량, 창의적 사고 역량, 공동체 역량 등 학생역량을, $Treat_i$ 는 실험집단 여부를, $Post_t$ 는 처치 이후 시기를, X_{it} 는 학생 및 학교 특성 벡터를, e_{it} 는 오차항을 의미함. β_3 은 이 연구에서 관심을 가지는 AIDT 효과에 관한 추정치라고 볼 수 있음.
- 둘째, 역확률가중치(Inverse Probability Weighting)를 활용한 회귀모형을 통해 AIDT 효과를 추정할 수 있음. 역확률가중치는 각 관측치가 처치 받을 확률(경향점수)을 산출한 후 해당 확률의 역수 값을 가중치로 부여하여 처치변수와 결과변수에 대한 회귀분석을 수행하는 방식임(Hirano & Imbens, 2001). 이를 AIDT 효과 분석에 적용할 경우 실험집단은 기준연도에 AIDT를 도입한 학교 또는 학급, 통제집단은 AIDT를 도입하지 않은 학교 또는 학급으로 구분됨. AIDT 도입 이전의 학생에게서 나타난 특징, 예컨대 성별, 학년, 사회경제적지위, 사전 학업성취도 등을 활용하여 경향점수를 추정함. 관측치의 경향점수를 활용

해서 역확률가중치를 산출함. 산출한 역확률가중치를 적용해서 $Y_i = \alpha + \tau T_i + \epsilon_i$ 라는 회귀모형을 추정함. 여기서 Y_i 는 결과변수를, T_i 는 AIDT 도입 여부를 각각 의미함. 연구의 관심 회귀계수는 AIDT 정책 효과를 의미하는 τ 임.

- 셋째, 경향점수모형과 결과모형을 동시에 사용하는 이중강건추정법(doubly robust estimation)을 통해 AIDT를 추정하는 방법도 가능함. 이중강건추정법은 경향점수모형과 결과모형 중 하나만 올바르게 지정될 경우 효과 추정치가 편향되지 않는다는 강력한 특성이 있음(Funk et al., 2011). 여기서 경향점수모형은 로지스틱 회귀모형이나 프로빗 회귀모형을 통해 AIDT 도입 여부를 예측하는 경향점수를 추정하는 모형을, 결과모형은 실험집단과 통제집단별로 결과변수를 예측하는 모형을 각각 의미함. 두 모형에서 추정한 값들을 활용해서 다음과 같은 이중강건추정량을 산출함.

$$-\hat{\tau}_{DR} = \frac{1}{n} \sum_{i=1}^n \left[\frac{W_i(Y_i - \hat{\mu}_1(X_i))}{\hat{e}(X_i)} + \hat{\mu}_1(X_i) - \frac{(1 - W_i)(Y_i - \hat{\mu}_0(X_i))}{1 - \hat{e}(X_i)} - \hat{\mu}_0(X_i) \right],$$

여기서 W_i 는 처치 여부를, Y_i 는 결과변수를, X_i 는 공변량 벡터를, $\hat{e}(X_i)$ 는 경향점수를, $\hat{\mu}_1(X_i)$ 는 실험집단의 조건부 기댓값, $\hat{\mu}_0(X_i)$ 는 통제집단의 조건부 기댓값을 각각 의미함. 이렇게 이중강건추정법을 적용할 경우, 모형 설정의 오류(model misspecification)로부터 비교적 자유롭다는 이점이 있음. 경향점수모형과 결과모형과 같은 두 개 모델을 결합해, 둘 중 하나의 모델만 정확하면 편향되지 않은 추정치를 산출 가능함.

- (필요 변수) 경향점수매칭 방식을 활용해서 AIDT 효과를 분석하기 위해서는 다음과 같은 몇 가지 변수의 정보가 필요함.
 - 첫째, 가장 기본적으로 AIDT 처치 여부 정보가 필요함. 이는 각 학생이나 학교 단위에서 AIDT를 실제로 도입했는지 여부를 보여주는 변수여야 함.
 - 둘째, AIDT의 성과로 볼 수 있는 결과변수를 설정하고 이에 관한 정보를 수집해야 함. AIDT의 교육적 효과를 평가하기 위해서는 학업성취도

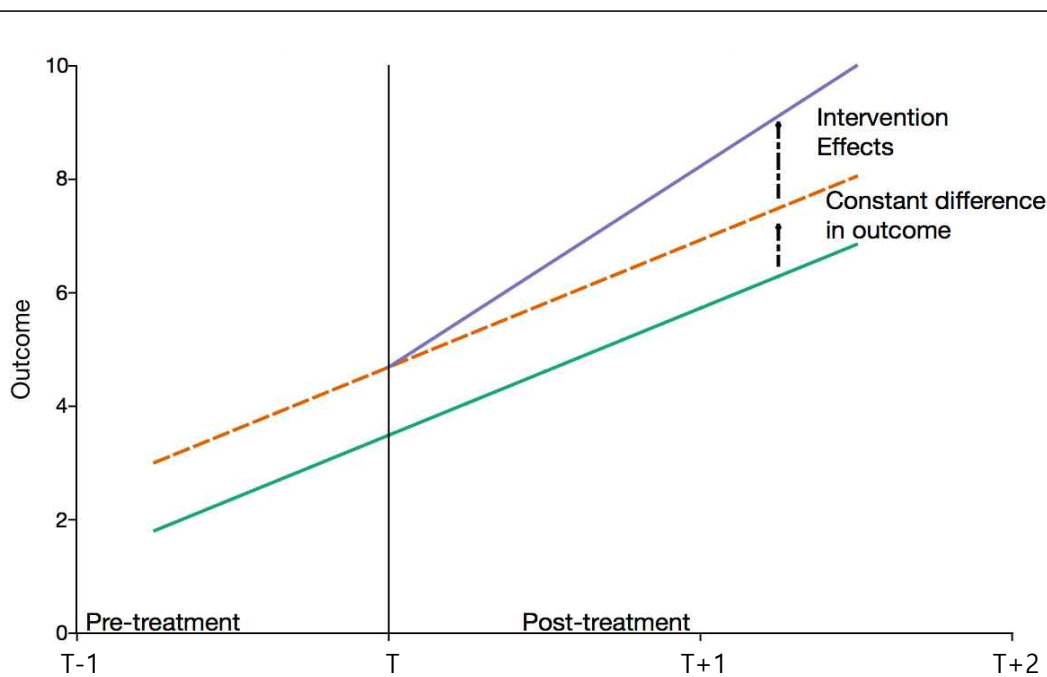
점수, 자기주도학습 역량, 협력적 소통 역량 등 다양한 지표를 결과변수로 포함할 필요가 있음. 이러한 결과변수는 AIDT 도입 이전(pre-treatment)과 이후(post-treatment) 모두에서 조사되어야 하며, 최대한 동일하거나 유사한 방식으로 측정되어야 사전 및 사후 비교가 가능함.

- 셋째, 사전 특성 변수(Pre-treatment Covariates)에 관한 정보를 수집해야 함. 이는 처치와 결과 모두에 영향을 줄 수 있는 학생이나 학교의 기본 특성들을 의미함. 예컨대, 성별, 학년, 가정의 사회경제적 지위(SES), 초기 학업성취도 수준, 학교 규모, 지역 특성(도시/농어촌 여부) 등과 같은 변수들을 고려해 볼 수 있음. 이러한 변수들은 경향점수 추정이나 회귀 분석 과정에서 필수적으로 사용된다는 점에서 정확한 수집이 요구됨.
- 넷째, 관측 시점이 명확히 구분되어야 함. AIDT 도입 시기(ex. 연도 및 학기)에 관한 정보 역시 함께 수집되어야 AIDT 도입 전후에 나타나는 주요 변수들의 변화를 측정할 수 있음. 아울러 다수의 관측 시점(예: 도입 1년 전, 도입 직후, 도입 1년 후 등)에서 관련 정보를 확보할 경우 시간에 따른 변화(dynamics)도 세밀하게 분석할 수 있을 것임.

○ (준실험설계 모형) 이중차분법(Difference-in-Differences, DID) 기반 준실험설계

- (개념) 실험집단과 통제집단 간 AIDT 도입 전후의 결과 변화를 비교하는 이중차분모형을 활용하여 AIDT 활용의 교육적 효과를 추정함. 이중차분모형은 정책 도입 전후의 결과 변화를 실험집단과 통제집단 간에 비교함으로써 순수한 정책 효과를 추정하는 방법임(Angrist & Pischke, 2009). 즉, 실험집단의 변화량에서 통제집단의 변화량을 차감하여 정책의 인과적 효과를 식별함. AIDT 효과성 분석에 이중차분모형을 적용하면, AIDT를 사용한 실험집단의 성과 변화량에서 통제집단의 성과 변화량을 차감하여 AIDT 활용의 교육적 효과를 추정하게 됨.
- DID를 적용하는 데 있어서 몇 가지 중요한 가정에 관한 검토가 필수적임

(Angrist & Pischke, 2009). 우선, AIDT 도입이 없었더라도 실험집단과 통제 집단이 비슷한 변화 추세를 보였을 것이라는 평행추세(parallel trends) 가정을 전제로 함. 다음으로, 분석 기간 동안 표본의 이탈(sample attrition)이 없어야 하며, 분석 시점에 외부로부터 발생하는 또다른 정책 개입이 존재하지 않아야 함. 아울러, AIDT 처치 여부는 결과변수와 독립적이어야 한다는 외생성의 가정이 충족되어야 함.



[그림 17] 이중차분법(DID)을 활용한 AIDT 효과 분석 모형도

- DID는 정책 처치가 결과 변수에 미치는 영향을 분석하는 기법으로, 실험집단과 통제집단의 시간에 따른 변화를 비교하여 순수한 개입 효과를 추정함. 아래의 그림은 DID를 활용해서 AIDT 효과를 분석하는 개념도임. 이 그림에서 수평축은 시간, 수직축은 결과를 나타냄. 아래의 녹색 선은 통제집단의 결과변수 값을, 위의 보라색 선은 실험집단의 결과변수 값을 의미하며, 점선은 처치가 없었을 경우 실험집단에서 나타날 것으로 예상되는 결과변수 값을 나타냄. DID는 AIDT 도입 전 두 집단의 추세가 평행(parallel trend assumption)하다고 가정함. 이후 실험집단과 통제집단의 변화 폭 차이를 계산하여 AIDT의 효과를 산출함. 즉, ‘실험집단의 사후-사전 결과변

수 변화 값’에서 ‘통제집단의 사후-사전 결과변수 변화 값’을 뺀 값으로 산출함. 그림의 ‘Intervention Effects’가 바로 이 DID 추정치에 해당하는데, 이를 통해 AIDT 효과를 통계적으로 식별할 수 있음.

- (조사 설계) 조사설계의 핵심은 AI 디지털교과서의 도입 시기가 지역 및 학교마다 다르다는 사실임. 이 연구는 이를 고려해서 이중차분모형을 활용한 준실험설계를 적용함. AIDT 효과 분석을 위한 구체적 설계는 다음과 같이 구성 가능함. 실험집단은 AIDT를 도입한 교육청이나 AIDT를 실제 수업에 활용한 학교나 학급으로 설정하고, 통제집단은 종전과 동일하게 서책형 교과서를 활용하는 집단으로 설정함. 시간적으로는 AIDT 도입 전(Pre-treatment)과 후(Post-treatment)를 구분하여 데이터를 구축함. 이후, 실험집단과 통제집단의 사전 특성을 비교하여 집단 간 동질성을 점검하고, 평행추세 가정을 검토한 후 DID 모형을 적용하여 정책 효과를 추정함. 분석 후에는 위약 처치(placebo test)나 통제집단합성법(Synthetic Control Model, SCM) 등을 통해 견고성(robustness)을 검증함(Abadie, 2021; Kennedy-Shaffer, 2024).
- 실제 교육감 또는 교장의 재량, ICT 인프라 차이, 교사의 선호 등으로 인해 지역, 학교, 학급마다 AIDT 도입 및 사용 시기에 차이가 있음. 지역 및 학교, 학급마다 AIDT 도입 및 사용 시기가 다른 상황에서 AIDT 도입 시기와 관찰 시점에 따라 다음과 같은 몇 가지 접근 방식이 가능함.
 - 첫째, 2개 시점의 경우 <표 12>와 같은 방식으로 준실험설계가 가능함. 이에 따르면, AIDT를 사용한 실험집단과 사용하지 않는 통제집단을 구분하고, 2025년 AIDT 도입 전후 나타난 변화를 비교하는 방식으로 AIDT의 효과를 분석할 수 있음. DID에서는 무작위 추출과 무작위 배정으로 선정된 집단 구분이 아니라는 점에서 실험설계와 차이가 있음. 또한 AIDT의 도입 여부가 영향평가의 목적이 아닌, 교육청, 학교, 교사 등의 재량에 의해 결정된 사항이라는 점에서도 두 방식에는 분명한 차이가 존재함(Hull & Dutch, 2019).

