

## 기계학습을 활용한 성취 가속도 개념 타당화

지정민 (한국교육과정평가원 부연구위원)\*

### 요약

본 연구는 학생의 학습 성장을 보다 정밀하게 포착하기 위한 지표로서, 성취 가속도 (Achievement acceleration) 개념의 타당성을 검증하고자 하였다. 기존 성장 참조 평가는 주로 성취 수준 변화량에 초점을 두어, 학생 성장의 미세한 방향성 변화나 동태적 특성을 충분히 반영하는 데 한계가 있었다. 성취 가속도는 비선형적이고 장기 의존적인 구조를 지닌 고차적 지표로, 이를 타당하게 측정하기 위해서는 해당 특성을 포착할 수 있는 특수한 모형이 요구된다.

이러한 특성에 부합하는 측정 도구로서, 본 연구는 비선형성, 장기 의존성, 구조화된 결측 패턴을 동시에 고려할 수 있는 Deep-LSTM 기반 시계열 모형을 분석에 적용하였다. 분석자료로는 대구교육종단연구(DELS) 2017-2024 자료의 수직적도 성취도 점수를 사용하였다.

분석 결과, Deep-LSTM 모형은 성취 가속도의 정의에 부합하는 예측 구조를 갖추고 있음이 확인되었다. 또한 장기 누적 성장 맥락 정보가 가속도 예측에 핵심적으로 기여함을 확인하였다. 나아가 준거타당도 검토에서, Deep-LSTM이 산출한 성취 가속도 지표는 기존 성취 수준과 변화율을 통제한 이후에도 이후 성취 변화율에 대한 추가적인 설명력을 보였다. 이는 성취 가속도가 기존 지표와 수렴하되 변별되는 독립적 지표임을 의미한다. 또한 성취 가속도가 성장 참조 평가에서 해석 가능하고 교육적으로 유용한 평가 정보로 활용될 수 있음을 여러 사례들로부터 확인할 수 있었다.

본 연구는 성장 궤적의 미세한 변화 구조를 반영한 가속도 지표가 성취 수준이나 단순 성장량을 넘어, 장기 교육종단자료에서 의미 있는 평가 정보로 기능할 수 있을 가능성을 보여준다.

주제어: 성장 참조 평가, 성취 가속도, 기계학습, 대구교육종단연구, 타당화

\* 제1저자, 교신저자, [jnjm0903@kice.re.kr](mailto:jnjm0903@kice.re.kr)

## I. 서론

2022 개정 교육과정은 표준화된 교육에서 벗어나 학습자 맞춤형 교육과 성장 참조 평가를 핵심 가치로 설정하였다(교육부, 2022). 이는 타인과의 비교를 통한 서열화에서 탈피하여, 학습자의 과거 성취 수준 대비 현재의 향상도를 측정함으로써 모든 학생의 실질적 성장을 보장하고자 하는 의도로 해석된다. 이러한 교육적 지향점을 구현하기 위해서는 성장을 과학적으로 측정하고, 성장을 저해하거나 촉진하는 요인을 정밀하게 진단할 수 있는 평가 도구의 개발이 같이 진행되어야 한다.

기존의 성장 평가 지표는 주로 고전검사이론(Classical Test Theory, CTT)이나 선형 회귀 분석에 의존해 왔다. 선형 모델은 학습자의 발달이 일정한 속도로 진행된다는 가정 하에 설계되어, 실제 교육 현장에서 발생하는 비선형적 발달-정체기나 비약적 도약-을 포착하는 데 한계가 있다. 또한, 기존의 횡단적 데이터는 집단 간 비교에는 유용하나, 개별 학습자의 시계열적 변화를 추적하여 성장 궤적을 산출하는 데는 근본적인 제약이 따른다. 따라서 개정 교육과정이 지향하는 개별화된 성장을 실질적으로 측정하기 위해서는 종단적인 데이터 확보와 이를 분석할 고도화된 비선형 모형이 요구된다.

또한 기존 평가는 특정 시점의 성취도를 측정하는 정적인 방식에 머물러 있어, 학습 결손이 발생한 이후에야 개입이 이루어지는 사후 대처적 성격이 강했다. 진정한 의미의 맞춤형 교육을 달성하기 위해서는 학습자의 성장 궤적을 실시간으로 모니터링하고, 기대 궤적에서의 이탈 징후를 사전에 포착하여 예방적 개입을 제공하는 동적 평가 체제로 전환되어야 한다. 본 연구에서는 이를 위해 Deep-LSTM(Long Short-Term Memory) 기반의 분석법을 제안하고자 한다. AI 기반 시계열 학습 모형은 학습자의 성장 패턴을 학습하여 개별화된 예측을 제공할 수 있으며, 이를 통해 교육적 개입이 필요한 시점에 대한 단서를 제공할 가능성이 있다.

본 연구는 성적의 결과치를 예측하는 것을 넘어, 성장의 속도와 가속도라는 역동적 지표를 도입하여 학습 지원의 시의성을 확보하고자 한다. 최근 추진되고 있는 AI 디지털 교과서 보급은 학습자의 실시간 학습 데이터를 확보할 수 있는 최적의 환경을 제공한다. 이러한 교육 도구의 변화와 더불어 실시간 피드백이 가능한 AI 기반 성장 예측 지표의 개발은 학습자 개인의 성취 궤적 이탈을 조기에 진단하고 교육적 자원을 적시에 제공하는 데 기여할 것이다. 성취 가속도의 타당화는 AERA(American Educational Research Association) 외(2014)가 제시하는 타당도 증거 유형 중 예측 준거와의 관계에 초점을 둔 준거 관련 증거를 중심으로 검증하였다. 또한 모형이 가속도 예측에 활용하는 정보의 구조를 순열 중요도 분석과 절제 실험을 통해 검토하였다. 본 연구는 학습

자의 성장 궤적을 측정하는 지표로서 가속도를 점검하고 이를 검증하는데 Deep-LSTM 모형을 활용하여, 다음과 같은 연구 문제를 설정하였다.

연구 문제 1. 성취 가속도의 특성(비선형성, 장기 의존성)을 타당하게 측정하기 위해 요구되는 모형 조건은 무엇인가?

연구 문제 2. 성취 가속도 예측에 기여하는 핵심 정보는 무엇인가?

연구 문제 3. 성취 가속도 지표는 기존 성취 수준 및 변화율과 변별되는 독립적 지표로서 준거타당도를 갖는가?

## II. 성장 참조 평가와 기계학습

### 1. 성장 참조 평가

많은 연구들이 이전의 성취가 이후의 성취에 지대한 영향을 미친다고 보고하고 있다 (Bloom, 1976; Caspi et al., 1998; Duncan et al., 2007). 과거 성취의 영향력에 대해서는 크게 두 가지 상반된 이론적 관점이 제시되어 왔다. 첫 번째는 누적이득가설 (Cumulative Advantage Hypothesis)로, 초기 성취 수준이 높은 학생일수록 더 우수한 교육 자원과 긍정적 학습 경험을 확보함으로써 시간이 지남에 따라 성취 격차가 확대되는 현상을 설명한다. 이러한 현상은 흔히 ‘부익부 빈익빈’으로 표현되는 매튜 효과 (Matthew effects)로도 알려져 있으며, 다수의 연구자들이 이를 교육적 맥락에 적용해 왔다(Caspi et al., 1998; DiPrete & Eirich, 2006; Stanovich et al., 1986). 예컨대 Stanovich 외(1986)는 읽기 능력에서의 초기 격차가 시간이 흐를수록 전반적인 인지 능력의 격차로 확장되는 과정을 규명하였고, Caspi 외(1998)는 과거 성취가 미래 성취를 예측하는 가장 강력한 요인임을 주장하였다. 이러한 결과는 초기 성취 수준이 높은 학생들이 더 높은 학업적 자기효능감을 형성하고, 그로 인해 성취 격차가 점진적으로 심화되는 과정으로 해석될 수 있다(DiPrete & Eirich, 2006).

반면, 또 다른 시각은 평균회귀(Regression to the Mean)의 관점이다. 평균회귀는 극단적인 성취가 이후 측정에서 평균에 가까워지려는 통계적 경향을 강조하며, 이는 성취의 일부가 일시적 요인이나 측정 오차에 의해 과대·과소 추정되었을 가능성을 전제로 한다. 즉, 초기의 매우 높은 혹은 매우 낮은 성취는 다음 시점의 측정에서 자연스럽게 전체 평균 수준으로 수렴할 가능성이 크다는 것이다. Lohman과 Korb(2006)는 초기에

영재로 분류된 학생들의 성취가 시간이 경과함에 따라 상대적으로 평균으로 회귀하는 현상을 평균회귀의 원칙으로 설명하였으며, Downey(2020)는 학교 교육이 성취 격차를 확대하기보다는, 방학 기간 동안 발생한 격차를 학기 중에 완화·수렴시키는 역할을 수행한다고 주장하였다. 이러한 논의는 성취 변화가 구조적 누적 효과라기보다는 일시적 변동의 결과일 수 있음을 시사한다.

이와 같이 학생의 성취는 이전 성취에 대한 경로 의존적 누적의 결과일 수도 있고, 평균으로 회귀하는 통계적 현상의 산물일 수도 있다. 그러나 두 이론은 모두 단일 시점의 성취 수준만으로는 성취 변화를 이해하기 어렵다는 한계를 지닌다. 이에 따라 성취 변화의 방향과 속도를 보다 정밀하게 포착할 수 있는 개념의 도입이 필요하다. 이러한 문제의식 속에서 교육계는 학생의 학습 성장을 보다 정교하게 반영할 수 있는 새로운 평가 지표의 개발과, 이를 가능하게 하는 분석 기술의 도입을 요구받고 있다.

일반적으로 교육평가의 종류는 학생의 성취 수준을 타인과 비교하는 규준참조평가(Norm-referenced assessment)와, 사전에 설정된 기준의 도달 여부를 판단하는 준거참조평가(Criterion-referenced assessment)로 구분된다. 이에 비해 성장 참조 평가는 외부 기준이나 타인과의 비교가 아니라, 학생 개인의 현재 성취를 과거의 자기 자신과 비교함으로써 학생의 학습 성장에 초점을 둔다. 이러한 평가는 개별화된 학습 성장을 강조하고 학습 동기를 촉진한다는 점에서, 2022 개정 교육과정이 지향하는 맞춤형 교육의 방향성과 부합한다. 특히 하위 성취 학생에게도 성장이라는 성공적인 경험을 제공함으로써 학업적 자기효능감을 제고할 수 있다는 점에서 큰 교육적 의의를 지닌다.