<표 12> 이중차분법(DID)의 성과 측정 방법(2개 시점)

구 분	Before (AIDT 도입 전 성과측정)	After (AIDT 도입 후 성과측정)
실험집단(AIDT 사용)	A	B
통제집단(AIDT 미사용)	C	D

※ 성과(Impact) = (B-A)-(D-C)

- 둘째, 3개 시점 이상일 경우 교육청, 학교, 학급이 AIDT 사용에 재량권을 가지고 AIDT 도입 시기를 결정함에 따라 2년 이상 AIDT를 사용한 교육청, 학교, 학급(실험집단 A), 1년 동안 AIDT를 사용한 교육청, 학교, 학급(실험집단 B, 실험집단 C), AIDT를 전혀 사용하지 않은 교육청, 학교, 학급(통제집단) 등으로 집단을 구분할 수 있으며, 이를 통해 AIDT 사용의 효과와 함께 AIDT 사용 연한에 따른 강도 효과도 추정이 가능함.

<표 13> 이중차분법(DID)의 성과 측정 방법(3개 시점 이상)

구분	연구 기간		
	시점 1	시점 2	시점 3
실험집단 A	AIDT 미사용	AIDT 사용	AIDT 사용
실험집단 B	AIDT 미사용	AIDT 미사용	AIDT 사용
실험집단 C	AIDT 미사용	AIDT 사용	AIDT 미사용
통제집단	AIDT 미사용	AIDT 미사용	AIDT 미사용

- 이 연구에서 적용하는 실험설계방식은 동일한 학교 안에서 학급을 단위로 삼아 처치를 적용하는 방식이나, 여기서 제시한 DID는 AIDT 도입 및 사용에 관한 재량권을 가진 교육청, 학교, 교사 단위로 처치가 이루어진다는 점에서 학교 단위에서도 AIDT 도입에 따른 효과 분석이 가능함. 이런 이유에서 학교 단위에서 AIDT 효과를 분석할 경우 실험설계 방식보다는 준실험/비실험 설계가 적절함.

○ (효과 분석 방법) 우선, 2개 시점을 활용할 경우 다음과 같은 분석모형을 활용해서 AIDT의 교육적 효과를 분석함.

$$- Y_{it} = \alpha + \beta_1 Treat_i + \beta_2 Post_t + \beta_3 (Treat_i \times Post_t) + X_{it}\gamma + e_{it},$$

여기서 β_3 은 이 연구에서 관심을 가지는 AIDT 효과에 관한 추정치임. 여기서 Y_{it} 는 학생의 학업성취도, 자기주도성, 협력적 소통 역량, 창의적 사고 역량, 공동체 역량 등 학생역량을, $Treat_i$ 는 실험집단 여부를, $Post_t$ 는 처치 이후 시기를, X_{it} 는 학생 및 학교 특성 벡터를, e_{it} 는 오차항을 의미함.

- 다음으로 3개 시점 이상의 분석자료를 활용하여 AIDT의 효과를 분석할 경우 일반화된 이중차분법(Generalized Difference-in-Differences, GDID)를 통해 AIDT의 교육적 효과를 분석할 수 있음. 일반화된 이중차분법은 다수의 처치 시점(staggered adoption)이나 다수의 집단(multiple groups)이 존재하는 상황에서 정책의 인과적 효과를 추정하는 데 활용 가능한 분석 방법임(Callaway & Sant' Anna, 2021). 3개 시점의 조사를 통해 AIDT 사용 여부에 관한 정보를 수집할 경우 학교, 지역별로 처치 시점이 다르게 진행되는 경우가 존재함. 학교 혹은 지역마다 AIDT 도입 시점일 다를 수 있기 때문임. 이럴 경우 GDID를 통해 다양한 집단 차이를 고려해서 AIDT의 순차적 도입에 따른 인과적 효과를 추정할 수 있음.

- 집단별, 시점별로 처치효과(τ)가 달라질 수 있다는 점을 고려해서 다음과 같은 함수식($Y_{it} = \alpha_i + \lambda_t + \tau_{g(i)t}D_{it} + \epsilon_{it}$)을 활용해서 AIDT 사용의 교육적 효과를 추정함. Y_{it} 는 학생의 학업성취도, 자기주도성, 협력적 소통 역량, 창의적 사고 역량, 공동체 역량 등 학생역량을, α_i 는 개체 고정효과를, λ_t 는 시점별 고정효과를, ϵ_{it} 는 오차항을 각각 의미함. 여기서 t 시점 집단 $g(i)$ 의 처치 효과를 의미하는 $\tau_{g(i)t}$ 가 집단별, 시점별 AIDT의 처치 효과를 보여주는 추정치임.

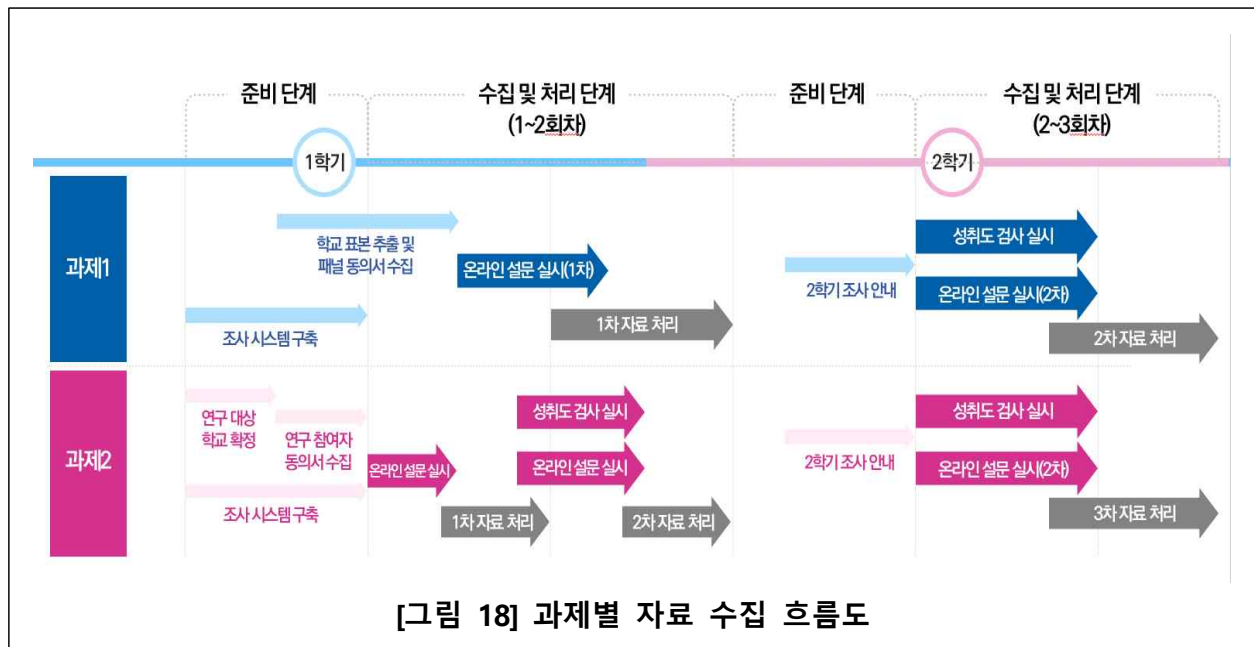
- (필요 변수) AIDT 효과를 분석하기 위해서는 다음과 같은 몇 가지 변수의 정보가 필요함. 우선, 각 교사(학급) 단위로 AIDT 처치 여부에 대한 정보가 있어야 함. 조사 시점에 각 학교가 AIDT를 활용한 수업을 제공하였는지에

관한 정확한 정보가 필요함. 또한, 결과변수로 학업성취도 점수(영어, 수학 등), 자기주도학습 역량, 협력적 소통 역량 등과 같은 학생 역량에 관한 정보 역시 필요함. 이 결과변수들은 AIDT 도입 전과 후 모두 수집되어야 하며, 학생과 학교의 기초적 특성(성별, 학년, SES, 학교 규모, 지역 등)에 관한 정보도 마련될 필요가 있음.

나. 학생, 교사, 학부모, 학교 대상 조사 시행

1) ‘[과제 1] 디지털 교육 정책 성과 진단’ 과의 차이점

- ‘[과제 2] 2025년 디지털교육 효과 분석’의 경우, 조사 홈페이지 구축부터 온라인 설문 개발, 조사 자료 검증 및 처리까지 [과제 1] 조사와 대부분 유사한 형태로 진행됨(목차 3-1. 연구의 내용 및 범위: [과제 1] 디지털 교육 정책 성과 진단 중 항목 ‘라’ 참고). 다만 조사 횟수 및 시기에 양 과제(조사) 간 차이가 있으며, 대략적인 과제별 자료 수집 프로세스는 아래와 같이 예상됨.



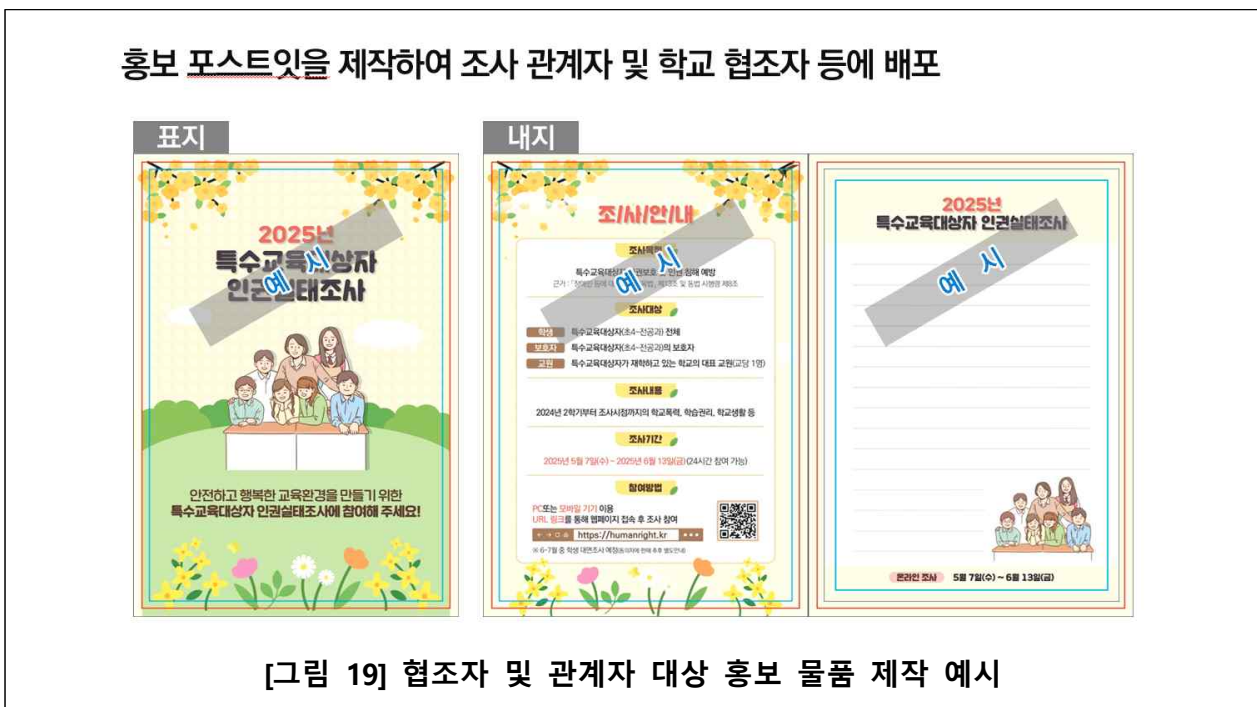
- [과제 2]의 경우, 조사 횟수가 시기별 최대 3차수로 긴박한 일정으로 운영을 요하기 때문에, 성취도 검사는 [과제 1]에서 개발된 성취도 검사 도구 사용하고자 하며, 사전 성취도 검사는 맞춤형 학업성취도 자율평가(KICE)를 활용할 예정임. 이 외에도 각 조사의 시점 및 세부 집행 내용에 대한 협의가 필요할 것으로 예상됨.