다만 성장 참조 평가가 실질적으로 기능하기 위해서는 누적이득가설이 시사하는 교육 불평등의 가능성을 고려할 필요가 있다. 상위 성취 학생은 반복적인 긍정적 피드백을 통해 성장의 가속이 가능하지만, 하위 성취 학생은 상대적으로 완만한 성장 궤적을 보일 수 있다. 이러한 구조적 차이가 성취 자료에 내재되어 있다는 점을 간과해서는 안 된다. 더불어 평균회귀라는 통계적 현상 역시 성장 평가에서 중요한 고려 요소이다. 측정 오차로 인해 일시적으로 높거나 낮은 점수를 받은 학생은 이후 측정에서 점수 변동이 나타날 가능성이 크기 때문이다. 또한 성장 참조 평가는 저성취 학생에게 과도한 낙인 효과를 유발하거나, 반대로 단기적 개선에 대한 과도한 낙관을 초래할 위험도 내포한다. 따라서 단일 시점 간 변화량보다는 시간적 맥락을 반영한 장기적 성장 지표의 활용이 요구된다. 이를 위해 다회차의 종단적 성취 자료를 수집하고, 통계적 오차를 통제함으로써 측정의 타당성을 확보하려는 노력이 병행되어야 한다.

이러한 맥락에서 성장 참조 평가의 핵심적 의의는 단순한 성취 수준의 비교를 넘어, 학생의 성취 변화 양상을 시간의 흐름 속에서 해석할 수 있는 평가 정보를 제공하는 데 있다. 특히 초기 성취 수준이 유사하더라도 성장 속도나 변화 패턴이 상이한 학습자를

구분함으로써, 성취가 정체되거나 둔화 국면에 진입한 학생을 보다 정밀하게 식별할 수 있다. 이러한 성장 변화 중심의 평가는 형성 평가의 관점에서 학습자의 성취를 지속적으로 모니터링하고, 일반 교육과정 내에서 성장이 지연되는 학생을 조기에 발견하여 적절한 교육적 지원을 설계하는 데 기여할 수 있다(Reschly & Bergstrom, 2009).

한편, 최근 심재권(2025)은 대구교육중단연구 자료를 활용하여 CatBoost 알고리즘 기반의 학업부진 예측 모형을 제시하고, 학습자를 성취 변화 유형에 따라 분류하여 각 유형에 속할 확률을 산출하였다. 이러한 연구는 학생의 학업부진 가능성을 사전에 탐색하고, 학습자의 성취 변화 양상을 유형화하는 데 유용한 시사점을 제공한다. 이에 본 연구는 Deep-LSTM 모형을 활용하여 단순 성취 수준이나 변화량을 넘어, 시간에 따른 성장 변화의 동태적 특성을 반영할 수 있는 성취 가속도를 성장 참조 평가의 지표로 제안하고자 한다. 이는 개인별 성취 궤적의 장기적 맥락과 비선형적 변화를 통합적으로 반영함으로써, 이산적인 시점 간 비교의 한계를 보완하고 연속적인 성장 과정을 해석가능한 평가 정보로 확장하려는 시도라는 점에서 기존 연구와 차별성을 지닌다. 비록 성장 참조 평가 관련 연구는 다수 존재하지만, 성취 가속도라는 고차 변화량을 평가 지표로 다루고 이를 예측 기반으로 검증한 연구는 아직 드물다. 따라서 본 연구는 성취 가속도라는 고차 변화량을 평가 지표로 제시함으로써 성장 참조 평가 관련 연구의 지평을 넓히고, 또한 예측된 가속도가 성장 참조 평가에서 지니는 의미를 검증하고자 한다.

## 2. 가속도의 개념과 성장 참조 평가 지표로서의 장점

교육평가연구에서 성장은 보통 특정 시점의 성취 수준, 혹은 시점간 변화량으로 정의되어 왔다(Betebenner & Linn, 2010; Scammacca et al., 2020). 따라서 성장 참조 평가는 동일 시점의 준거 비교와는 달리 개인 내 성취 변화에 초점을 두었다. 하지만 기존 성장 지표는 학생들간의 점수 차이, 평균 성장률, 선형 또는 저차 다항식 기반 성장계수에 의존해 왔다(Singer & Willett, 2003; Bollen & Curran, 2006).

성적 변화량은 얼마나 빨리 성장하는가를 나타내는 지표로 특정 시점의 성취 수준보다 학습의 역동성을 잘 설명한다는 점에서 의미가 있지만, 변화량 자체가 시간에 따라 일정하다고 가정되기 쉽다는 점에서 한계가 있다. 다시 말해, 두 학생이 동일한 성장률을 보여도 한 학생은 가속, 다른 학생은 감속 국면일 수 있는데, 변화량만으로는 이를 구분할 수 없다. 성적 변화량은 성장의 현재 상태는 설명할 수 있으나 보다 긴 맥락에서 성장 궤적의 방향성 변화는 충분히 포착하지 못한다는 한계가 있다.

따라서 본 연구에서 제시하는 성장 지표로서의 성취 가속도는 변화량의 변화, 즉 성장이 얼마나 빨라지거나 느려지고 있는가를 나타내는 지표로 학습이 촉진되는 국면과

정체되는 국면을 구분할 수 있는 지표라는 점에서 의의가 있다. 이는 단일 시점이나 단기 변화가 아닌 시간적 맥락 속에서 누적된 성장 패턴의 구조적 변화를 반영한다.

성취 가속도는 성장 참조 평가 지표로서 다음과 같은 장점이 있다. 먼저, 가속도는 미래 성취 변화를 예측하는 조기 신호가 될 수 있다. 가속도를 통해 우리는 향후 학생의 성장 경로에 대한 예측을 할 수 있으며, 교육적 개입시점을 적시에 판단할 수 있다. 평균 성적이나 성적 변화량의 단순 기울기만 보면 아직 괜찮은 학생들도, 성취 가속도가 음으로 꺾이는 순간 이미 성적 하락의 전조가 나타난 경우가 많다. 학업 중단이나 이탈은 갑작스런 사건이라기보다 점진적으로 전개되는 경우가 많으며, 완전한 학업 이탈 이전에 이를 알리는 신호가 오랫동안 선행된다는 점이 일관되게 보고된다(Balfanz et al., 2007; Henry et al., 2012). 성취 가속도는 이러한 기울기의 꺾임을 수량화해서 조기 경보 시스템의 정밀도를 높여줄 수 있다.

또한, 성취 가속도는 현실적인 학생 성적 변화의 비선형성을 반영한다. 학생의 학습은 선형적으로 나타나지 않는다. 급성장기, 정체기, 준비기, 재도약기 등 다양한 차원의 비선형적 모습들이 등장한다. 미국 학생 대다수가 반복적 재교육에 머물러 있다는 지적이 있는데(TNTP, 2018), 이는 속도만 측정하는 평가 체계가 성장 궤적의 질적 변화를 놓칠 수 있다는 문제와 맞닿아 있다.

마지막으로, 성취 가속도 지표는 교육 격차 분석의 정교화를 실현한다. 학생의 격차가 벌어지고 있는지, 좁혀지고 있는지를 성취 가속도를 통해서 알 수 있다. 박희진과 남궁지영(2019)의 연구는 중·고등학생의 수학성적 변동에 부모 SES보다 학교 수준 요인(학교 평균 SES, 교사의 교육기대 등)이 시간이 지남에 따라 더 중요해짐을 보여주었다. 이런 변동 분석은 본 연구가 제안하는 가속도 지표의 개념적 토대와 맞닿아 있으며, 가속도 지표를 적용하면 이러한 궤적의 질적 변화를 보다 정교하게 포착할 수 있다.

### 3. 가속도 지표를 위한 모형

교육적 맥락에서 중단 데이터는 주로 학생의 성취도, 심리 요인, 사회적 배경 등의 ‘변화’를 추적하기 위해 활용되어 왔다. 학습자의 성장을 측정하기 위한 대표적인 분석 기법으로는 잠재성장모형(Latent Growth Modeling, LGM)과 성장혼합모형(Growth Mixture Modeling, GMM)이 널리 쓰여왔다(Meredith & Tisak, 1990; Muthén, 2001). 잠재성장모형은 모든 학생이 하나의 평균적인 성장 궤적을 공유한다는 전제하에 개인 간 차이를 분석하는 방식이며, 성장혼합모형은 집단 내에 서로 다른 성장 패턴을 지닌 다수의 잠재 계층이 존재함을 가정하고 이를 모형화하는 방법이다.

관련 선행연구를 살펴보면, 백수진 외(2023)는 청소년기 학업 스트레스의 변화를 성

장혼합모형과 SEM Tree를 통해 비교 분석하였고, 김수정 외(2020)는 잠재성장모형을 적용하여 학교에 대한 학생 인식의 종단적 변화를 고찰하였다. 국외 연구에서는 McArdele(2009)이 교육 및 심리학 데이터의 종단적 변화를 모형화하는 과정을 잠재성장모형을 통해 체계적으로 제시하였으며, Muthén과 Muthén(2000)은 성장혼합모형을 활용하여 집단별 궤적을 분석하는 방법론적 기틀을 마련하였다. 또한 서미정(2009)은 초기 청소년의 공격성 및 비행 변화를, 김소영과 윤기봉(2016)은 청소년 삶의 만족도 변화와 그 개인차에 영향을 미치는 요인을 각각 잠재성장모형으로 검증하였다. 노언경 외(2017) 역시 다변량 다층 잠재성장모형을 사용하여 학교폭력 가해 및 피해 경험의 변화 양상을 추정한 바 있다.

이러한 전통적 모형들은 현상에 대한 높은 설명력을 제공한다는 장점이 있으나, 동시에 몇 가지 측면에서 보완될 수 있다. 첫째, 전통적 모형들은 선형 혹은 이차 함수에 의존하기에 학습 성장의 불규칙하고 역동적인 측면을 반영하는데 한계가 있다. 둘째, 대규모 수준의 비정기적 혹은 실시간 단위 데이터를 처리하는 데 어려움이 있다. 셋째, 모델 예측값과 실제값의 차이인 잔차를 단순한 측정 오차로 간주하여 분석에서 제외한다. 마지막으로, 대개 연구 종료 후 사후 분석을 목적으로 시행되기에 학습 과정에서 발생하는 미세한 징후를 포착하여 실시간으로 개입하는 데에는 한계가 존재한다.