2) 데이터 연계 및 데이터클리닝, 온라인 조사 실시

- 성취도검사 및 설문조사 등 수집 데이터*의 데이터 연계 및 데이터클리닝 실시.
- AIDT 로그데이터 중 활용 가능한 데이터 수집 방안 및 데이터클리닝, 데이터 제공·분석 방안 제시.
 - AIDT를 도입한 이후에 사용 빈도가 낮거나 학생의 주도적인 참여가 부족할 경우에 긍정적인 효과를 기대하기 어려운 반면에, AIDT 로그데이터에서 능동적이고 상호작용적인 활동이 규칙적으로 나타날 경우 학습 효과가 크게 향상될 것으로 기대됨.
 - 연구에 동의한 학생을 대상으로 KERIS를 통해 AI 디지털 교과서에서 생성되는 로그데이터를 제공받음. KERIS에서 제공받는 학습 활동 데이터를 수집함. 민간 AIDT 개발사에서 일괄 처리 데이터로 처리하여 AIDT 포털로 전송한 xAPI 포맷의 학습 활동 데이터를 미디어, 평가, 과제, 경로, 질의, 교수활동 등의 영역에서 수집함. 학생 데이터는 개인식별정보가 노출되지 않도록 익명화된 고유 ID를 부여해서 관리함.
 - AIDT 로그데이터를 전처리한 다음 수업 중 AIDT가 얼마나 자주 사용되었는지, 어떤 기능을 주로 사용하였는지, 학습자가 AIDT와 어떻게 상호작용하는지 등에 대한 정보를 분석함. 그리고 AIDT 로그데이터는 설문조사 및 학업성취도 데이터와 연계하여 AIDT의 활용 양상이 학생의 학업성취, 핵심역량, 디지털 리터러시 등에 어떠한 영향을 미쳤는지 분석함. AIDT 로그데이터와 설문조사, 학업성취도 데이터를 서로 연결하고 가명처리하는 것은 KERIS의 협조를 받아 실시함.

3) 연구 참여 유인 및 연구 대상 관리 방안 제시

- 조사 횟수, 유형의 증가 등으로 인해, [과제 1] 대비 협조 관계자 및 응답자에게 보다 과중한 부담이 주어지는 점을 고려하여 연구 참여 유인 방안을 마련할 필요가 있음.
- [과제2] 조사 대상 중 일반학교의 경우, 연구학교 대비하여 본 연구의 참여 유인 요소가 상대적으로 떨어질 수 있음. 이를 고려하여 일반학교 대상으로 조사 협조 감사 및 진행 지원을 위한 일정 금액에 상당하는 학교별 조사 물품 제공을 고려할 수 있으며, 조사 인지도 제고를 위해 홍보 물품을 제작 및 배포하는 방법도 고려할 수 있음.



- 학생 및 학부모 대상으로는 1학기 조사 이후, 2학기 조사의 연속성 유지를 위해 방학 기간 또는 2학기 개학 시점에 경품 이벤트를 개최하여 조사를 안내하는 방법을 고려할 수 있음. 또한, 조사 응답에 대한 답례를 제공하는 방법에 있어서도 2학기 마지막 조사까지 모두 참여했을 때, 한 번에 제공하는 방법과 하나의 조사를 완료할 때마다 답례를 지급하는 방법 등으로 다양한 차원적 모색도 가능함.

조사 응답율 제고 방안

학기별로 실시되는 조사 특성을 감안하여
조사 응답 연속성 유지를 위해
 응답자 답례 추가 및 과거 답례 관리

- ☑ 경품이벤트 진행을 통해
 참여도 제고할 수 있음
- ☑ 지난 조사 답례 미사용자 확인하여
 답례 재발송을 통해 조사 환기



[그림 20] 응답자 대상 경품이벤트 예시

다. 조사결과 분석을 통한 디지털교육 효과성 분석

1) AI 디지털교과서 사용 효과 분석

○ AIDT 사용이 학생성과에 미치는 효과 분석

- 기초통계 분석: 실험집단과 통제집단 간 학업성취도 및 문해력, 비인지적 특성(사회정서역량, 학생과 교사의 상호작용 등) 평균 차이 검정, AIDT 연구학교와 일반학교 간 평균 차이 검정 등 기초통계 분석 실시.
- 회귀분석 및 다층분석: AIDT 시행의 효과를 추정하기 위해 학생·교사·학교 특성을 통제한 다중회귀분석 및 다층분석을 실시함.

$$Y_{ic} = \beta_0 + \beta_1 R_c + \beta_2 X_{ic} + \epsilon_{ic},$$

여기서 Y_{ic} 는 학급 c 가 AIDT 사용 학급으로 배정되었을 때 학생 i 의 성과, R_c 는 학급 c 가 AIDT 사용 학급으로 배정되었는지 여부, X_{ic} 는 학생의 성과에 영향을 미칠 수 있는 학생·학급·학교의 특성 벡터를 나타냄.

- 하위집단 분석: AIDT 사용의 효과가 학생 및 학교의 사회경제적 배경에 따라 어떻게 달라지는지 분석하기 위해 하위집단 분석을 실시함.

$$Y_{ic} = \beta_0 + \beta_1 R_c + \beta_2 W_{ic} + \beta_3 (R_c \times W_{ic}) + \beta_4 X_{ic} + \epsilon_{ic},$$

여기서 W_{ic} 는 학생 및 학교의 사회경제적 배경 특성을 나타내며, AIDT 사용 여부와의 상호작용을 통해 AIDT 사용이 학생의 성과에 미치는 영향이 사회경제적 배경에 따라 어떻게 차별화되는지 분석할 수 있음.

- 도구변수 분석: 실험 진행 과정에서 중도 이탈, 비순응, AIDT 활용 수준 차이 등이 발생하는 경우 AIDT 학급 배정과 실제 AIDT 활용 간에 불일치가 나타날 수 있음.

- 이 경우 다음과 같은 2단계 최소제곱법(Two-Stage Least Squares, 2SLS)을 적용하여 ‘실제 처치를 받은 집단에 대한 처치 (Treatment-on-the-Treated, TOT) 효과’를 추정하고자 함. 이때 2SLS 분석의 도구변수로는 학급의 무작위 배정 여부를 활용함.

$$1\text{단계 모형: } AIDT_c = \pi_0 + \pi_1 R_c + \pi_2 X_c + \nu_c ,$$

여기서 $AIDT_c$ 는 학급 c 에서 AIDT를 실제 활용했는지 여부를, R_c 는 학급 c 가 무작위 배정을 통해 AIDT 활용 학급으로 할당되었는지 여부를, X_c 는 학급 특성 벡터를, ν_c 는 오차항을 의미함. 2단계 모형에서는 1단계 모형에서 추정한 \widehat{AIDT}_c 값을 활용하여, AIDT 활용 수업의 학생성과에 대한 순수 효과를 추정함.

$$2\text{단계 모형: } Y_{ic} = \beta_0 + \beta_1 \widehat{AIDT}_c + \beta_2 X_{ic} + \epsilon_{ic} ,$$

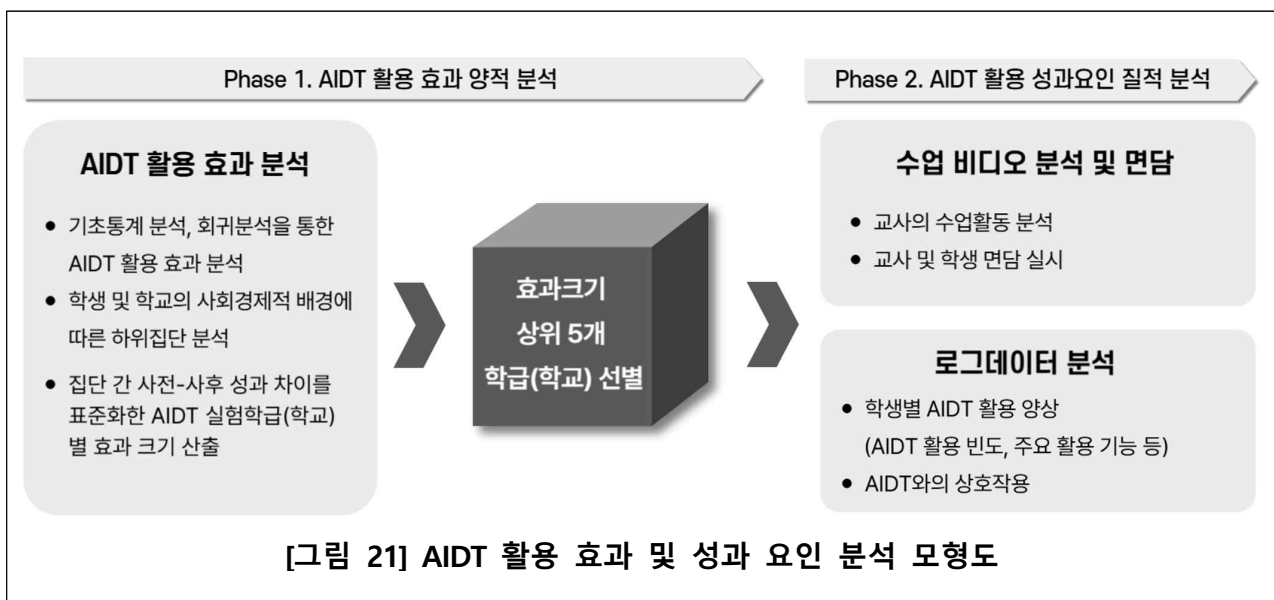
여기서 Y_{ic} 는 학급 c 에 속한 학생 i 의 성과를, \widehat{AIDT}_c 는 1단계 모형에서 예측한 처치 확률을 의미함. 이렇게 해서 추정한 AIDT 수업의 효과 추정치(β_1)는 AIDT를 활용한 수업이 학생의 성과에 미치는 인과적 효과를 나타냄. 본 연구에서는 주로 ‘처치-의도(Intent-to-treat, ITT) 효과’를 분석하나, TOT 효과를 분석함으로써 처치의 과소 추정 문제를 보완하고자 함.

2) AI 디지털교과서 활용 성과 요인 분석

- 전문적 학습 공동체 실험 효과 분석을 위한 실험 설계가 불가능할 경우 대안으로 질적 분석을 병행하여 혼합 연구를 실시함.
- 혼합 연구를 위하여 AI 디지털교과서 사용 효과크기 상위 5개 학교(학급)를

대상으로 질적 분석을 실시하여 AIDT 활용의 주요 성과 요인을 탐색함.

- 먼저, 선정된 학교(학급)의 AIDT 활용 수업을 촬영한 영상자료를 PICRAT 모형에 따라 분석함. 두 명의 연구자가 ATLAS.ti 소프트웨어를 이용하여 독립적으로 자료를 분석하고, 그 결과를 비교하여 평정자 간 신뢰도를 산출하며, 불일치 항목은 연구자 간 논의를 통해 최종 결정함.
- 다음으로, 해당 학교(학급)의 교사 및 학생을 대상으로 심층 면담을 실시하여 AIDT의 효과, 교사의 수업 설계 방식, 학생과 AIDT의 상호작용 양상 및 경험 등을 분석함.
- 마지막으로, AIDT 로그데이터 분석을 통해 학습자의 AIDT 활용 빈도, 주요 활용 기능 등 학습자와 AIDT 간 상호작용 특성을 질적으로 분석함.



3) 교사의 전문적 학습 공동체 참여 효과 분석

- 전문적 학습 공동체 참여에 따른 집단 간 차이 분석
 - 실험집단과 통제집단 간 활동 패턴의 차이를 분석하기 위해 ENA(Epistemic Network Analysis)를 실시함. ENA를 통해 교수학습 활동 간 연결성을 분석하고 실험집단과 통제집단의 차이뿐만 아니라 각 집단에 있어 사전검사와 사후검사의 차이를 시각화하여 비교함.
 - 교사와의 면담 내용에 대한 주제 분석을 통해 AIDT 전문적 학습 공동체

활동이 수업 개선에 어떤 영향을 미치는지와 학생의 역량 향상에 어떤 기여를 하는지 질적으로 분석함.