전통적 잠재성장모형에서 가속도는 2차 성장요인(quadratic slope)으로 간접적으로 표현된다. 그러나 이러한 접근은 성장 궤적 전체를 저차 다항식으로 요약하고 개인별 국소적 가속, 감속 패턴을 직접적으로 추정하기 어렵다는 점에서 한계가 있다. 또한 관측 시점이 불규칙하거나 결측이 많은 장기 교육자료에서는 가속도 추정의 안정성이 저하될 수 있다. 따라서 성장 지표로서 가속도는 현재 성취가 유사한 학생간 성장 방향성의 차이를 식별하여, 향후 성취 변화에 대한 예측적 단서를 제공할 수 있어야 한다. 다만 가속도는 직접 관측되는 값이 아니라 시계열 구조를 통해 추정되어야 하는 잠재적 지표로서 안정적 추정 방법이 선행되어야 한다. 본 연구에서는 이러한 맥락에서 장기 종단자료에 적합한 기계학습 기반의 시계열 모형을 통해 가속도를 새로운 성장 참조 평가지표로 제안하고자 한다.

#### 4. 가속도 지표를 위한 모형으로서 기계학습의 활용

기계학습이란 인공지능의 한 분야로서, 수십년간 급격한 성장을 거두어왔다. 초기 기계학습은 컴퓨터 알고리즘을 통한 문제 해결에 주력했으며, 1967년에는 패턴 인식의 기반이 된 최근접 이웃(K-Nearest Neighbor) 기법이 등장하며 그 지평을 넓혔다(Alzubi et al., 2018). 특히 2006년 딥러닝 기술의 본격적인 도입은 인공지능이 기업의 실질적

인 문제 해결 도구로 자리 잡는 결정적인 전환점이 되었다(Alzubi et al., 2018).

교육분야에서도 다양한 교육 데이터를 이용한 예측 모형의 적용이 활발히 이루어지고 있다(Baker & Inventado, 2014; Romero & Ventura, 2020). 특히 많은 연구들이 교육적 자료에 기계학습을 결합하여 모형의 예측력을 높이는 데에 많은 관심을 가져왔다. Baker와 Inventado (2014)는 교육에서의 데이터 마이닝의 적용과 그 발전을 분석하였다. 이 논문은 예측 모형, 구조 발견, 관계 마이닝, 그리고 모형을 활용한 발견으로 나누어서 교육분야에서의 데이터 마이닝의 발전을 알아보았다. 회귀분석, 분류 모형, 클러스터링 등 다양한 기계학습 관련 모형들이 교육자료를 분류하고 예측하기 위해 사용되었다.

RNN(Recurrent Neural Network)은 일반적인 인공신경망이 데이터 사이의 순서를 고려하지 않는 것과 달리, 과거의 정보를 취합하여 미래를 예측한다. 즉, 현재시점의 입력값과 이전시점의 은닉상태(과거의 정보를 압축한 기억 형태로 새로운 입력이 들어올 때마다 갱신된다)를 결합하여 현재의 은닉상태를 계산한다. 하지만, RNN의 경우 장기 의존성라는 치명적인 약점이 존재한다. 이는 시계열 데이터가 길어질수록 앞부분의 정보가 점차 희석되고, 가중치를 갱신할 때 앞쪽까지 전달되는 신호가 0에 가까워져 기울기가 소실되어 학습이 거의 이루어지지 않게 되는 현상이다.

이를 해결하기 위해 도입된 LSTM(Long Short-Term Memory)은 ‘긴 단기 기억’이라는 의미로 시간에 따라 변하는 정보를 기억하고 학습하는 학습 방법을 뜻한다. LSTM은 시계열 자료를 처리하기 위한 기존 RNN의 변형 알고리즘으로서 인간의 다양하고 복잡한 행동을 인식한다. LSTM은 RNN의 정보 손실을 보완하기 위해 만들어진 방법으로서 정보를 선택적으로 기억하고 삭제하여 긴 데이터라도 앞부분의 중요한 정보를 보존할 수 있다. 다시 말해, 무엇을 잊을 것인지, 무엇을 새로 기억할 것인지, 무엇을 다음 층으로 출력할 것인지를 함수를 통해 결정한다.

이러한 LSTM은 하나의 은닉층을 통해 원본 데이터에서 중요한 정보를 찾아내는 역할을 한다. 그러나 하나의 은닉층은 단순 구조로 인해 복잡한 데이터를 처리하기에는 여전히 한계가 있었고, 이를 보완하기 위해 Deep-LSTM이 제안되었다. Deep-LSTM은 여러 은닉층을 수직으로 쌓아 올림으로써 자료의 계층적 특징을 학습할 수 있기에, 복잡하고 다층적인 자료를 파악할 수 있다. Deep-LSTM에서는 한 층의 출력이 다음 층의 입력이 된다. 예를 들어, 입력층을 통해 첫 번째 은닉 상태를 계산하면, 첫 번째 층의 출력이 두 번째 층의 입력값이 되고, 두 번째 층의 출력이 세 번째 층의 입력값이 된다. 이러한 과정을 거쳐 마지막 층의 출력값이 최종 예측값을 도출한다. 층이 깊어질수록, 복잡한 시계열 데이터의 비선형 관계를 더 잘 근사할 수 있으며, 예측 오차를 유의미하게 줄일 수 있다.

이 방법은 다수 개체의 종단 자료가 누적되어 충분한 학습 표본을 확보할 수 있는 상

황에서 잘 작동한다. 본 연구는 Deep-LSTM이 종단자료에 적합한 이유로 자료의 불규칙성, 장기 누적 효과, 구조적 결측치 존재를 제시한다. 본 연구는 Deep-LSTM이 이와 같은 장기적이고 불규칙적인 자료를 잘 보완할 수 있을 것이라 기대한다.

### III. 분석 방법

#### 1. 대구교육종단연구의 주요 내용

본 연구에서는 대구교육종단연구(Daegu Educational Longitudinal Study, DELS)에서 구축된 수직척도 성취도 점수를 학생 성장의 준거 지표로 채택 및 활용하였다. 연구 대상은 대구교육종단연구 2017-2024 패널 데이터에 포함된 학생들이다. 대구교육종단연구는 대구 지역 학생들의 인지적, 정의적 특성, 가정 배경, 학교 환경 및 학업 성취도 등 다양한 교육적 요인을 장기적으로 추적하여 학생의 변화에 미치는 교육적 영향을 체계적으로 분석하기 위해 구축된 종단 자료이다. 특히 본 연구에서 활용한 수직척도 점수는 학년간 성취 수준을 동일한 척도 상에서 비교할 수 있게 함으로써, 개별 학습자의 연속적인 성장 궤적을 정밀하게 산출하는데 적합한 지표라 할 수 있다. 본 연구는 대구광역시교육청 대구미래교육연구원의 데이터 활용 승인을 거쳐 제공된 자료를 사용하였으며, 모든 분석은 식별정보가 제거된 익명화 데이터를 대상으로 수행되었다(대구광역시교육청 대구미래교육연구원, 2024).

<표 1> 대구교육종단연구의 주요 내용

| 구분     | 주요 내용                                     | 비고             |
|--------|---|----------------|
| 조사 목적  | 대구 학생들의 성장에 미치는 교육적 영향 분석 및 정책 수립 기초자료 제공 |                |
| 조사 설계  | 초4, 중1, 고1 시점의 3개 패널을 대상으로 한 장기 추적 조사     |                |
| 연구 패널  | 2017 중 패널(2017년 당시 중학교 1학년 학생)            |                |
| 분석 기간  | 2017년(1차) ~ 2024년(8차)                     | 총 8개년도 시계열 데이터 |
| 조사 영역  | 인지적 성취도(수학), 심리적 특성, 가정 환경, 학교 생활 등       | 다각적 배경 변인 포함   |
| 조사 대상  | 매년 대구 지역 내 약 3,000여 명의 학생 및 해당 학부모        | 패널별 상이         |
| 성취도 측정 | 수직척도 점수 활용                                | 학년 간 성취 비교 가능  |

〈표 1〉은 대구교육종단연구 자료의 주요 구성 내용을 제시하고 있다. 본 자료는 학생의 심리적 특성, 가정환경, 학교생활 요인 등 다각적인 조사 영역을 포함하며, 교과 성취도에 대한 종단적 조사를 수행한다.

본 연구에서는 분석 자료의 신뢰성을 확보하기 위해 가속도 산출에 필요한 최소 3개년 연속 성취도 자료가 확보된 학생만을 분석 대상으로 포함하였다. 우선 수직척도 점수가 누락되었거나 가속도 산출에 필수적인 3개년 연속 성취도 자료가 확보되지 않은 사례를 제외하였다. 모형 학습 과정에서는 결측 마스크를 입력에 반영하여 결측 패턴을 정보로 활용하였다.