○ 교사의 전문적 학습 공동체 참여에 따른 AIDT 사용의 차별적 효과 분석

- ‘AI 디지털교과서 사용 효과 분석을 위한 실험’ 연구에서 수집된 학업성취도 검사와 설문조사 결과를 사전검사(7월 시행)와 사후검사(11-12월 시행) 자료로 활용하여 교사의 AIDT 전문적 학습 공동체 참여가 학생에게 미치는 영향을 통계적으로 분석함.
- 교사의 전문적 학습 공동체 참여 여부에 따른 AIDT 사용 효과의 차이를 확인하기 위해 하위집단 분석(subgroup analysis)을 실시함.

$$Y_{ic} = \beta_0 + \beta_1 R_c + \beta_2 PLC_{jc} + \beta_3 (R_c \times PLC_{jc}) + \beta_3 X_{ic} + \epsilon_{ic},$$

여기서 PLC_{jc} 는 c 학급 교사 j 의 전문적 학습 공동체 참여 여부를 나타내며, c 학급의 AIDT 사용 여부(R_c)와의 상호작용을 통해 AIDT 사용이 학생의 성과에 미치는 영향이 교사의 전문적 학습 공동체 참여 여부에 따라 어떻게 차별화되는지 분석할 수 있음.

- 도구변수 분석: 실험 진행 과정에서 교사의 중도 이탈, 비순응, 전문적 학습 공동체 참여 수준 등으로 인해 무작위 배정을 통한 교사의 전문적 학습 공동체 할당 여부와 실제 참여 간에 불일치가 발생할 수 있음. 이 경우 다음과 같은 2단계 최소제곱법(Two-Stage Least Squares, 2SLS)을 적용하여 ‘실제 처치를 받은 집단에 대한 처치(Treatment-on-the-Treated, TOT) 효과’를 추정하고자 함.

$$1\text{단계 모형: } PLC_j = \pi_0 + \pi_1 R_j + \pi_2 X_j + \nu_j,$$

여기서 PLC_j 는 교사 j 가 전문적 학습 공동체에 실제 참여했는지 여부를, R_j 는 교사 j 가 무작위 배정을 통해 전문적 학습 공동체 참여 집단에 할당되었는지 여부를, X_j 는 교사 특성 벡터를, ν_i 는 오차항을 의미함. 2단계 모형에서는 1단계 모형에서 추정한 \widehat{PLC}_j 값을 활용하여, AIDT 활용 수업의 학생성과에 대한 인과적 영향을 추정함.

$$2\text{단계 모형: } Y_{ic} = \beta_0 + \beta_1 R_c + \beta_2 \widehat{PLC}_j + \beta_3 (R_c \times \widehat{PLC}_j) + \beta_3 X_{ic} + \epsilon_{ic},$$

여기서 Y_{ic} 는 학급 c 에 속한 학생 i 의 성과를, \widehat{PLC}_j 는 1단계 모형에서 예측한 처치 확률을 의미함. 이렇게 해서 추정한 AIDT 사용 여부와 전문적 학습 공동체 참여 여부의 상호작용 효과 추정치(β_3)는 교사의 전문적 학습 공동체 참여 여부에 따른 AIDT 활용 수업이 학생의 성과에 미치는 순수 효과를 나타냄. 본 연구에서는 주로 ITT 효과를 분석하나, TOT 효과를 분석함으로써 처치의 과소추정 문제를 보완하고자 함.

라. 조사 참여학교에 대한 학교별 분석 리포트 및 학생용 성적표 제공

- 성취도 검사와 학생 역량 변인 등을 중심으로 조사 참여 학교에 대한 학교별 분석 리포트 제공.
 - 검사 개요, 문항별, 각 하위 요소별 검사 결과 분석 및 강·약점 분석과 더불어, 사전검사와 1학기 말 시점에서 실시되는 중간검사 간 결과 성장에 대한 비교 분석 정보 제공.
 - 학교별 분석 리포트는 한글 파일로 작성하되, 리포트에 담긴 분석 결과를 담은 표나 그래프는 엑셀 파일로 함께 제공.
 - 분석 리포트는 2025. 9~10월 제공(단, 연구 추진 일정 등에 따라 조정 가능).
- 조사에 참여한 학생을 대상으로, 성취도 검사와 학생 역량 변인(디지털 리터러시, 자기주도성, 협업적 소통 역량 등)에 대하여 개인별 성적표 작성하여 제공.

4. 관련 연구에 대한 내용 분석

가. 디지털 교육 효과 분석 연구

1) 디지털 교육 효과 분석 연구의 필요성

- 4차 산업혁명 이후 인공지능(AI), 빅데이터, 사물인터넷(IoT) 등 디지털 기술이 사회 전반에 확산되면서 다양한 분야에서 디지털 전환이 가속화됨. 이러

한 변화는 교육 분야에도 영향을 미쳐, 학생 중심 맞춤형 학습 실현에 대한 기대와 함께 디지털 기술에 기반한 교육이 새로운 교육 패러다임으로 부상하고 있음.

- 디지털 교육은 실시간 피드백, 적응형 콘텐츠, 에듀테크 등을 활용하는 교육 방식으로, 전통적인 수업 방식의 한계를 보완하고, 학습 기회를 확대함으로써 학습 동기와 성취도 향상에 기여할 수 있을 것으로 기대됨(문화원, 정동욱, 2025).
- 우리나라는 2025년 AI 디지털 교과서 도입을 계기로 교육의 디지털 전환을 본격적으로 추진 중이며, 교원의 디지털 교수·학습 역량 강화를 위한 연수 프로그램 운영, 디지털 역량 인증제 도입 검토 등 다양한 정책이 활발히 진행되고 있음. 또한, 맞춤형 콘텐츠 및 플랫폼 제공과 함께, 개별화 학습을 지원하기 위한 초·중등학교의 물리적 환경 개선도 병행되고 있음(교육부, 2024; 2025).
- 그러나 디지털 교육이 실제로 학생의 학업 성취, 교사의 교수·학습 변화, 교육의 형평성과 질 제고에 어떤 영향을 미치는지에 대한 실증적 분석은 아직 충분하지 않은 실정임.
- 디지털 교육의 효과를 다차원적으로 평가하기 위해서는 교육 정책의 성과를 과학적으로 검증할 수 있는 체계적 기반 마련이 필요함.
- 디지털 교육 효과 분석은 단순히 기술의 활용 여부를 평가하는 것을 넘어, 학습자 중심 교육의 실현 가능성, 교육 형평성 확보, 디지털 기반 교육 정책의 타당성에 대한 근거를 마련하는 데 중요한 역할을 하게 될 것으로 기대됨.

2) 디지털 역량 및 디지털 리터러시 관련 대규모 조사연구

가) 국내 조사 연구

① 디지털 리터러시 검사 도구 개발

- 디지털 교육이 새로운 교육 패러다임으로 자리 잡으면서, 이를 효과적으로 활용할 수 있는 디지털 리터러시에 대한 관심도 함께 증가하고 있음.
- 디지털 리터러시는 ‘디지털 사회의 구성원으로서 자율적인 삶을 영위하기

위해 필요한 기본 소양' 으로 정의됨(계보경 외, 2021). 이는 윤리적 태도를 바탕으로 디지털 기술을 이해하고 활용하여, 정보 탐색, 관리, 창작, 문제 해결까지 아우르는 실천적 역량을 포함함.

- 한국교육학술정보원(KERIS)은 2007년부터 초·중학생을 대상으로 'ICT 리터러시 검사' 를 운영해옴.
- 2018년에는 이를 '디지털 리터러시 검사' 로 개편하고, ICT와 컴퓨팅 사고력(CT)을 포함하는 이원적 구조로 재구성하여 실시함(계보경 외, 2022).
- 2023년부터 초등학교 4학년~중학교 3학년을 대상으로 수행형 평가 도구를 도입하여, 디지털 리터러시를 보다 실제적인 맥락에서 측정함.
- 해당 평가는 시나리오 기반 모듈로 구성되어 있으며, 측정 영역은 디지털 도구, 디지털 정보·데이터, 디지털 의사소통 및 협력, 디지털 자원 생산, 디지털 안전과 건강의 5가지 영역으로 구성됨.

② 교원 AI·디지털 역량 측정도구 개발

- 디지털 교육 환경에서 교사는 교육 혁신을 주도하는 핵심 주체로 여겨지며, 이에 따라 교원의 AI 및 디지털 역량 강화가 중요한 과제로 떠오르고 있음.
- 교원의 AI·디지털 역량은 'AI 및 디지털 기술에 대한 기초 지식과 교육적 의미를 바탕으로, 직무 수행에 적절히 기술을 적용할 수 있는 능력과 윤리적 태도' 로 정의됨(서정희 외, 2023).
- AIEDAP 사업팀은 2022년에 이루어진 1차년도 연구에서 교사의 교수·학습 활동에 초점을 두고 공통 역량을 도출함. 또한, 국어, 수학, 사회, 과학, 미술 등 5개 교과를 중심으로 교과별 역량 체계 및 행동 지표를 비교·분석함. 이어서 2023년 2차년도 연구에서는 초·중등, 예비·현직, 교과 구분 없이 통합된 공통 역량 및 행동 지표를 구성함.
- 이 과정에서 교원의 AI·디지털 공통 역량은 '교육적 목적을 위해 AI 및 디지털 기술을 비판적이고 창의적이며 윤리적으로 활용할 수 있는 지식, 기술, 가치·태도' 로 재정의됨(서정희 외, 2023).

③ 학교 디지털 전환 실태조사 및 지표 개발

- 디지털 기술의 발전은 교수·학습 방법뿐 아니라 교사, 학생, 학부모 간 상호작용 방식까지 변화시키며, 학교 교육의 디지털 전환을 가속화하고 있음.
- 이러한 흐름에 대응하여, 학교 현장의 디지털 전환 수준을 진단하고 정책 방향을 구체화하기 위해 다양한 연구가 추진되고 있음.
- 한국교육개발원은 2022년 ‘교육 및 데이터 환경 변화에 따른 교육 지표 개선 방안’ 연구를 통해 학교 교육 디지털 전환 지표와 학생 자기주도성 증진을 위한 환경 지표를 개발함(권희경 외, 2022).
 - 해당 연구에서는 초·중학교를 중심으로 디지털 기술 도입에 따른 교육 변화와 교사의 수용 정도를 모니터링하기 위해 ‘학교 교육 디지털 전환 조사(안)’을 설계함.
 - 조사는 디지털 인프라, 교수·학습, 데이터 수집·관리·활용, 구성원 상호작용 및 교육 기회 보장 4개 영역으로 구성되었으며, 현재와 미래의 디지털 전환에 대한 인식을 비교 분석할 수 있도록 고안되었음.
- 또한, 한국교육학술정보원(KERIS)은 기존 ‘교육정보화 실태조사’를 확장하여 초·중등학교 디지털 전환 실태조사를 실시함.
 - 2023년에는 전국 12,223개교의 디지털 교육 환경을 분석하였고, 344개교의 교사 1,074명을 대상으로 설문조사를 진행함.
 - 이를 통해 학교 단위의 디지털 전환 수준을 진단하기 위한 8개 영역의 분석 프레임워크(디지털 준비, 참여, 학습 안전, 교육담당자 역량, 교육혁신, 정책 및 규정, 협업, 평가관리)를 개발함.
- 해당 연구는 교육 현장에서 디지털 전환이 실제로 어떻게 이루어지고 있는지를 종합적으로 진단하고, 정책 개발을 위한 기초 자료를 제공한다는 점에서 의미가 큼.

나) 국외 조사 연구

① 국제 학업성취도 평가(PISA)의 ICT 활용 관련 조사

- PISA(Programme for International Student Assessment)는 OECD가 주관하는 3년 주기의 국제 학업성취도 평가로, 만 15세 학생들이 사회·경제적 삶에 필요한 핵심 역량을 얼마나 갖추고 있는지를 측정하는 대규모 조사임(OECD,

2023a).

- 주요 평가 영역은 읽기, 수학, 과학이며, 이와 함께 학생들의 ICT 활용 능력과 디지털 학습 환경에 대한 조사도 병행됨.
- 뿐만 아니라, 학생 대상 설문에서 컴퓨터 친숙도 문항을 통해 ICT 사용 능력, 사용 환경, 태도 등을 측정하고, 학교장 및 교사를 대상으로는 학교의 ICT 인프라, 정책, 교사 연수, 교수법에서의 ICT 통합 현황 등을 조사하기도 함(OECD, 2023b).
- 특히 PISA 2022부터는 ICT 활용 실태를 보다 심층적으로 파악하기 위해 ICT 활용 이론 프레임워크가 새롭게 도입되었음.
- 프레임워크는 ICT 접근성, ICT 사용, 학생의 ICT 역량의 3가지 차원으로 구성됨.
 - ‘ICT 접근성’은 디지털 자원의 가용성과 품질, 네트워크 환경 등 기술적 인프라 수준을 의미함.
 - ‘ICT 사용’은 학생이 수업 안팎에서 ICT를 활용하는 빈도, 유형, 목적 등을 포함함.
 - ‘ICT 역량’은 디지털 리터러시와 관련된 지식, 기술, 태도는 물론, ICT 기반 학습 효능감과 문제 해결 능력 등을 포괄함.
- 이와 같은 ICT 문항은 학업 성취(읽기, 수학, 과학)는 물론, 정서적 웰빙, 자기주도성, 디지털 격차 등 다양한 교육 결과 변수와의 관련성을 분석할 수 있는 기초 자료를 제공함.
- 이를 통해 국가 간 비교는 물론, 교육 시스템 수준에서의 디지털 전환 현황과 정책 효과를 종합적으로 진단할 수 있음.

② 국제 컴퓨터·정보 리터러시 연구(ICILS)

- ICILS(International Computer and Information Literacy Study)는 국제교육성취도평가협회(IEA)가 주관하는 국제 비교 연구로, 8학년 학생들의 컴퓨터 및 정보 기술 활용 역량을 측정하기 위해 5년 주기로 실시됨.
- 2013년부터 시작된 본 조사는 학생들이 디지털 환경 속에서 학습, 일상생

활, 직업 활동에 얼마나 잘 대비되어 있는지를 평가하기 위해 설계됨 (Fraillon et al., 2013).

- 학생의 ICT 역량은 컴퓨터·정보 리터러시(CIL)와 컴퓨팅 사고력(CT) 2가지 차원으로 구분하여 측정됨.
 - ‘CIL’은 ‘학교, 가정, 직장 등 다양한 맥락에서 컴퓨터를 활용하여 정보를 조사하고, 창작하며, 의사소통하는 능력’을 의미함.
 - ‘CT’는 ‘현실 문제를 인식하고 컴퓨터를 활용해 실행 가능한 알고리즘적 해결 방안을 개발하고 평가하는 능력’으로 정의됨.
- 이와 더불어, 학생의 배경, ICT 활용 경험, ICT 활용 맥락과의 관련성에 대한 데이터도 함께 수집됨.
- 학교 차원에서는 ICT 접근성, ICT 사용 경험, CIL 및 CT 관련 교수·학습 전략, 학교의 리더십 및 지원 체계 등을 중심으로 자료를 수집하며, 개인 차원에서는 학생의 ICT 사용 빈도, 목적, 태도 등 개인적 특성과 일상적 맥락을 조사함.
- 이를 통해 각국의 교육 시스템이 학생들의 ICT 역량 강화를 위해 제공하는 환경과 정책, 자원 등을 비교·진단할 수 있는 기초 자료가 마련됨.
- 특히 2023년 조사에서는 CIL과 CT의 평가 구조가 보다 세분화 되었음.
 - ‘CIL’은 컴퓨터 사용에 대한 이해, 정보 수집, 정보 생산, 디지털 의사소통의 4가지 영역으로 구성되며, 각 영역은 정보 처리, 창작, 공유, 윤리적 활용 등의 세부 역량을 포함함.
 - ‘CT’는 문제 개념화, 해결책 실행의 두 영역으로 구성되며, 문제 구조화, 알고리즘 개발, 프로그램 구현, 사용자 상호작용 이해 등 실질적 문제 해결 능력을 측정하는 데 중점을 두고 있음.

3) 디지털 교육 효과에 대한 연구

- OECD(2023a)는 디지털 교육의 이론적 효과로 맞춤형 학습, 포용성 및 형평성 증진, 교사 업무 효율화, 행정 효율성 향상, 데이터 기반 정책 개발 등을 제시하며, 디지털 전환이 단순한 도구의 변화가 아니라 교육 시스템 전반의 구조적 혁신으로 이어질 수 있음을 강조하고 있음(문희원, 정동욱, 2025).

- 이러한 관점은 디지털 교육이 단순한 기술 도입의 차원을 넘어, 교육의 질과 형평성을 포괄하는 근본적인 변화의 계기가 될 수 있음을 시사함.
- 그러나 디지털 교육에 대한 긍정적인 기대와는 달리, 일각에서는 디지털 교육이 초래할 수 있는 부정적인 결과에 대한 우려도 제기되고 있음.
- 이와 같은 맥락에서 디지털 교육이 실제로 학생의 학업 성취, 교수 학습, 교육의 형평성 확보 등에 어떠한 영향을 미치는지를 실증적으로 검토하는 것은 여전히 중요한 과제임.
- 국내·외에서는 디지털 기반 수업, AI 튜터 활용, 온라인 학습 플랫폼 등 다양한 형태의 디지털 교육이 학습자와 교사, 학교 수준에서 어떤 변화를 유도하고 있는지에 대한 실험연구, 비교연구, 사례연구들이 축적되고 있으며, 이는 크게 학업 성취, 문해력, 학습 격차 해소 등 3가지 영역에 대한 연구로 나누어 살펴볼 수 있음.

가) 학업 성취

- 디지털 교육의 학업 성취에 대한 효과와 관련된 연구는 크게 학생 맞춤형 학습 프로그램 활용과 디지털 기기 제공의 2가지 측면으로 구분할 수 있음 (문희원, 정동욱, 2025).
- 먼저, 학생 맞춤형 학습 프로그램은 학습자의 개별 특성에 맞춰 피드백을 제공하고 학습 과정을 조정하며, 정규 수업 외의 시간에도 자기주도적 학습 활동이 지속될 수 있도록 돕는다는 점에서 의의가 있음.
 - 미국의 맞춤형 학습 프로그램인 ASSISTments는 문제 해결 과정에서 학생에게 단계별로 유도 질문과 피드백을 제공하여 학습자의 이해를 도우며, 특히 수학 성취 향상에 긍정적인 영향을 미치는 것으로 나타남(Escueta et al., 2020).
 - Roschelle 외(2016)의 연구에서도 미국 Maine 주 내 43개 학교에서 ASSISTments를 주 3~4회 하루 평균 10분 내외로 활용한 학생 집단이 그렇지 않은 집단에 비해 수학 성취도에서 유의미한 향상을 보인 것으로 나타남.

- 다음으로, 디지털 기기와 안정적인 인터넷 환경은 학습 공간을 교실에 국한하지 않고 다양한 장소로 확장시키며, 자료의 탐색과 활용에 소요되는 시간과 비용을 절감함으로써 학습자의 역량 형성을 보다 효율적으로 지원할 수 있는 환경을 조성함.
- 그러나 단순히 노트북과 인터넷 등 디지털 기기를 제공하는 것만으로 학업 성취에 유의미한 변화가 생기지 않는다는 연구 결과가 보고됨(Fairlie & Robinson, 2013; Goolsbee & Guryan, 2006; Leuven et al., 2007).
- 또한, 디지털 기기 제공이 수학, 영어 학업 성취에 부정적인 영향을 미칠 수 있으며, 이는 부모의 적절한 지도와 감독을 통해 완화될 수 있다는 연구도 보고된 바 있음(Malamud & Pop-Eleches, 2011).
- 선행연구를 종합하였을 때, 단순히 기기 보급만으로는 학업 성취 향상을 기대하기 어려우며, 학습자의 수준과 특성에 맞춘 적응형 학습 시스템이나 개별 피드백 중심의 프로그램이 함께 제공될 때 교육적으로 유의한 성과를 기대할 수 있음.

나) 문해력

- 디지털 교육 도입 초기에는 전자책이나 태블릿 사용이 주의력 분산, 눈의 피로, 과도한 자극 등으로 문해력 저하를 유발할 수 있다는 우려가 제기되어옴(Honma et al., 2022; Jeong, 2012)
- 그러나, 최근 메타분석 연구에서는 에듀테크 활용이 오히려 초등학생의 문해력 향상에 효과적이라는 결과가 지속적으로 보고됨(Silverman et al., 2024).
- 2010년부터 2023년까지 119개 실험연구 결과를 종합하여 에듀테크 활용이 문해력의 4가지 영역(독해, 언어 이해, 읽기 이해, 작문)에 미치는 영향을 분석한 결과, 영역별로 차이는 있지만 모든 영역에서 긍정적인 효과가 있는 것으로 나타남.
- 가장 효과가 큰 영역은 작문 영역(Hedge's $g = 0.81$)이었으며, 뒤이어 독해(Hedge's $g = 0.33$), 언어 이해(Hedge's $g = 0.30$), 읽기 이해(Hedge's $g = 0.23$) 순으로 향상 효과가 있었음.
- 한편 게임화(gamification), 적응형 학습(adaptivity), 피드백 제공 등의 추가

기능 제공은 독해와 읽기 이해에 대한 효과를 유의하게 감소시키지 않았으나, 언어 이해에서는 에듀테크 효과가 일부 감소되는 것으로 나타남.

- 각 문해력 영역별로 에듀테크의 기능이 미치는 영향이 달라지는 이유는 밝혀지지 않았으나, 각 영역별로 효과적인 에듀테크 기능이 달라질 수 있으므로 설계 단계에서 내용 요소에 따라 기능 선택에 신중을 기할 필요가 있음.

다) 학습격차 해소

- 디지털 교육의 보편화는 학생 간 기술 접근성 및 활용 능력의 차이를 확대시켜, 교육의 형평성에 어긋난다는 비판이 제기되어 왔음.
- 일부 연구에서는 학생 간 디지털 활용 역량 차이는 디지털 교육 효과의 편차로 이어져, 학습 격차의 심화로 이어질 가능성을 지적하며, 디지털 교육이 불러올 부정적인 결과에 대한 우려를 제기함(Jeong et al., 2024).
- 그러나, 실증 연구에서는 디지털 교육이 오히려 소외계층 학생에게 더 긍정적인 효과가 나타난다는 결과가 보고됨(Feng et al., 2023).
 - 학습 속도나 수준에 따라 개별화된 학습을 제공하는 에듀테크는 초기 성취도가 낮았던 소수 집단에서 더 큰 효과를 발휘했으며, 그 효과는 단기적으로 그치지 않고 장기적으로 학습 성과를 지속적으로 향상시키는 데 기여한 것으로 나타남.
- 한편, 디지털 교육의 효과 크기가 매우 작게 나타난 메타분석 결과(Cohen's $d = 0.029$)도 존재하는 만큼(Di Pietro & Munoz, 2025), 맥락적 조건과 실행 방식에 대한 정교한 고려가 병행될 필요가 있음.
- 이처럼 디지털 교육은 학생 간 격차를 줄이고 공정한 학습 기회를 확대하는데 있어 유효한 대안이 될 수 있음.
- 다만 디지털 교육의 효과는 적용 환경과 설계 방식에 따라 달라질 수 있으므로, 형평성 관점에서의 지속적이고 체계적인 검증이 요구됨.

4) 시사점

- 상기한 내용을 종합하여, 디지털 교육 효과 분석을 위해 도출한 시사점은 다음과 같음.

- 첫째, 디지털 교육 효과 분석을 위한 체계적인 조사 기반이 요구됨.
 - 디지털 교육의 효과를 과학적으로 진단하고 정책적 방향성을 설정하기 위해서는, 교육 주체별 디지털 역량과 학교 단위의 전환 수준 등을 정기적으로 점검할 수 있는 조사 체계가 마련되어야 함.
 - 최근 국내·외에서는 디지털 리터러시, 교원 역량, 디지털 전환 환경 등을 측정하는 다양한 조사연구가 진행되고 있으며, 이는 디지털 교육 효과 분석의 기반으로 작용할 수 있음.
 - 이러한 대규모 조사연구는 정량적·정성적 데이터를 통해 디지털 교육 정책의 수립과 실행을 뒷받침하는 데 중요한 역할을 하므로, 지속적인 확장고도화가 필요함.
- 둘째, 학업 성취, 문해력, 학습격차 해소 등 주요 교육성과를 중심으로 한 실증연구의 축적이 필요함.
 - 디지털 교육은 교육성과에 긍정적인 영향을 미칠 수 있는 가능성을 보여주고 있으나, 그 효과는 기술의 종류나 적용 방식, 학습자의 특성에 따라 상이하게 나타남.
 - 예를 들어, 맞춤형 학습 프로그램은 학업 성취 향상에 기여하는 반면, 단순한 기기 보급은 유의미한 성과를 보이지 않음.
 - 문해력의 경우에도 영역에 따라 효과 크기에 차이가 존재하며, 학습격차 해소 측면에서도 적용 맥락에 따라 결과가 달라지는 경향이 있음.
 - 한편, 디지털 기기의 과도한 사용으로 인한 주의력 분산, 사회적 상호작용 감소, 디지털 접근성이나 활용 역량의 격차가 오히려 기존의 교육 불평등을 심화시킬 수 있다는 지적 등 디지털 교육의 부작용에 대한 우려가 공존하는 만큼 디지털 교육 효과를 보다 신중하고, 다각적인 관점으로 검토할 필요가 있음.
- 셋째, 디지털 교육 효과를 극대화하기 위한 교육 주체별 역량 강화가 병행되어야 함.
 - 학생의 디지털 활용 능력은 정보탐색, 협력, 창작, 안전 등 복합적 역량을 포함하므로, 이를 체계적으로 진단하고 교육적으로 강화할 수 있는 방안이

필요함.