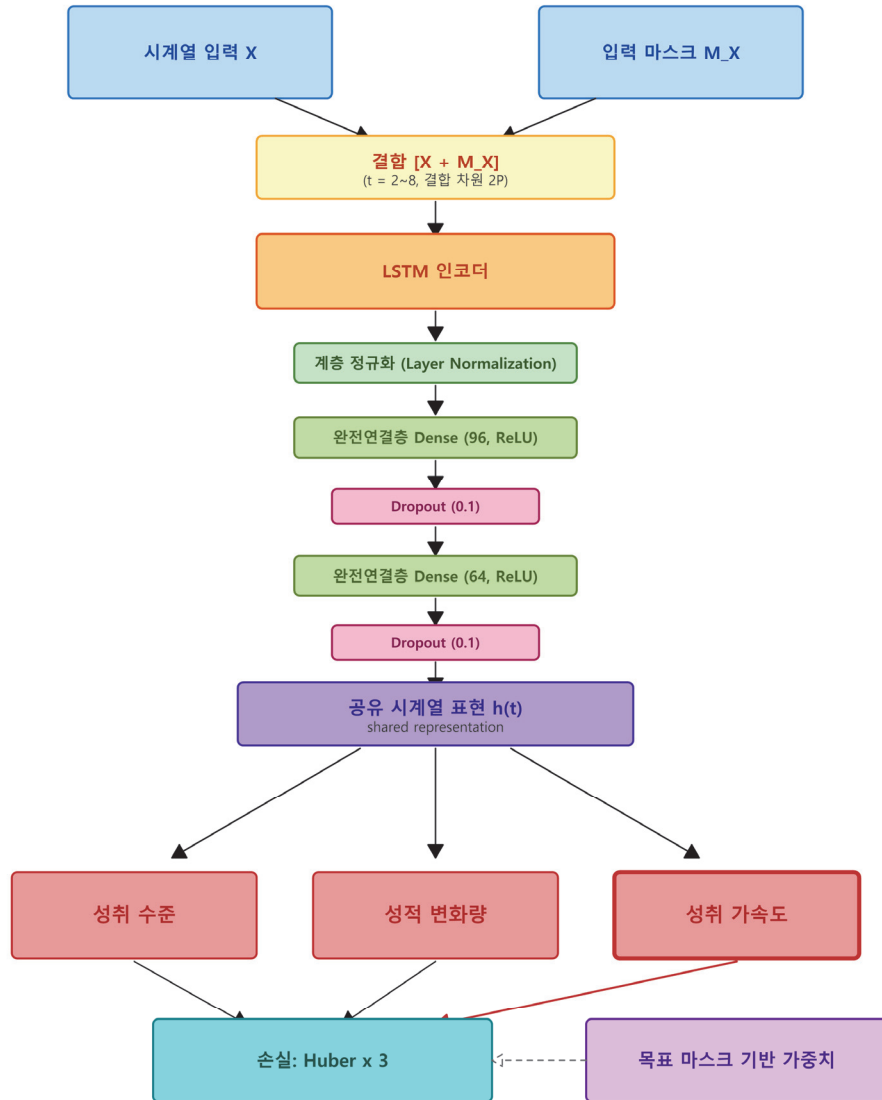
## 2. Deep-LSTM의 이론과 적용

본 연구는 장기 교육종단자료에서 성취 가속도를 예측하기 위해 Deep-LSTM 기반 시계열 모형을 적용하였다. 분석 단위는 학생 시점이며, 입력 시계열은 각 시점  $t$ 에 대해 과거 관측을 누적적으로 활용하며, 예측 목표는 다음 시점( $t+1$ )의 성취 수준, 성적 변화량, 성취 가속도이다. 입력 구성은  $t=2\sim 8$  시점의 관측값을 활용하되, 예측 대상이 관측 가능한  $t+1$  시점이 있는 경우( $t=2\sim 7$ )에만 손실에 기여하도록 마스크로 처리하였다. 즉, 한 학생의 여러 시점에 걸친 과거 성취 흐름을 바탕으로, 다음 시점의 성장 상태와 변화 방향을 예측하는 구조이다.

자료는 수직척도 점수를 기반으로 구성하였다. 각 시점의 성취 수준은 해당 연도의 수직척도 점수이며, 변화율은 인접 시점 간 점수 차이, 가속도는 변화율의 차이로 계산하였다. 이후 각 시점  $t$ 의 관측값으로부터 다음 시점의 수준, 변화율, 가속도를 생성하였다. 장기 종단자료에서 발생하는 결측은 제거하지 않고 학습 과정에 반영하였다. 구체적으로, 입력 단계에서는 각 시점·변수의 결측 여부를 이진 마스크  $M_X$ 로 구성하여 관측값  $X$ 와 함께 모형에 입력함으로써 모형이 결측 자체를 하나의 정보로 학습하도록 하였다. 또한 손실 계산 단계에서는 목표값의 결측 마스크를 가중치로 활용하여, 실제로 관측된 시점만 손실에 기여하도록 하였다.

Deep-LSTM의 입력은 직전 점수만을 사용하는 대신, 장기 누적 맥락과 변화 이력을 함께 반영하도록 확장하였다. 입력 특성에는 누적평균 기반의 장기 정보, 여러 시점의 지연값, 개인 수준 요약 통계(개인 평균 및 분산), 그리고 시간항( $time, time^2$ )이 포함된다. 이를 통해 모형은 학생들의 비선형 성장 패턴과 학생 개인 간 차이를 동시에 포착하고자 하였다. 또한 각 시점별 배경 변인은 결측률 98% 미만, 고유값 2개 이상의 조건을 충족하는 변수 중에서 결측률이 낮은 순으로 최대 30개를 선별하였으며, 장기 맥락 정보(누적평균, 지연값, 개인 요약통계)를 포함한 전체 입력 차원은 최대 160개로 제한하였다.

모델 입력에서 결측값은 0으로 대체하되, 각 변수에 대해 관측 여부를 나타내는 입력 마스크를 함께 생성하여 값 정보와 관측 정보를 동시에 학습에 투입하였다. 이로써 시간적으로 반복되거나 누적되는 결측 패턴 자체가 예측에 활용될 수 있도록 설계하였다. [그림 1]은 이러한 입력 구성과 공유 인코더, 다중 출력 구조를 개념적으로 나타낸 것이다.



[그림 1] 다중과제 Deep-LSTM의 구조

다중과제 학습의 안정성을 높이기 위해, 목표변수(성취 수준, 성적 변화량, 성취 가속도)는 학습 집합을 기준으로 1/99 백분위 원저화로 극단값(전체 관측치의 약 2%)을 완

화한 뒤 z-점수 표준화를 적용하여 학습에 사용하였다. 성능 평가는 예측치를 원래 단위로 역변환한 뒤 RMSE(Root Mean Squared Error, 평균 제곱근 오차)와 MAE(Mean Absolute Error, 평균 절대 오차)를 산출하였다. 두 지표 모두 값이 작을수록 오차가 작음을 나타내며, RMSE는 큰 오차에 더 민감하게 반응하여 극단적 오차의 영향을 함께 반영하고, MAE는 모든 오차에 동일한 가중치를 부여하여 평균적 정확성을 보여준다.

모형은 입력 시계열로부터 공유 시계열 표현을 학습한 뒤, 이를 바탕으로 세 개의 목표 변수를 동시에 예측하는 다중과제 Deep-LSTM 구조로 구현하였다. 먼저, LSTM 층을 통해 시계열 정보를 추출한 후, 정규화와 비선형 층(Dense)을 거쳐 표현력을 강화하였다. 과적합을 방지하기 위해 Dropout과 조기종료를 적용하였고, 손실함수는 Huber loss, 최적화는 Adam 알고리즘을 사용하였다.

자료는 학생 ID 단위로 층화 분할하여 학습 집합과 테스트 집합을 80:20 비율로 구성하였다. 학습 집합 내부에서는 10%를 검증 집합으로 분할하여 조기종료와 하이퍼파라미터 탐색에 활용하였으며, 테스트 집합은 최종 성능 평가에만 사용하였다. 동일 학생의 여러 시점 관측치가 분할 집합 간에 누출되는 것을 방지하기 위해 학생 단위 분할을 적용하였다.

주요 하이퍼파라미터는 LSTM 기반 시계열 예측 선행연구(Reimers & Gurevych, 2017)에서 일반적으로 활용되는 범위를 참조하여 은닉 단위 수 48, Dense 층 구성 96→64, Dropout 비율 0.1, 배치 크기 256, 반복 횟수 50으로 설정하였다. 학습 집합 내부의 10%를 검증 자료로 분할하여 조기종료의 기준으로 활용하였으며, 이를 통해 모형의 과적합을 방지하였다.

학습은 5개의 무작위 시드로 반복 수행하여 성능의 평균과 변동성을 함께 보고하였다. 특히 본 연구의 핵심 목표인 가속도 예측을 강조하기 위해, 손실 가중치를 성취 수준 0.2, 성적 변화량 0.5, 성취 가속도 3.0으로 부여하여 가속도 예측에 가중치를 집중하였다.

비교 기준으로는 전통적 성장모형인 잠재성장모형을 사용하였다. 여러 모형 중 잠재성장모형을 선택한 이유는 본 연구의 비교 대상이 성장 궤적 전체를 구조화된 형태로 추정하는 전통적 성장모형이기 때문이다. ARIMA 계열 모형은 본질적으로 단일 개체(N=1)의 긴 시계열을 모델링하기 위한 도구로, 본 연구처럼 다수 학생의 짧은 패널 자료를 통합 분석하는 구조와는 부합하지 않는다. 다층성장모형은 잠재성장모형과 수학적으로 동형이므로 별도 비교를 수행하지 않았다. 반면, 잠재성장 모형은 개인 수준 성장 궤적을 구조 모형 형태로 명시적으로 표현하며, 교육 평가 연구에서 표준적으로 사용되어 온 대표적 방법이라는 점에서 비교 기준으로 타당하다고 보았다.

잠재성장모형은 관측된 시점을 활용해 성장 요인을 추정하며, 2차 모형을 우선 적용

하되 수렴이 불안정할 경우 선형 모형으로 대체하는 절차를 따랐다. 그리고 본 연구 자료에서는 2차 모형의 수렴이 불안정하여 최종적으로 선형 모형을 채택하였다. 적합한 잠재성장모형으로부터 시점별 예측 점수를 산출한 뒤 변화율과 가속도를 계산하여, 동일한 테스트 자료에서 Deep-LSTM과 가속도 예측 성능을 비교하였다.

마지막으로, 본 연구가 가속도를 단일 과제로 학습하지 않고 성취 수준과 성적 변화량을 함께 학습한 이유는, 가속도가 이들 하위 성장 지표의 시간적 축적 위에서 만들어지는 고차 지표이기 때문이다. 본 연구는 모든 성취 가속도 관련 성장 정보를 모형이 함께 학습하도록 설계함으로써, 가속도 예측에 필요한 신호를 보다 안정적으로 추출하고 장기 종단자료의 비선형적 성장 구조를 효과적으로 반영하고자 하였다.

본 연구의 분석에는 R(4.5.2)을 사용하였다. Deep-LSTM 모형은 keras3와 tensorflow R 패키지를 통해 구현하였으며, 잠재성장모형은 lavaan 패키지의 growth() 함수로 적합하였다. 데이터 전처리는 tidyverse 계열 패키지로 수행하였다.

### 3. 수직척도의 활용

본 연구는 Deep-LSTM 모형을 적용하기 위해 수직척도를 활용하였다. 수직척도란 학년이 서로 다른 학생들의 성취 수준을 동일한 척도상에서 직접 비교하기 위해 설계된 공통 점수 체계이다. 이는 시점 간 점수의 직접 비교를 가능하게 하여, 시간이 흐름에 따라 학생의 성취도가 얼마나 상승했는지를 정량적으로 파악할 수 있다.