- 또한, 디지털 교육을 설계하고 운영하는 교사는 AI 및 디지털 기술을 비판적이고 창의적으로 활용할 수 있는 능력과 윤리적 태도를 갖추어야 하며, 이를 위한 실천적 역량 기반의 연수와 지원이 요구됨.
- 역량 강화는 단순한 기능 습득을 넘어서, 디지털 환경에 적응하고 변화를 주도할 수 있는 전문성을 개발하는 방향으로 이루어져야 함.
- 넷째, 디지털 교육 효과 분석은 기술 활용 여부를 넘어 교육 시스템 전반의 구조적 전환과 연계되어야 함.
- 디지털 교육의 실질적인 효과는 개별 수업이나 기술 도입 여부만으로 설명되기 어려우며, 교육과정 운영 방식, 학습 환경, 상호작용 구조 등 전반적인 시스템의 변화와 밀접하게 연결됨.
- 따라서 효과 분석은 이러한 구조적 전환의 맥락을 함께 고려해야 하며, 기술 중심의 단편적 분석을 넘어 교육 전반의 질적 변화를 포착할 수 있는 분석 틀의 설계가 요구됨.

나. 교육에서의 실험연구

1) 무작위 대조 실험 (Randomized Control Trial, RCT)

- 무선으로 표집된 표본을 처치와 통제 집단으로 무선 할당하는 방식
- 사례: STAR(Student/Teacher Achievement Ratio) 프로젝트 (미국 테네시주)
 - (목적) 학급 규모가 학업 성취도에 미치는 인과적 영향 분석.
 - (방법) 총 11,600명의 학생들을 소규모 학급(처치), 일반 규모 학급(통제1), 보조교사 학급(통제2)에 무작위 배정.
 - ※ 처치 집단과 통제 집단 모두 기존 평균 이상의 교육 환경을 유지하여, 어느 학생도 실험으로 인해 불리한 조건에 처하지 않도록 설계
 - (결과) 소규모 학급 학생들이 읽기·수학 성취도에서 더 높은 성과, 보조교사 배치는 큰 효과 없음.
 - (정책적 영향) 미국 여러 주(예: 캘리포니아)에서 학급 규모 축소 정책 추진
 - (한계) 학생 이동, 학급 변경 등 표본 이탈 문제 발생 → 후속 연구를 통해

주요 결과에 영향 없음 확인.

2) 유도 설계 실험 (Encouragement Design)

○ 정책(처치)을 희망하는 지원자를 대상으로 처치와 통제 집단을 무선 할당하는 방식

○ 사례: 밀워키 실험 (Milwaukee Parental Choice Program)

- (목적) 학교 선택 바우처(School choice voucher)*의 효과 분석.

* 저소득층 가정의 학생들이 공립학교 외에 사립학교를 선택할 수 있도록 지원하는 제도

- (방법) 바우처 신청자가 정원을 초과하는 경우 신청자를 바우처 수혜(처치), 바우처 비수혜(통제) 집단으로 무작위 배정.
- (결과) 바우처 수혜 학생들의 일부 학업 성취 개선 보고.
- (한계1) 학교/학년 간 차이 통제 어려움, 종교 학교 제외 등의 학생의 실질적 선택권 제약으로 잠재적 효과 제한 가능성.
- (한계2) 장기추적 어려움 및 학생 이동 문제 존재.

3) 디지털 교육 관련 실험연구

○ 디지털 교육 관련 실험연구는 RCT와 Encouragement Design을 활용한 연구로 구분됨.

- RCT 연구의 경우, 무작위 배정 이후 실험집단과 통제집단의 차이를 확인하기 위하여 인구학적 특성을 비교하고, 집단 간 차이가 나타날 경우 해당 변수를 분석 모델에 포함하여 통제함(Rutherford et al., 2014; Snipes et al., 2015; Tatar et al., 2008).
- 한편, RCT의 윤리적, 실행적 한계를 보완하기 위하여 다수의 연구에서는 참여자들의 신청을 받아서 진행하는 Encourage Design 방식으로 실험연구를 수행함(Cabalo, Ma, & Jaciw, 2007; Campuzano et al., 2009; Deault, Svage, & Abrami, 2009; Faber & Visscher, 2018).
- 디지털 교육의 효과성을 메타분석한 Escueta et al.(2020)에 포함된 컴퓨터 보조 학습(Computer-Assisted Learning) 관련 총 30편의 실험 논문을 분석한 결과, RCT 실험이 18편(60%), Encouragement Design 실험이 12편(40%)

으로 나타남. 실험 단위는 개인(학생·교사 등) 단위가 56.7%로 가장 많았고, 다음으로 학교 단위(23.3%), 학급 단위(20.0%) 순으로 높게 나타남. 개인 단위 실험의 경우 대부분 RCT 설계가 활용된 반면, 학교 및 학급 단위 실험 설계의 경우 Encouragement Design이 많이 활용됨.

<표 14> 컴퓨터 보조 학습 관련 실험 연구 분석

(논문 편수, %)

실험 설계 방법	실험 단위			
	학교 단위	학급 단위	개인 단위 (학생, 교사)	계
RCT	3	4	11	18 (60%)
Encouragement Design	4	2	6	12 (40%)
계	7 (23.3%)	6 (20.0%)	17 (56.7%)	30 (100%)

※ Escueta et al.(2020)에 포함된 컴퓨터 보조 학습(Computer-Assisted Learning) 관련 총 30편의 실험 논문 분석

○ 실험 목적에 따른 다양한 단위의 실험 설계 사례

<표 15> 실험 목적에 따른 실험 단위별 실험 설계 사례

실험 단위	실험 목적 및 실험 단위 설정	선행연구
학교	<ul style="list-style-type: none"> - 프로그램 도입이 학생의 성취도에 어떠한 영향을 미쳤는지 분석 - 동일 학교 내 교사 간의 상호작용으로 인한 파급효과(spillover effect)를 방지하고자 학교 단위 실험 설계 	Cavalluzzo et al., 2012; Rockoff, 2015
학급	<ul style="list-style-type: none"> - 특정 학급에서 일괄적으로 적용되는 프로그램(커리큘럼)이 학생의 성취도에 미치는 영향을 분석 - 동일 교사가 여러 학급을 가르치고, 개별 학생에게 프로그램을 다르게 적용하기 어려움에 따라 학급 단위 실험 설계 	Hegedus, Dalton, & Tapper, 2015; Wijekumar, Meyer, & Lei, 2012; Wijekumar et al., 2014
교사	<ul style="list-style-type: none"> - 교사의 프로그램 활용이 학생 성취도에 미치는 영향 분석 - 프로그램 사용 및 실행이 교사 단위로 이루어지기 때문에 교사 단위 실험 설계 	Campuzano et al., 2009; Dynarski et al., 2007; Tatar et al., 2008
학생	<ul style="list-style-type: none"> - 개별 학생이 수행할 수 있는 프로그램이 학생 성취도에 미치는 영향 분석 - 교육 기술의 개입이 개별 학습자에게 직접 제공되고, 효과 또한 학생의 성취 수준에서 측정됨에 따라 학생 단위 실험 설계 	Deault, Svage, & Abrami, 2009; Faber & Visscher, 2018; Kelly et al., 2013

○ 기타 실험 설계 사례

- 지연-처치 방식을 통해 1년차는 RCT로 실험집단과 통제집단 간의 효과를 검증하고, 2년차는 통제집단이었던 교사들도 처치 집단에 포함하여 교사의 기기 사용 여부에 따른 변화 정도를 검증(Roschelle et al., 2010; Roschelle et al., 2016).
- 학교, 학급 단위의 연구의 경우, 비슷한 학교, 학급의 특성을 공유하는 쌍(pair)을 만들어 매칭 후, 각각 실험집단과 통제집단에 무작위 배정하는 방법(cluster RCT, matched-pairs design)을 활용(Cabalo, Ma, & Jaciw, 2007; Rockoff, 2015; Pane et al., 2014).

○ 실험연구에서 발생하는 편향(Bias) 등의 문제를 해결하기 위한 사례

① 처치 충실도 및 일관성 관리

- 대체 편향(Substitution bias)이 발생하지 않도록 프로그램 사용, 시험지 등을 철저히 통제(Van Klaveren, Vonk, & Cornelisz, 2017).
- 수업 주체인 교사들의 교육과정 변경을 위하여 시간적 여유를 두고 실험 진행(Pane et al., 2014; Roschelle et al., 2016).
- 처치의 효과를 높이기 위해 처치 그룹의 교사들에게 기기 활용 교육 워크숍 수행(Ritter et al., 2007; Roschelle et al., 2010; Roschelle et al., 2016).

② 교사 영향 통제

- 교사의 특성을 통제하기 위해 집단별로 사전 교육을 수행하여 교사 간 영향이나 수업 방식의 차이를 최소화(Mitchell & Fox, 2001)
- 동일 교사가 실험집단과 통제집단의 수업을 모두 진행(Morgan & Ritter, 2002; Cabalo, Ma, & Jaciw, 2007).
- 학급의 특성과 교사의 특성을 통제할 수 있다는 장점이 있지만, 처치하는 학습 기기가 다르거나, 코스웨어의 알고리즘이 다른 경우에 한정적으로 사용 가능함(Van Klaveren, Vonk, & Cornelisz, 2017).
- 각 학교의 두 명의 교사에게 서로 다른 시기에 두 가지 수업(처치, 통제집단)을 각각 배정하여 교사의 영향력을 통제(Pane et al., 2010).

③ 중도 탈락 제외

- 중도 탈락이 발생한 경우, 해당 사례를 연구에서 제외하고 학교, 학급 단위의 연구인 경우에도 쌍(pair)을 이룬 실험집단 또는 통제집단의 자료를 제외(Ritter et al., 2007; Rockoff, 2015; Roschelle et al., 2016).
- 학교 개인 사정으로 인해 실험 전에 이탈한 경우, 학교 간 연구에서 학교 내 연구로 조정하여 연구를 유연하게 진행(Rockoff, 2015).

④ 사전-사후 평가

- 대체로 학생의 사전 성취도 점수와 사후 성취도 점수를 비교하여 디지털 기기(또는 프로그램)의 효과를 확인(Hegedus, Dalton, & Tapper, 2015; Rutherford et al., 2014).
- 디지털 교육에서 기기를 활용한 학습은 이전과는 다른 경험을 제공하기 때문에 기존의 표준화된 평가 방식 외에 별도의 평가 도구를 마련하여 효과를 분석(Roschelle et al., 2010; Wijekumar, Meyer, & Lei, 2012; Wijekumar et al., 2014).

⑤ 효과 추정 및 분석의 엄밀성 확보

- 연구의 무결성을 위하여 실험을 포기하는 참여자가 나오더라도 원래의 배정 그대로 성과를 측정하는 Intent-to-Treat (ITT) 방식을 활용하고, 실제 처치를 받은 효과를 추정하는 Treatment-on-the-Treated (TOT)를 추가적으로 활용하여 과소 추정 문제를 보완(Borman, Benson, & Overman, 2009; Rouse & Krueger, 2004; Rutherford et al., 2014; Schenke, Rutherford, & Farkas, 2014; Wang & Woodworth, 2011).
- 집단 간 프로그램 사용이 혼재되거나, 실험집단 학생들의 프로그램 활용 수준이 다르기 때문에 도구변수(실험집단 배정 여부)를 활용하여 TOT를 추정(Barrow, Markman, and Rouse, 2009; Borman, Benson, and Overman, 2009).
- 학급의 구성이 다른 집단을 대상으로 할 때, 학급 및 교사의 영향을 분리하고자 위계적 선형 모델(Hierarchical Linear Modeling)을 활용(Dynarski et al., 2007; Ritter et al., 2007; Roschelle et al., 2010; Roschelle et al., 2016;

Wijekumar, Meyer, & Lei, 2012; Wijekumar et al., 2014).