그리고 본 연구는 이 수직척도를 활용하여 학생의 성장궤적을 추적하고자 하였다. 먼저, 성장 변화량이란 시점  $t$ 와 시점  $t+1$  사이의 수직척도 점수 변화량으로, 특정 기간 동안 학습자가 습득한 지식의 양적 변화를 의미한다. 성취 가속도는 이전 시기 대비 성장 변화의 변화율을 의미하며, 학습자의 성장 동력의 변화를 나타내는 지표이다. 본 연구는 외부의 특이 개입이 없을 때, 학습자가 보일 것으로 기대되는 기대 성장 궤적을 추정하고자 하였다. 성장 변화량과 성취 가속도 차이는 다음과 같이 계산된다.

$$v_t = y_{t+1} - y_t$$

$$a_t = y_{t+1} - 2y_t + y_{t-1}$$

여기서  $v_t$ 는 시점  $t$ 에서의 성장 변화량을 나타내며,  $y_{t+1}$ 은 당해연도( $t$ )의 수직척도 기반 성취 수준이다.  $y_t$ 는 전년도( $t-1$ )의 수직척도 기반 성취 수준이다.  $a_t$ 는 본 연구의 핵심인 성취 가속도를 의미한다.  $a_t$ 가 양의 값을 가진다는 것은 학생의 성적이 관성에서 벗어나 예상보다 도약하고 있음을 의미한다. 그리고 이때, 학습자의 성취 자료는 장기적

인 환경적 요인과 개인적 요인이 복합적으로 작용하는 시계열 데이터이다. 본 연구에서 채택한 2층 구조의 Deep-LSTM은 하위 연결층에서 단기적인 성취 변동을, 상위 연결층에서 장기적인 학생의 학습 패턴과 성취 궤적을 계층적으로 학습함으로써 기대 궤적 산출의 정밀도를 높이고자 했다.

#### IV. 결 과

본 절에서는 세 연구 문제에 대응하여 분석 결과를 제시한다. 먼저 연구 문제 1과 관련하여, 성취 가속도의 비선형성과 장기 의존성을 타당하게 측정하기 위해 요구되는 모형 조건을 검토한다. 이를 위해 잠재성장모형과 Deep-LSTM을 비교하여 어떤 구조적 특성이 가속도 측정에 필수적인지를 경험적으로 확인한다. 다음으로 연구 문제 2는 순열 중요도 분석과 절제 실험을 통해 가속도 예측에 기여하는 핵심 정보를 규명한다. 마지막으로 연구 문제 3은 예측된 가속도 지표가 기존 성취 지표를 통제한 이후에도 이후 성취 변화에 대해 독립적 설명력을 지니는지를 준거타당도 관점에서 검증한다.

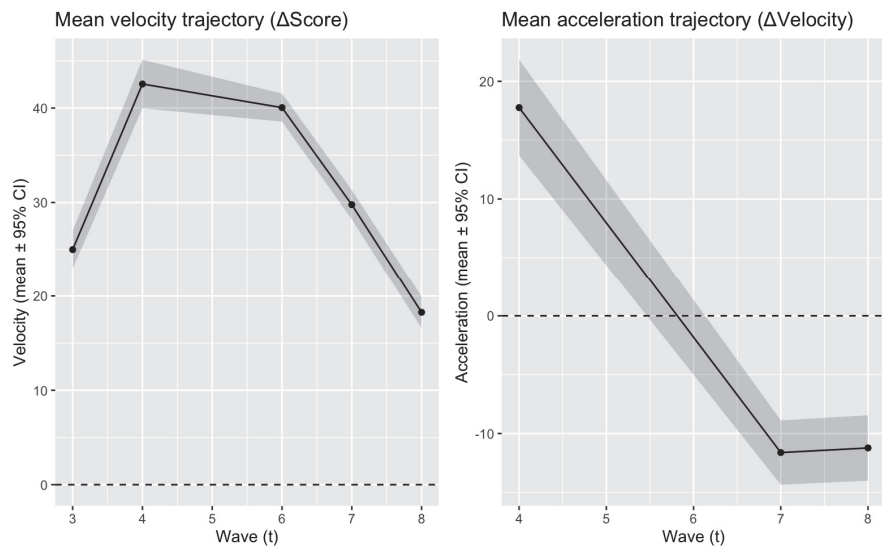
〈표 2〉는 본 연구의 분석에 활용된 최종 유효 표본의 연도별 분포를 보여준다. 본 연구는 대구교육종단연구(DELS)의 2017년 기초 데이터를 바탕으로 하되, 성취 가속도 산출을 위해 최소 3개년( $t, t-1, t-2$ )의 연속된 성취 수준 자료가 확보되어야 하는 분석 모델의 특성을 고려하여 실질적인 분석 구간을 2020년부터 설정하였다. 2021년에 표본 수가 4,004명으로 급증한 것은 2021년 당시 중학교 2학년 학생 약 1,000명을 대상으로 실시된 추가 표집 인원이 3년치 데이터를 확보함에 따라 분석 대상에 포함되었기 때문이다. 본 연구는 매년 약 2,000명에서 3,000명 규모의 표본 크기를 유지하고 있어, 통계적 유의성을 확보하기에 충분한 자료적 토대를 갖추고 있다.

〈표 2〉 연도별 수직척도 점수의 기술통계

| 연도   | 학생수   | 결측률 | 평균 수직척도 점수 | 수직척도점수 표준편차 | 수직척도점수 최솟값 | 수직척도 점수 최댓값 |
|------|-------|-----|------------|-------------|------------|-------------|
| 2018 | 2,769 | 0   | 870        | 54.9        | 676        | 1008        |
| 2019 | 2,685 | 0   | 895        | 63.1        | 703        | 1041        |
| 2020 | 2,275 | 0   | 937        | 72.1        | 710        | 1104        |
| 2021 | 4,004 | 0   | 796        | 43.8        | 649        | 902         |
| 2022 | 3,901 | 0   | 836        | 59.3        | 635        | 971         |
| 2023 | 3,903 | 0   | 866        | 53.5        | 694        | 967         |
| 2024 | 3,912 | 0   | 885        | 63.6        | 697        | 1035        |

〈표 2〉는 대구교육종단연구 자료를 기반으로 연도별 학생 수, 결측률, 그리고 수직척도 점수의 분포 특성을 요약한 것이다. 모든 연도에서 성취 점수의 결측률은 0으로 나타나, 본 연구에서 활용한 성취 자료가 연도별로 완전 관측되었음을 확인할 수 있다. 연도별 학생 수는 2018-2020년에는 약 2,200-2,700명 수준이었으나, 2021년 이후 약 4,000명으로 증가하였는데, 이는 여러 학년 패널이 통합된 종단자료의 특성을 반영한 결과로 해석된다.

연도별 평균 수직척도 점수는 2018년부터 2020년까지 점진적인 상승을 보이다가, 2021년에 상대적으로 낮은 수준으로 변화한 이후 다시 완만한 회복 양상을 나타낸다. 이와 함께 수직척도 점수의 표준편차 역시 연도별로 일정하지 않게 변화하며, 특히 일부 시점에서는 학생 간 성취 격차가 확대되는 양상이 관찰된다. 이러한 결과는 학생 성취의 성장 과정이 단순한 선형 증가로 설명되기보다는, 시점에 따라 성장의 속도와 방향이 달라지는 비선형적 특성을 지님을 시사한다.



[그림 2] 시간에 따른 평균 속도 변화량과 가속도 변화량

〈표 2〉와 [그림 2]에 제시된 이러한 비선형적 성장 패턴과 시점별 분포의 변화는, 장기 종단자료에서 성장 궤적의 동태적 변화를 포착할 필요성을 강조한다. 이는 성취 수준이나 변화량에 기반한 접근을 넘어, 시간에 따른 성장 변화의 구조를 함께 고려할 수 있는 분석 방법의 적용이 요구됨을 보여주며, 본 연구에서 Deep-LSTM 모델을 활용하여 성취 가속도를 분석하고자 한 배경을 뒷받침한다.

[그림 2]의 왼쪽 그림은 연도별 평균 성취 변화율(속도)의 궤적을 제시한 것으로, 전

반적으로 성취는 모든 시점에서 증가하는 양상을 보인다. 그러나 변화율의 크기는 시점에 따라 차이를 보이며, 이후 시점으로 갈수록 성장 속도가 둔화되는 경향이 관찰된다. 이에 비해 [그림 2]의 오른쪽 그림은 성취 변화율의 변화, 즉 가속도의 평균 궤적을 나타낸다. 초기 시점에서는 가속도가 양의 값을 보여 성장 속도가 촉진되는 국면임을 시사하나, 이후 가속도는 0을 기준으로 감소하여 후반 시점에서는 음의 값으로 전환된다. 이는 성취 수준이 여전히 증가하고 있음에도 불구하고, 성장 과정은 이미 감속 국면에 진입하였음을 의미한다. 이러한 결과는 성장의 동태적 국면을 이해하기 위해 변화율뿐 아니라 가속도 수준의 분석이 필요함을 보여준다.

<표 3> 모형별 예측 성능 비교

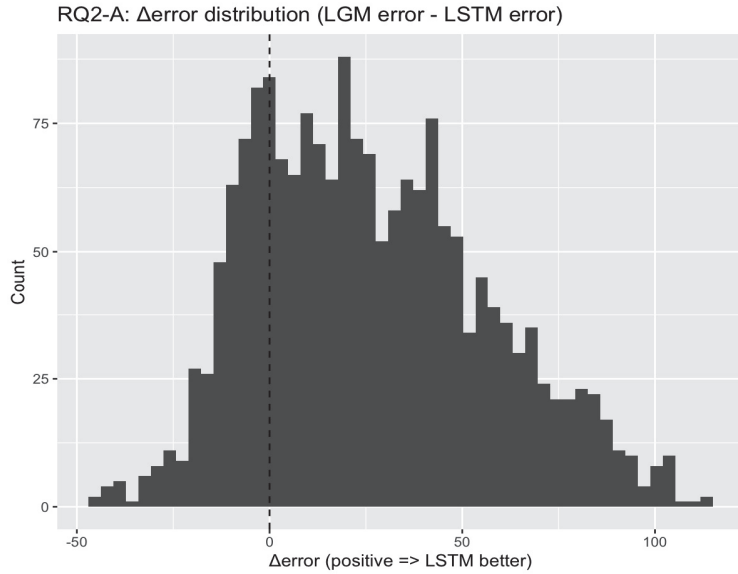
| 모형             | 예측 과제           | RMSE<br>평균 | RMSE<br>표준편차 | MAE<br>평균 | MAE<br>표준편차 | 반복<br>횟수 | 평가 표본 수 |
|----------------|-----------------|------------|--------------|-----------|-------------|----------|---------|
| 잠재성장모형 (선형)    | 다음 시점 가속도       | 87.8       | -            | 69.3      | -           | -        | 1,836   |
|                | 다음 시점 가속도       | 54.8       | 25.4         | 42.4      | 20.5        | 5        | 1,836   |
| 다중과제 Deep-LSTM | 다음 시점 성취 수준     | 46.7       | 17.8         | 36.4      | 14.4        | 5        | 3,176   |
|                | 다음 시점 성적<br>변화량 | 41.2       | 15.8         | 32.4      | 13.1        | 5        | 3,176   |

<표 3>은 모형별 예측 변수에 따른 성능을 비교한 결과를 제시한다. 먼저, 본 연구의 분석 표본은 예측 과제별로 상이하다. 성취 수준 예측과 성적 변화량 예측은 각각 단일 시점(t+1) 및 인접 두 시점(t, t+1)의 관측값만을 요구하므로, 해당 조건을 충족하는 학생 3,176명이 분석 대상에 포함되었다. 반면 성취 가속도 예측은 정의상 연속 3개년의 성취도 자료가 모두 확보되어야 하므로, 이 조건을 충족하는 학생 1,836명만이 가속도 예측 평가 표본으로 활용되었다. 이러한 표본 구성의 차이는 가속도라는 고차 변화량 지표의 산출 요건에서 비롯되는 것으로, 자료 가용성에 따른 자연스러운 결과이다.

다음으로, 다음 시점 가속도 예측에서 전통적인 잠재성장모형은 RMSE 87.8, MAE 69.3으로 나타난 반면, 다중과제 Deep-LSTM 모형은 RMSE 54.8, MAE 42.4로 상대적으로 우수한 예측 성능을 보였다. Deep-LSTM 모형은 다음 시점 성취도와 예측에서도 안정적인 성능을 보여줌으로써 다중과제 학습 구조의 유용성을 확인할 수 있었다.

RMSE와 MAE 간 차이의 크기는 오차 분포의 형태에 대한 정보를 제공한다. 일반적으로 RMSE가 MAE보다 같거나 크며, 두 값의 차이가 클수록 일부 극단적 예측 오차가 존재함을 의미한다. 본 연구에서는 Deep-LSTM의 RMSE(54.8)와 MAE(42.4) 차이가

12.4로, 잠재성장모형의 RMSE-MAE 차이(18.5)보다 작아 예측 오차의 분포가 상대적으로 균일함을 시사한다. 또한, 개인 단위 오차 비교를 위해 5개 시드에 걸친 예측값의 평균을 사용하여, 시드별 확률적 변동성을 완충하고 Deep-LSTM의 안정적 예측 특성을 반영하였다.



[그림 3] 모형의 오차 분포 히스토그램

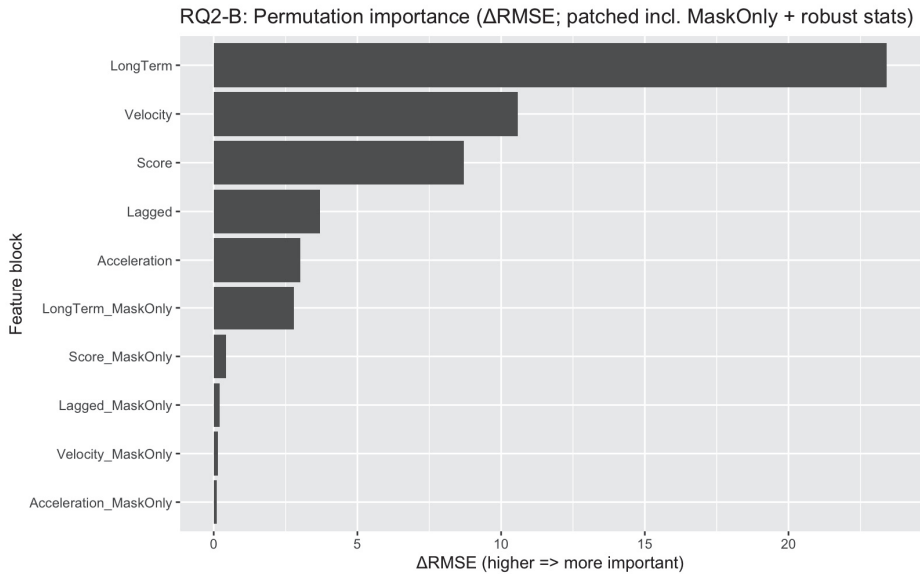
[그림 3]은 모형의 오차 분포 히스토그램을 보여준다. 전반적으로 Deep-LSTM이 0보다 오른쪽에 치우쳐 있는데, 이는 Deep-LSTM이 잠재성장모형에 비해 좋은 성능을 나타낸다는 점을 시사한다. 하지만 [그림 3]에서 보듯이 분산이 크고 0보다 작은 영역도 역시 존재함을 보여줌으로써 항상 Deep-LSTM이 잠재성장모형에 비해 우위에 있는 것은 아니고, 학생에 따른 차이가 있을 수도 있음을 보여준다.

<표 4> 기초 성취 사분위별 결과

| 사분위 | 학생 수 | 오차 변화 평균 | 오차 변화 표준편차 | Deep-LSTM이 더 우수할 확률 |
|-----|------|----------|------------|---------------------|
| 1   | 433  | 24.8     | 27.3       | 0.794               |
| 2   | 501  | 24.9     | 29.7       | 0.772               |
| 3   | 499  | 26.6     | 31.6       | 0.772               |
| 4   | 403  | 29.0     | 31.3       | 0.782               |

〈표 4〉는 모든 4분위에서 Deep-LSTM의 오차 변화의 평균값은 모두 양의 값을 나타냄을 보여준다. 특히 최하위 사분위부터 최상위 사분위까지 오차 변화의 평균은 24.8에서 29.0 수준으로 유사하게 분포하여 특정 성취 집단에 국한되지 않은 성능 향상을 시사한다. 또한 각 사분위에서 오차 변화가 양의 값을 가질 확률이 0.77~0.79로 나타나, Deep-LSTM이 대부분 더 정확한 예측을 제공함을 확인할 수 있었다.

다음으로는 Deep-LSTM의 성능 차이가 어떠한 요인에 의해 설명되는지를 알기 위해 순열 중요도(Permutation Importance) 분석을 실시하였다. 입력변수를 성취 수준, 성적 변화량, 성취 가속도, 시차변수, 장기 맥락 정보 등으로 나누어 각 변수를 개인 축으로 무작위 재배열 한 다음, 다음 시점의 성취 가속도를 예측할 때의 RMSE 증가량을 산출하였다.



[그림 4] 가속도 예측에 있어서의 변수 중요도 분석 결과

[그림 4]는 이 분석에 따른 결과를 제시한다. 분석 결과, 장기 맥락 정보(Long Term)을 교란했을 때 RMSE 증가량이 가장 크게 나타났으며, 그 다음으로 변화율, 점수 수준 순서로 중요도가 확인되었다. 이것은 장기 맥락 정보가 무의미 해졌을 때 모형의 성능이 가장 떨어진다는 의미로, 장기 맥락 정보가 이 모형에서 가장 중요함을 보여준다. 반면, 시차 변수 및 현재 시점의 가속도 정보는 상대적으로 제한적인 기여도를 보였다.

<표 5> 가속도 예측에 있어서의 변수 제거시 RMSE 변화량 비교

| 요인              | $\Delta$ RMSE 평균 | $\Delta$ RMSE 표준편차 | $\Delta$ RMSE 중앙값 | IQR  | 시드 수 |
|-----------------|------------------|--------------------|-------------------|------|------|
| 장기 누적 특성        | 23.4             | 24.6               | 15.8              | 9.00 | 5    |
| 성장 변화량          | 10.6             | 7.65               | 8.34              | 1.54 | 5    |
| 성취 수준           | 8.71             | 7.79               | 5.12              | 4.49 | 5    |
| 시차 변수           | 3.70             | 2.84               | 3.49              | 2.03 | 5    |
| 성취 가속도          | 3.01             | 1.85               | 2.26              | 1.94 | 5    |
| 장기 누적 특성-결측 패턴만 | 2.79             | 4.01               | 1.22              | 3.62 | 5    |
| 성취 수준-결측 패턴만    | 0.42             | 0.21               | 0.34              | 0.24 | 5    |
| 시차 변수-결측 패턴만    | 0.20             | 0.55               | 0.10              | 0.47 | 5    |
| 성장 속도-결측 패턴만    | 0.15             | 0.12               | 0.15              | 0.18 | 5    |
| 가속도-결측 패턴만      | 0.10             | 0.08               | 0.05              | 0.10 | 5    |

주: 1. RMSE = Root Mean Squared Error(평균 제곱근 오차); IQR = Interquartile Range(사분위범위, Q3 - Q1);  $\Delta$ RMSE = 순열 후 RMSE - 기준 RMSE.; 모든 오차 지표는 값이 작을수록 예측 성능이 우수함을 의미한다.  
 2. '결측 패턴만' 계열은 해당 변수의 결측 마스크 채널만 순열하여, 결측 정보 자체의 기여도를 단독으로 측정할 수 있다. 나머지 변수(장기 누적 특성, 성장 변화량 등)는 값 채널과 결측 마스크 채널을 함께 순열하여 전체 기여도를 측정할 수 있다.

<표 5>에서는 추가적으로 실시한 각 변수의 결측 패턴만을 단독으로 무작위 배열하고 중요도를 확인하는 분석을 실시하였다. 그 결과, 장기 맥락 정보에 해당하는 결측정보에서는 의미 있는 RMSE 증가가 관찰된 반면, 점수, 변화율, 가속도에서의 결측치는 예측 성능에 거의 영향을 미치지 않았다. 이는 결측 정보 자체가 성취 가속도 예측에 중요한 것이라기보다는, 장기적인 시간 맥락 속에서 구조화된 결측 정보일 때에만 결측 정보가 의미 있는 정보로 활용될 수 있음을 시사한다. 또한 특정 초기값에 따른 변동성 영향을 점검하기 위해 중앙값 기반 요약물 실시한 결과, 장기 맥락 정보와 변화율의 중요도 순위는 일관되게 유지됨으로써 순열 중요도의 강건성을 확인할 수 있었다.