⑥ 윤리 문제

- 실험연구에서의 윤리적 문제를 해소하고자 실험이 끝난 후 통제집단에도 동일한 프로그램을 사용하도록 제공(Roschelle et al., 2010; Wijekumar, Meyer, & Lei, 2012).

4) 시사점

- 디지털 교육 정책의 인과적 효과를 분석하기 위한 실험연구에서 무작위 배정을 통해 처치-통제집단을 비교하는 RCT가 가장 이상적인 실험연구이나, 윤리적 고려와 참여자 반발을 고려하여 무작위 배정 대신 Encouragement Design 적용 가능.
- 무작위 배정이 잘 이루어지더라도 대체 편향(Substitution Bias), 표본 탈락 등의 문제로 편의(bias)가 발생할 가능성이 존재하기 때문에 실험연구 설계 시 엄밀한 설계가 필요하고, 실험 후 분석 시에도 다양한 보정 노력을 통해 명확한 정책 효과 추정이 필요.
- 공정한 실험을 위해 처치 또는 통제집단이 불리한 조건에 처하지 않도록 세심한 연구설계 필요.

다. 국가수준 학업성취도 평가의 검사 개발, 동등화, 수준설정

1) 국가수준 학업성취도 평가의 개요

- 학생들의 학업 성취 수준을 국가 차원에서 진단하고 기초학력 보장 및 학습 결손을 조기 발견, 교육과정 운영 실태 파악 및 개선, 시도/학교 유형 간 학력 수준 비교, 교육정책 수립 및 학력 신장 지원에 필요한 기초자료 제공 등의 목적에 따라 시행.
- 대상 및 교과
 - 대상 학년: 중학교 3학년, 고등학교 2학년
 - 평가 교과: 국어, 수학, 영어(공통) / 중학교는 사회·과학 포함

- 평가 방식: 2016년까지 전수 → 2017년부터 표집 방식 운영
- 2023년 예시: 중3 14,642명, 고2 10,193명 참여

○ 시기 및 운영

- 매년 9월경 평가 실시(과거에는 6월)
- 교육청과 학교 단위로 시행, KICE(한국교육과정평가원) 주관
- 인지적 평가 + 비인지 설문조사 동시 실시

2) CBT(컴퓨터 기반 평가) 전환

○ 도입 배경

- 기존 지필평가의 한계 극복: 고정된 문항 유형, 정답 위주의 평가
- 실생활 기반 문제해결 능력 평가 요구 증대
- 기술 기반 평가 확산(PISA, NAEP 등 글로벌 추세 반영)

○ 도입 시기 및 과정

- 2020년 일부 학교 대상 CBT 시범 도입
- 2022년 전면 전환, 2023년 안정화
- 각 학교에서 사전 장비 점검 및 모의평가 실시

○ CBT 문항 유형 예시

- 시뮬레이션형: 실험 도구 조작
- 자료탐색형: 정보 비교, 그래프 해석
- 대화형 인터페이스: 역할극, 의견 교환
- 멀티미디어형: 영상, 음성 기반 문항

○ 장점

- 복합역량(문해력, 문제해결력) 평가 가능
- 채점 자동화로 결과 제공 신속화
- 문항 다양화로 학생 흥미 유발

- 환경친화적(종이 사용 감소)

○ 단점 및 고려사항

- 디지털 격차: 기기 활용 경험 차이
- 초기 시스템 오류 및 서버 부하 가능
- CBT에 대한 교사/학생의 사전 훈련 필요

3) 문항 개발

○ 문항 유형

- 선다형: 객관식 문항, 자동 채점 용이
- 서답형: 단답·서술형으로 사고력 측정 가능
- 정보활용형: 웹 자료 분석 및 응용
- 혼합형: 복합 스킬을 요구하는 시나리오 기반 문항

○ 개발 절차

- 평가틀 개발: 교육과정 기반, 영역/문항 수 등 구조 설계
- 문항 집필: 교과 전문가 및 교사들이 문항 초안 작성
- 전문가 검토: 타당도·난이도 검토 및 편향 요소 제거
- 예비검사 시행: 소규모 학생 대상 시범 운영
- 문항 분석: 문항반응이론 및 고전검사이론 활용
- 최종 선별: 통계 지표 기준 부적합 문항 제거
- 문항 품질 관리
- 문항은행(item bank) 관리
- 교육과정 변화 시 문항군 재구성
- 오답지 매력도 및 오답 유형 분석
- 문항반응 시간 및 클릭 패턴 분석(CBT의 경우)

4) 검사 동등화

○ 목적

- 해마다 다른 문항 사용 → 난이도 차이 보정 필요
- 여러 시험지 유형 간 점수 비교 가능성 확보

○ 방법

- 공통 문항 활용
- 전년도 평가 문항 일부 재사용하여 검사 간 연결
- 문항반응이론 기반 동등화

4) 수준설정(standard setting)

○ 목적

- 점수에 의미 부여: 학생 수준별 해석 기준 마련
- 성취기준 도달 여부 판단 기준점(cut score) 설정

○ 방법

- 수정된 앵고프 방법
- 문항별로 최소능력보유자의 정답 확률 추정
- 패널 평정 → 평균화하여 컷 점수 결정
- Bookmark 방법: 문항 난이도순 정렬 → 수준별 북마크 설정

○ 절차 및 운영

- 패널 구성: 현장 교사, 교수, 평가 전문가 등
- 2~3회 반복 회의, 통계 확인 후 확정
- 필요시 정책적 판단으로 조정 가능

5. 연구결과의 활용방안

- (국가 수준) 본 연구를 통해 수집된 디지털 교육 중단연구 자료는 국가 수준에서 디지털 교육 관련 핵심의제를 설정하고 교육의 디지털 전환을 지원하기 위한 다양한 정책 방안을 마련하는 데 의미 있는 자료로 활용될 수 있을

것임. 즉, 데이터 기반 의사결정을 도울 수 있는 자료가 본 연구를 통해 수집된다는 점에서 디지털 교육 종단연구 조사 체계를 타당하고 신뢰롭게 수립하는 것은 매우 중요한 과업임.

- 수집된 자료를 전문적으로 분석하여 그 결과를 정책적으로 활용하는 것과 함께, 학부모, 교사, 학생들에게 공유하고 확산함으로써 디지털 교육에 대한 신뢰를 높이는 데 활용될 수 있음.

○ (학생, 교사, 학교 등 교육 현장) 2022 개정 교육과정은 디지털 소양을 기초 소양 중 하나로 설정하고 디지털 환경에 맞는 교수학습 및 평가를 혁신적으로 도입하여 이를 함양할 수 있도록 하고 있음. 즉, 디지털 소양(디지털 리터러시)은 학교 교육을 통해 길러내야 할 중요한 미래 역량으로 여겨지고 있음. 관련하여 본 조사는 학생의 디지털 리터러시 등 미래 역량도 측정하고 있어 이를 학교 단위나 학생 개인별로 피드백을 제공하는 등의 방법을 통해 교육적으로 활용할 수도 있을 것임.

- 뿐만 아니라, 교사의 디지털 리터러시도 함께 조사되며 교사의 수업을 들여다 볼 수 있는 다양한 지표가 조사된다는 점에서 교사의 디지털 기반 교수학습 및 평가 역량을 함양하기 위한 연수(컨설팅) 또는 전문적 학습 공동체를 운영하는 데 있어서 유용하게 활용될 수 있음.

○ (학계) 기존 국가 및 지역 수준에서의 패널 조사 등에서는 디지털 교육 관련 일부 지표가 조사되는 데 그쳐, 디지털 교육의 성과 전반을 충실히 볼 수 있는 자료는 거의 존재하지 않았다고 볼 수 있음. 이러한 면에서 본 연구를 통해 수집된 자료가 공개되어 다양한 분야의 전문가나 현장 교사들이 분석할 수 있게 된다면 학술적으로도 활용도가 매우 높을 것으로 기대됨.

- 구축된 데이터셋을 활용한 정책연구 주제 및 방법론 등을 제안함으로써 향후 조사 결과를 활용한 2차 분석 연구의 활성화를 도모할 수도 있을 것임.

6. 연구 추진 일정

○ 연구 추진 일정: 2025. 5. 계약일 ~ 2025. 12. 19. (약 8개월)

연구 내용		추진 일정							
		5	6	7	8	9	10	11	12
연구개시 및 착수보고회 개최									
과제 1	- 학교표본 추출 및 동의서 수집								
	- 조사시스템 구축 및 조사원 교육								
	- 사전 설문조사 실시								
	- 데이터클리닝								
	- 결과 분석								
	- 조사시스템 구축 및 조사원 교육								
	- 조사 실시								
	- 데이터클리닝 및 결과분석								
중간보고									
성취도 개발	- 1차 조사 문항 개발 및 검토								
	- 예비조사 실시, 본조사 문항 확정, CBT 구축								
	- (1차 조사 실시)								
	- 수준 설정 및 동등화								
	- 2차 조사 문항 개발 및 검토								
	- 예비조사 실시, 본조사 문항 확정, CBT 구축								
	- (2차 조사 실시)								
	- 수준 설정 및 동등화								
과제 2	- 연구 설계 및 대상 확정								
	- 동의서 수집								
	- (사전 설문조사 실시)								
	- 1차 조사 실시								
	- 데이터클리닝 및 결과분석								
	- 학교별 리포트 및 학생용 성적표 작성								
	- 2차 조사 실시								
	- 데이터클리닝 및 결과보고서 작성								
연구보고서 작성 및 완료 보고									

※ 상기 일정은 연구 추진 상황에 따라 변동될 수 있음

7. 참고문헌

- 계보경, 김혜숙, 차현진, 김현진, 한나라, 문범섭(2022). **수행형 기반 디지털리터러시 평가도구 개발 연구**. 연구보고 RR 2023-6. 한국교육학술정보원.
- 계보경, 이현숙, 한나라, 김혜숙(2021). **2021년 국가수준 초·중학생 디지털 리터러시 수준 측정 연구**. 연구보고 RR 2021-6. 한국교육학술정보원.
- 교육부 (2021). 2022 개정교육과정 총론 주요사항.
- 교육부 (2023.2). **디지털 기반 교육혁신방안**. 교육부 보도자료 (2023. 2. 23.).
- 교육부 (2024.4). **디지털 기반 교육혁신 역량 강화 지원방안**. 교육부 보도자료 (2024. 4. 15.).
- 교육부 (2024.5). **초중등 디지털 인프라 개선계획(안)**. 교육부 보도자료 (2024. 5. 15.).
- 권희경, 김나영, 김혜자, 남궁지영, 박근영, 이강주, 이윤서, 최인희, 최은아(2022). **교육 및 데이터 환경 변화에 따른 교육지표 개선 방안 연구**. 한국교육개발원 연구보고서 RR 2022-14.
- 김미화, 정동욱(2013). 초등학교 학급담임 배정 영향 요인 분석. **한국교원교육연구**, 30(3), 1-20.
- 문희원, 정동욱(2025). **디지털 교육 효과에 대한 메타연구 고찰**. 서울대학교 학습과학 연구소 이슈페이퍼 2025년 제1호.
- 박소미, 신부경, 김미송, 조영환 (2024). 온라인 협력적 문제해결에서 고수행 집단과 저수행 집단 간 상호작용 패턴의 차이. **교육정보미디어연구**, 30(3), 837-859.
- 박현정, 정광훈, 길혜지, 정동욱, 조영환, 임철일, 최미애, 서정희, 김은지(2025). **한국 디지털 교육 종단연구 설계 및 실행방안 연구**. 한국교육학술정보원 RR 2025-01.
- 서정희, 허희옥, 임규연, 이현우, 임경희, 박세영, 공현아, 권미영, 최미애, 정광훈 (2023). **예비·현직 교원의 AI·디지털 역량 측정도구 개발**. 한국교육학술정보원 연구자료 KR 2023-08.
- Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of economic literature*, 59(2), 391-425.
- Angoff, W. H. (1971). Scales, norms and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (pp. 508-600). Washington, DC: American Council on

Education.

- Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.
- Barrow, L., Markman, L., & Rouse, C. E. (2009). Technology's edge: The educational benefits of computer-aided instruction. *American Economic Journal: Economic Policy*, 1(1), 52-74.
- Bärnighausen, T., Tugwell, P., Røttingen, J. A., Shemilt, I., Rockers, P., Geldsetzer, P., & Atun, R. (2017). Quasi-experimental study designs series —paper 4: uses and value. *Journal of clinical epidemiology*, 89, 21-29.
- Bishop, D. V., & Thompson, P. (2023). *Evaluating what works: An intuitive guide to intervention research for practitioners*. CRC Press.
- Borman, G. D., Benson, J. G., & Overman, L. (2009). A randomized field trial of the Fast ForWord Language computer-based training program. *Educational Evaluation and Policy Analysis*, 31(1), 82-106.
- Cabalo, J. V., Ma, B., & Jaciw, A. (2007). Comparative Effectiveness of Carnegie Learning's“ Cognitive Tutor Bridge to Algebra“ Curriculum: A Report of a Randomized Experiment in the Maui School District. Research Report. *Empirical Education Inc*.
- Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, 22(1), 31-72.
- Callaway, B., & Sant' Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of econometrics*, 225(2), 200-230.
- Campuzano, L., Dynarski, M., Agodini, R., & Rall, K. (2009). Effectiveness of Reading and Mathematics Software Products: Findings From Two Student Cohorts. NCEE 2009-4041. *National Center for Education Evaluation and Regional Assistance*.
- Cavalluzzo, L., Lowther, D. L., Mokher, C., & Fan, X. (2012). Effects of the Kentucky Virtual Schools' Hybrid Program for Algebra I on Grade 9 Student Math Achievement. Final Report. NCEE 2012-4020. *National*

- Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational evaluation and policy analysis*, 25(2), 119–142.
- Datnow, A., & Castellano, M. (2000). Teachers' responses to Success for All: How beliefs, experiences, and adaptations shape implementation. *American educational research journal*, 37(3), 775–799.
- Deault, L., Savage, R., & Abrami, P. (2009). Inattention and response to the ABRACADABRA web-based literacy intervention. *Journal of Research on Educational Effectiveness*, 2(3), 250–286.
- Di Pietro, G., & Muñoz, J. C. (2025). A meta-analysis on the effect of technology on the achievement of less advantaged students. *Computers & Education*, 226, 105197.
- Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L., & Sussex, W. (2007). Effectiveness of reading and mathematics software products: Findings from the first student cohort.
- Escueta, M., Nickow, A. J., Oreopoulos, P., & Quan, V. (2020). Upgrading education with technology: Insights from experimental research. *Journal of Economic Literature*, 58(4), 897–996.
- Faber, J. M., & Visscher, A. J. (2018). The effects of a digital formative assessment tool on spelling achievement: Results of a randomized experiment. *Computers & education*, 122, 1–8.
- Fairlie, Robert W., & Jonathan Robinson. (2013). Experimental Evidence on the Effects of Home Computers on Academic Achievement among Schoolchildren. *American Economic Journal: Applied Economics*, 5(3), 211–240.
- Feng, M., Huang, C., & Collins, K. (2023, June). Promising long term effects of ASSISTments online math homework support. *In International Conference on Artificial Intelligence in Education* (pp. 212–217). Cham: Springer Nature Switzerland.
- Fraillon, J., Schulz, W., & Ainley, J. (2013). *International Computer and Information Literacy Study: Assessment Framework*. International Association for the Evaluation of Educational Achievement.
- Funk, M. J., Westreich, D., Wiesen, C., Stürmer, T., Brookhart, M. A., &

- Davidian, M. (2011). Doubly robust estimation of causal effects. *American journal of epidemiology*, 173(7), 761-767.
- Geldsetzer, P., & Fawzi, W. (2017). Quasi-experimental study designs series—paper 2: complementary approaches to advancing global health knowledge. *Journal of clinical epidemiology*, 89, 12-16.
- Goolsbee, A., & Guryan, J. (2006). The impact of internet subsidies in public schools. *The Review of Economics and Statistics*, 88(2), 336-347.
- Hahn, J., Todd, P., & Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1), 201-209.
- Hegedus, S. J., Dalton, S., & Tapper, J. R. (2015). The impact of technology-enhanced curriculum on learning advanced algebra in US high school classrooms. *Educational Technology Research and Development*, 63, 203-228.
- Hirano, K., & Imbens, G. W. (2001). Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. *Health Services and Outcomes research methodology*, 2, 259-278.
- Honma, M., Masaoka, Y., Iizuka, N., Wada, S., Kamimura, S., Yoshikawa, A., Moriya, R., & Izumizaki, M. (2022). Reading on a smartphone affects sigh generation, brain activity, and comprehension. *Scientific reports*, 12(1), 1589.
- Hull, M., & Duch, K. (2019). One-to-one technology and student outcomes: Evidence from Mooresville's digital conversion initiative. *Educational Evaluation and Policy Analysis*, 4(1), 79-97.
- Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of econometrics*, 142(2), 615-635.
- Jackson, K., & Makarin, A. (2018). Can online off-the-shelf lessons improve student outcomes? Evidence from a field experiment. *American Economic Journal: Economic Policy*, 10(3), 226-254.
- Jeong, D. W., Moon, H., Jeong, S. M., & Moon, C. J. (2024). Digital capital

- accumulation in schools, teachers, and students and academic achievement: Cross-country evidence from the PISA 2018. *International Journal of Educational Development*, 107, 103024.
- Jeong, H. (2012). A comparison of the influence of electronic books and paper books on reading comprehension, eye fatigue, and perception. *The Electronic Library*, 30(3), 390–408.
- Kelly, K., Heffernan, N., Heffernan, C., Goldman, S., Pellegrino, J., & Soffer Goldstein, D. (2013). Estimating the effect of web-based homework. *In Artificial Intelligence in Education: 16th International Conference, AIED 2013, Memphis, TN, USA, July 9–13, 2013. Proceedings 16* (pp. 824–827). Springer Berlin Heidelberg.
- Kennedy-Shaffer, L. (2024). Quasi-experimental methods for pharmacoepidemiology: difference-in-differences and synthetic control methods with case studies for vaccine evaluation. *American Journal of Epidemiology*, 193(7), 1050–1058.
- Kimmons, R., Graham, C. R., & West, R. E. (2020). The PICRAT model for technology integration in teacher preparation. *Contemporary Issues in Technology and Teacher Education*, 20(1), 176–198.
- Kolen, M. J. & Brennan, R. L. (2014). Test equating, scaling, and linking: Methods and practice (3rd ed.). New York: Springer.
- Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of economic literature*, 48(2), 281–355.
- Leuven, E., Lindahl, M., Oosterbeek, H., & Webbink, D. (2007). The effect of extra funding for disadvantaged pupils on achievement. *The Review of Economics and Statistics*, 89(4), 721–736.
- Malamud, O., & Pop-Eleches, C. (2011). Home Computer Use and the Development of Human Capital. *Quarterly Journal of Economics*, 126(2), 987–1027.
- Mitchell, M. J., & Fox, B. J. (2001). The effects of computer software for developing phonological awareness in low-progress readers. *Literacy Research and Instruction*, 40(4), 315–332.
- Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The bookmark

- procedure: Psychological perspectives. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 249–281). Mahwah, NJ: Lawrence Erlbaum Associates.
- Morgan, P., & Ritter, S. (2002). An experimental study of the effects of Cognitive Tutor Algebra I on student knowledge and attitude. *Retrieved from the Carnegie Learning website: http://carnegielearning.com/web_docs/morgan_ritter_2002.pdf*.
- OECD. (2023a). *OECD Digital Education Outlook 2023: Towards an Effective Digital Education Ecosystem*. OECD.
- OECD. (2023b). *PISA 2022 results (Volume I): The state of learning and equity in education*. PISA, OECD.
- OECD. (2023c). *Organisation for Economic Co-operation and Development. (2023). PISA 2022 Assessment and Analytical Framework*. OECD.
- OECD. (2024). *Curriculum Flexibility and Autonomy: Promoting a Thriving Learning Environment*. OECD Publishing, Paris, <https://doi.org/10.1787/eccbbac2-en>.
- Ong-Dean, C., Huie Hofstetter, C., & Strick, B. R. (2011). Challenges and dilemmas in implementing random assignment in educational research. *American Journal of Evaluation*, 32(1), 29–49.
- Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014). Effectiveness of cognitive tutor algebra I at scale. *Educational Evaluation and Policy Analysis*, 36(2), 127–144.
- Pane, J. F., McCaffrey, D. F., Slaughter, M. E., Steele, J. L., & Ikemoto, G. S. (2010). An experiment to evaluate the efficacy of cognitive tutor geometry. *Journal of Research on Educational Effectiveness*, 3(3), 254–281.
- Reeve, J., & Cheon, S. H. (2021). Autonomy-supportive teaching: Its malleability, benefits, and potential to improve educational practice. *Educational psychologist*, 56(1), 54–77.
- Ritter, S., Kulikowich, J., Lei, P. W., McGuire, C. L., & Morgan, P. (2007). What evidence matters? A randomized field trial of Cognitive Tutor Algebra I. *Frontiers in artificial intelligence and applications*, 162, 13.
- Rockoff, J. E. (2015). Evaluation report on the School of One i3 expansion.

Unpublished manuscript. *New York, NY: Columbia University.*

- Roschelle, J., Feng, M., Murphy, R. F., & Mason, C. A. (2016). Online mathematics homework increases student achievement. *AERA open*, *2*(4), 2332858416673968.
- Roschelle, J., Shechtman, N., Tatar, D., Hegedus, S., Hopkins, B., Empson, S., ... & Gallagher, L. P. (2010). Integration of technology, curriculum, and professional development for advancing middle school mathematics: Three large-scale studies. *American Educational Research Journal*, *47*(4), 833-878.
- Rouse, C. E., & Krueger, A. B. (2004). Putting computerized instruction to the test: A randomized evaluation of a “scientifically based” reading program. *Economics of Education Review*, *23*(4), 323-338.
- Rutherford, T., Farkas, G., Duncan, G., Burchinal, M., Kibrick, M., Graham, J., ... & Martinez, M. E. (2014). A randomized trial of an elementary school mathematics software intervention: Spatial-temporal math. *Journal of Research on Educational Effectiveness*, *7*(4), 358-383.
- Sariyatun, S., Suryani, N., Sutimin, L. A., Abidin, N. F., & Akmal, A. (2021). The Effect of Digital Learning Material on Students’ Social Skills in Social Studies Learning. *International Journal of Instruction*, *14*(3), 417-432.
- Schenke, K., Rutherford, T., & Farkas, G. (2014). Alignment of game design features and state mathematics standards: Do results reflect intentions?. *Computers & Education*, *76*, 215-224.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Quasi-experiments: interrupted time-series designs. *Experimental and quasi-experimental designs for generalized causal inference*, 171-205.
- Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, *3*(3), 9-45.
- Silverman, R. D., Keane, K., Darling-Hammond, E., & Khanna, S. (2024). The Effects of Educational Technology Interventions on Literacy in Elementary School: A Meta-Analysis. Review of Educational Research.
- Smith, J. A., & Todd, P. E. (2005). Does matching overcome LaLonde’s critique of nonexperimental estimators?. *Journal of econometrics*, *125*(1-2), 305-353.
- Snipes, J., Huang, C. W., Jaquet, K., & Finkelstein, N. (2016). The Effects of

the Elevate Math Summer Program on Math Achievement and Algebra Readiness. *Society for Research on Educational Effectiveness*.

Styles, B., & Torgerson, C. (2018). Randomised controlled trials (RCTs) in education research—methodological debates, questions, challenges. *Educational Research*, 60(3), 255–264.

Tatar, D., Roschelle, J., Knudsen, J., Shechtman, N., Kaput, J., & Hopkins, B. (2008). Scaling up innovative technology-based mathematics. *The Journal of the Learning Sciences*, 17(2), 248–286.

Van Klaveren, C., Vonk, S., & Cornelisz, I. (2017). The effect of adaptive versus static practicing on student learning—evidence from a randomized field experiment. *Economics of Education Review*, 58, 175–187.

Wang, H., & Woodworth, K. (2011). *Evaluation of Rocketship Education's use of DreamBox Learning's online mathematics program*. Menlo Park: SRI International Center for Education Policy.

Wijekumar, K. K., Meyer, B. J., & Lei, P. (2012). Large-scale randomized controlled trial with 4th graders using intelligent tutoring of the structure strategy to improve nonfiction reading comprehension. *Educational Technology Research and Development*, 60, 987–1013.

Wijekumar, K., Meyer, B. J., Lei, P. W., Lin, Y. C., Johnson, L. A., Spielvogel, J. A., & Cook, M. (2014). Multisite randomized controlled trial examining intelligent tutoring of structure strategy for fifth-grade readers. *Journal of Research on Educational Effectiveness*, 7(4), 331–357.