각 분석은 다섯 개의 무작위 초기값(시드)으로 반복 학습하여 성능의 평균과 변동성을 함께 보고하였다. 시드 수는 초기값 설정에 따른 학습 결과의 확률적 변동을 추정하기 위해 동일 모형을 반복 학습한 횟수를 의미한다. 변동성 지표로는 표준편차와 함께 IQR(Interquartile Range, 사분위범위; 제3사분위수와 제1사분위수의 차이)을 보고하였는데, IQR은 극단값의 영향을 배제하고 중앙 50% 범위의 분포 폭을 나타내는 강건한 지표이다.

다음으로는 절제(Ablation) 실험을 통해 순열 중요도 분석에서 확인된 구성요소들이 실제로 예측 성능에 필수적인지를 검증하였다. 이는 입력 정보의 일부를 체계적으로 제거함으로써 Deep-LSTM의 성능에 어떤 요인이 영향을 끼쳤는지를 체계적으로 보여준다.

<표 6> 절제 실험에 따른 Deep-LSTM 변형 모형의 예측 성능 비교

| 모형 변형       | RMSE 평균 | RMSE 표준편차 | MAE 평균 | MAE 표준편차 | 반복 횟수 | 평가 표본 수 |
|-------------|---------|-----------|--------|----------|-------|---------|
| 전체 모형       | 56.60   | 25.20     | 43.00  | 20.50    | 5     | 1,836   |
| 결측 패턴 미포함   | 64.90   | 3.00      | 48.80  | 3.32     | 5     | 1,836   |
| 장기 맥락 제거    | 65.40   | 1.12      | 48.70  | 1.16     | 5     | 1,836   |
| 단일과제 가속도 모형 | 52.80   | 23.80     | 39.70  | 19.20    | 5     | 1,836   |

주: 모든 변수는 표준화(z-점수)되어 투입되었다. 동일 학생 내 반복 측정 구조를 고려하여 학생 ID를 군집 단위로 하는 군집 강건 표준오차(cluster-robust standard error)를 적용하였다. 분석은 R의 lm() 함수와 lmtest::coefest()를 통해 수행되었다.

<표 6>에서 나타난 결과는 입력 정보의 구성에 따른 뚜렷한 성능차이를 보여준다. 모든 구성요소를 포함한 전체 모형은 평균 RMSE 56.60, MAE 43.00으로 안정적인 예측 성능을 보였다. 반면, 입력 변수의 결측 여부를 반영하지 않은 결측 패턴 미포함 모형과 장기 누적 성장 맥락 정보를 제거한 장기 맥락 제거 모형에서는 RMSE와 MAE가 모두 증가하여, 뚜렷한 모형의 성능 저하가 관찰되었다. 이는 Deep-LSTM의 성능 우위가 시계열 구조 자체보다는 시간적으로 구조화된 결측 패턴과 장기 성장 맥락 정보를 함께 활용하는 데에서 기인함을 시사한다.

한편 성취 가속도만을 단일 과제로 학습한 모형은 평균 RMSE 52.80, MAE 39.70로 전체 모형과 유사하거나 일부 지표에서는 더 낮은 평균 오차를 보였으나, 표준편차가 크게 나타나 변동성이 상대적으로 큰 것으로 확인되었다. 비록 표준편차가 전체 모형보다는 낮지만 입력 정보의 상당수를 제거함으로써 많은 정보적 손실이 있을 수 있음을 의미한다. 단일 과제 모형은 단일 과제 학습이 특정 조건에서는 높은 적합도를 보일 수 있으나, 일반화 측면에서는 다중과제 학습을 적용한 전체 Deep-LSTM 모형이 보다 풍부하고 상대적으로 일관적인 성능을 제공할 수 있음을 시사한다.

마지막으로 본 연구는, Deep-LSTM이 예측한 가속도 기반 성장참조지표가 교육평가 관점에서 타당한 정보를 제공하는지를 검증하고자 한다. 이를 위해 본 연구는 예측된 가속도 지표가 이후 학업 성취 변화와 어떠한 관계를 가지는지를 분석하였다. 다시 말해, 본 연구는 예측된 성취 가속도의 준거타당도를 검증하기 위해 다음 시점 학업 변화율을 종속변수로 하는 회귀분석을 수행하였다. 동일 학생의 여러 시점 관측치가 반복 측정되는 자료 구조를 고려하여, 학생 ID를 군집 단위로 하는 군집 강건 표준오차를 적용하였다. 이는 동일 학생 내 관측치 간 오차 상관을 허용함으로써 표준오차의 과소 추정을 방지하기 위한 절차이다. 모형 1은 현재 성취 수준과 성장 속도를 통제변수로 포함하였으며, 모형 2는 여기에 Deep-LSTM으로 예측한 성취 가속도를 추가하였다. 모든 예측변수는 표준화(z-점수)하여 투입하였다.

<표 7>을 보면, 다음 시점 학업 변화율을 종속변수로 한 분석에서 현재 성취 수준과 변화율을 통제한 이후에도 Deep-LSTM이 예측한 가속도 지표는 통계적으로 유의미한 설명력을 보였다. 구체적으로, 예측 가속도는 이후 학업 변화율을 유의하게 예측하였으며, 이는 가속도 지표가 기존의 성취 수준이나 변화율로 설명되지 않는 추가적인 정보를 포함하고 있음을 의미한다. 즉, 가속도는 단순한 성취 상태나 연간 성장량이 아니라, 학업 성장의 방향성과 속도가 어떻게 변화하고 있는지를 반영하는 지표로서 기능함을 보여주고 있다.

<표 7> 예측된 성취 가속도의 준거타당도 분석 결과

| 모형   | 변수               | 추정치    | 표준오차 | t값     | p값     |
|------|------------------|--------|------|--------|--------|
| 모형 1 | 절편               | 27.70  | 1.12 | 24.80  | < .001 |
|      | 성취 수준 (표준화)      | -7.45  | 1.23 | -6.04  | < .001 |
|      | 성장 속도 (표준화)      | -20.40 | 1.30 | -15.70 | < .001 |
| 모형 2 | 절편               | 27.70  | 0.76 | 36.30  | < .001 |
|      | 성취 수준 (표준화)      | -3.34  | 0.87 | -3.85  | < .001 |
|      | 성장 속도 (표준화)      | 21.80  | 1.52 | 14.30  | < .001 |
|      | 예측된 성취 가속도 (표준화) | 56.50  | 1.31 | 43.20  | < .001 |

주목할 점은 모형 1에서 음의 값(-20.40)을 보이던 성장 속도의 회귀계수가 예측 가속도를 투입한 모형 2에서는 양의 값(+21.80)으로 부호가 역전된다는 것이다. 이는 전형적인 억제효과로, 예측 가속도가 회귀식에 포함됨으로써 기존 변수(성장 속도)에 혼입되어 있던 분산 구조가 분리되어 각 변수의 고유한 설명력이 드러난 결과로 해석할 수 있다(Tzelgov & Henik, 1991). 변수 간 분산팽창지수(VIF)를 확인한 결과, 성취 수준 1.27, 성장 속도 2.70, 예측된 성취 가속도 2.59로 모두 10 미만이었다. 이는 부호 역전이 다중공선성에 의한 불안정한 추정이 아니라, 예측 가속도가 성장 속도 변수 내부의 이질적 정보를 분리해내는 보완적 기능을 수행함을 시사한다.

## V. 논의 및 제언

본 연구는 장기 교육종단자료를 활용하여 성취 가속도를 성장 참조 평가 지표로 개념화하고, 그 타당성을 검증하고자 하였다. 이를 위해 Deep-LSTM 기반 시계열 모형을 적용하여 전통적 성장모형과의 예측 성능을 비교하고, 그 성능 차이가 발생하는 구조적

원인을 분석하였다. 나아가 성취 가속도가 성장 참조 평가 지표로서 지니는 의미와 활용 가능성을 논의함으로써, 방법론 비교를 넘어 새로운 평가 지표 개념의 확장에 기여하고자 하였다.

먼저 첫 번째 연구 질문과 관련하여, 분석 결과 Deep-LSTM 모형은 전통적 성장모형에 비해 다양한 성능지표에서 가속도 측정에 대한 일관된 우위를 보였다. 예를 들어, 평가 표본에서 관측된 성취 가속도의 표준편차는 약 87.71점이었다. 이를 기준으로 환산하면, 잠재성장모형의 RMSE(87.8)는 약 1.00 SD 수준으로, 관측된 가속도의 변동성과 유사한 수준의 오차를 보였다. 반면 Deep-LSTM의 RMSE(54.8)는 약 0.62 SD 수준으로, 잠재성장모형 대비 약 0.38 SD의 예측 오차 감소에 해당한다. 특히 잠재성장모형은 정의상 각 개인의 가속도를 0으로 가정하므로, 잠재성장모형의 가속도 예측 오차가 관측된 가속도의 표준편차와 유사하게 나타난 것은 모형의 이론적 제약이 실제 자료에서 드러난 결과로 해석할 수 있다. Deep-LSTM은 이러한 근본적 제약을 넘어 가속도 변동의 일부를 포착할 수 있음을 보였다.

이러한 성능 우위는 특정 시점이나 특정 성취 집단에 국한되지 않고, 학생들의 학업 성취 수준 전반에 걸쳐 전체적으로 나타났다는 점에서 의의가 있다. 성취 가속도는 변화율의 변화를 의미하는 고차 변화량으로, 선형 혹은 저차 다항식 가정을 전제로 하는 전통적 성장모형과는 구조적으로 다른 예측 대상이다. 전통적 성장모형은 설명력과 해석 가능성, 그리고 이론적 적합성 측면에서 분명한 강점을 지니지만, 미래 예측이나 국소적 변화의 포착에는 한계가 있을 수 있다. 물론, 본 연구에서 확인된 Deep-LSTM의 적합성은 이 모형이 복잡한 AI 기반 모형이기 때문이라기보다는, 가속도라는 예측 대상의 특성에 보다 적합한 특성을 학습했기 때문으로 해석할 수 있다. 이러한 결과는 Deep-LSTM이라는 모형이 향후 학교 현장에서 형성평가 보조 자료로 활용될 가능성을 시사한다.

두 번째 연구 질문과 관련한 분석은 Deep-LSTM이 가속도 예측에서 우수한 성능을 보이는 이유를 체계적으로 설명한다. 개인 단위 오차 분석 결과, Deep-LSTM의 성능 우위는 평균적 지표에만 국한되지 않고 다수의 학생과 시점에 걸쳐 광범위하게 나타났다. 이는 일부 극단적 사례에 의해 평균 성능이 왜곡된 결과가 아님을 시사한다. 더불어 순열 중요도 분석과 절제 실험 결과는 Deep-LSTM의 성능 우위가 어떠한 정보적 기반 위에서 형성되는지를 명확히 보여준다. 전통적 성장모형이 성취 변화를 저차 다항식 형태의 잠재 성장 요인(절편, 기울기 등)으로 요약하는 데 비해, Deep-LSTM은 장기 누적 정보, 개인 수준 통계, 변화율 신호를 비선형적으로 결합하여 활용한다. 이러한 구조는 가속도와 같은 고차 변화량을 예측하는 데 특히 유리한 것으로 해석된다.

특히 교육평가적 맥락에서는 평균적으로 약간 더 나은 성능보다, ‘누구에게, 언제’ 모

형이 더 효과적으로 작동하는지가 중요하다. 이러한 점에서 개인 단위 오차 분포 분석은 중요한 시사점을 제공한다. Deep-LSTM의 성능 우위는 특정 소수의 학생이 아니라 다수의 학생에게서 반복적으로 관찰되었으며, 이는 교육 현장에서의 가속도 측정을 위한 Deep-LSTM의 실질적 활용 가능성을 뒷받침한다.

결측치 설계에 대한 분석 역시 중요한 함의를 제공한다. 본 연구 결과는 결측 여부 자체가 항상 강력한 예측 변수가 되기보다는, 장기적 시간 맥락 속에서 반복·누적되는 결측 패턴이 성취 가속도 예측에 의미 있는 신호로 작용함을 보여준다. 이는 대규모 교육 종단자료에서 결측이 단순한 통계적 문제를 넘어, 학습 경험의 구조적 특성을 반영할 수 있음을 시사한다.

또한 본 연구에서 장기 누적 정보는 예측 성능에 있어 압도적인 중요성을 보였다. 이는 단일 시점의 성취 점수보다 누적된 학습 이력이 더 많은 정보를 담고 있음을 의미하며, 성장 참조 평가의 이론적 전제와도 부합한다. 변화율과 가속도는 위계적 관계를 이루며, 가속도 예측은 변화율 정보를 전제로 한다. 이러한 관점에서 단일 점수 중심의 평가는 구조적 한계를 지니며, 가속도는 동일한 현재 성취 수준이라 하더라도 개인의 장기적 분포 맥락에 따라 해당 변화가 도약인지 일시적 변동인지를 구분할 수 있게 한다는 점에서 중요한 가치를 지닌다. 비록 기계학습 모형은 종종 이해하기 어려운 블랙박스로는 비판을 받지만, 본 연구는 중요도 분석을 통해 정보 기여 구조에 대한 설명 가능성을 제시하였고, 전통적 성장모형과의 비교를 통해 해석을 확장하였다.

절제 실험 결과 또한 이러한 해석을 뒷받침한다. 특히 장기 맥락 정보를 제거했을 때 성능이 크게 저하된다는 점은 학생의 성취 가속도가 단기 변화만으로는 설명될 수 없음을 의미한다. 더불어 결측 정보를 제거했을 때의 급격한 성능 저하는, 결측이 단순한 잡음이 아니라 의미 있는 맥락 정보임을 확인시켜 준다. 즉, Deep-LSTM의 가속도 예측 성능 우위는 장기 의존성, 비선형 변화 구조, 결측 정보를 포함한 시간 맥락을 활용할 수 있는 모형 구조와 관련되어 있는 것으로 해석 가능하다.

아울러 본 연구는 단일 과제 Deep-LSTM에 비해 다중 과제(multi-task) 구조가 갖는 이점을 확인하였다. 이는 성취 수준-변화율-가속도로 이어지는 성취 변화의 위계적 구조를 하나의 모형 내에서 통합적으로 학습하고 해석할 수 있기 때문이다. 따라서 본 연구에서 관찰된 성능 향상은 특정 기술 요소 하나에 기인하기보다는, 입력 설계, 결측치 처리 방식, 다중 과제 학습 구조 등이 데이터의 특성과 조화를 이루어 나타난 결과로 볼 수 있다.

세 번째 연구 질문과 관련하여, 본 연구는 가속도 기반 성장 참조 평가지표의 타당도를 검증하였다. 특히 준거 타당도의 관점에서, 가속도 지표가 기존에 타당성이 입증된 성취 지표 및 이후 학업 성취 변화와 유의미한 관련성을 지니는지를 분석하였다. 분석

결과, Deep-LSTM이 예측한 가속도 지표는 이후 성취 변화를 예측하는 데 있어 준거 타당도를 확보하고 있을 뿐만 아니라, 기존의 성취 수준이나 변화율 지표를 보완하는 추가적인 평가 정보를 제공함을 확인하였다. 이는 기존 성장 지표를 통제된 이후에도 가속도가 독립적인 설명력을 지닌다는 점에서 중요한 의미를 갖는다.

더 나아가 본 연구는 가속도가 무엇을 측정하는 지표인지를 개념적으로 정리하고자 하였다. 성취 수준이 학습자의 현재 위치를, 변화율이 현재의 속도를 의미한다면, 가속도는 성장의 방향성을 반영한다. 평가 지표로서 가속도의 강점은 단기적 성과보다는 장기적 추세를 포착할 수 있다는 점에 있다. 이는 교육적 개입의 적절한 시점을 판단하는데 유용한 정보를 제공한다. 즉, 학생의 성장 방향이 정체되거나 하락 국면에 진입할 가능성을 사전에 포착함으로써, 적절한 시점에 교육적 개입을 설계할 수 있는 근거를 제공한다.

이러한 맥락에서 가속도는 성장 참조 평가에서 실질적으로 활용될 수 있는 지표이다. 예를 들어, 성적표에 성취 점수와 함께 가속도의 방향성을 제시함으로써 학생, 학부모, 교사가 학생의 학습 상황을 입체적으로 이해할 수 있다. 또한 가속도 지표는 조기 학습 경고 신호로 기능하여, 학습자의 성장 둔화 가능성을 사전에 탐지하는 데 기여할 수 있다. 이는 2022 개정 교육과정이 강조하는 맞춤형 교육의 방향성에도 부합한다. 동일한 성취 점수를 받은 학생이라 하더라도 성장 동력은 상이할 수 있기 때문에, 형성 평가 관점에서 학생 개인의 성취 변화를 지속적으로 모니터링하는 데 가속도 지표가 활용될 수 있다. 나아가 가속도는 AI 기반 평가지표의 책임 있는 활용이라는 측면에서도 의미를 지닌다. 가속도가 학생을 단정적으로 분류하는 도구가 아니라, 교사의 전문적 판단을 보조하는 형성적 평가 도구로 활용될 때, AI 기반 평가지표의 윤리적 사용 가능성도 함께 제고될 수 있다.

교육 정책적 차원에서도 성취 가속도 지표는 다양하게 활용될 수 있다. 첫째, 학교 단위의 평가에서 성취 가속도는 기존의 성취 수준 중심 평가가 지닌 한계를 보완할 수 있다. 성취 수준만을 기준으로 학교를 평가할 경우, 학생의 사회경제적 배경이 유리한 학교가 구조적으로 높은 평가를 받는 편향이 발생할 수 있다. 반면, 성취 가속도 지표는 학생 개인의 성장 궤적을 준거로 삼기 때문에 초기 성취 수준이 낮은 학생이 다수인 학교라 하더라도 학생들의 성장 동력이 어떻게 변화하고 있는지를 포착할 수 있다. 이는 교육 자원 배분의 형평성을 제고하는 데 기여할 수 있다.

둘째, 기초학력 보장 정책 및 학습 부진 조기 지원 체계와 연계할 수 있다. 현재의 기초학력 진단은 특정 시점의 성취 수준에 초점을 두어, 이미 학습 결손이 누적된 이후에야 개입이 이루어진 사후 대처적 한계를 지닌다. 가속도 지표는 성취 수준이 아직 임계 기준을 벗어나지 않은 학생 중에서도 성장 동력이 둔화 국면에 진입한 학생을 조기에

식별할 수 있어, 예방적 개입의 근거 자료로 충분히 활용될 수 있다. 이는 기초학력 보장법의 취지와도 부합하며, 학습지원대상학생 선정의 판단 자료로 병행 활용될 여지가 있다.

셋째, 교육 격차 모니터링 및 정책 효과 평가에도 기여할 수 있다. 가속도 지표는 지역과 학교 간 성장 동력의 차이를 추적할 수 있게 함으로써, 교육 격차가 확대, 유지, 축소되는 국면을 정밀하고 세심하게 진단할 수 있다. 특히 교육 개입 정책(예: AI 디지털 교과서 도입 등)의 효과를 검증할 때, 단순한 사전-사후 성취 비교를 넘어 정책 도입 전후의 성장 가속도 변화를 함께 살펴봄으로써 정책의 장기적인 효과를 포착할 수 있다.

한편 본 연구는 몇 가지 한계를 지닌다. 첫째, 자료적 한계로서 대구 지역의 교육종단 자료에 한정되어 있어 연구 결과의 일반화에 제약이 있을 수 있다. 둘째, 모형적 측면에서 Deep-LSTM은 관측된 패턴에 기반한 예측 모형으로, 잠재적 인과 구조를 직접적으로 설명하는 데에는 한계가 있다. 셋째, 가속도 지표는 인과적 해석보다는 예측에 초점을 둔 지표이므로, 정책 결정의 직접적 근거로 사용되기보다는 보조적 판단 자료로 활용되어야 한다. 또한 지표의 오남용은 학생에 대한 낙인 효과나 과도한 예측 의존을 초래할 수 있다는 윤리적·실천적 한계도 존재한다. 마지막으로, 2021년에는 추가 표집된 코호트가 분석 대상에 포함되어 표본 수가 4,004명으로 급증하였다. <표 2>에서 확인되듯, 해당 시점의 평균 성취 수준과 표준편차는 다른 시점과 다소 상이한 분포를 보였다. 이는 추가 표집 코호트의 학년 구성 및 진입 시점이 기존 코호트와 다르기 때문으로 추정되며, 본 연구의 분석에서는 시간항(time,  $time^2$ )과 결측 마스크를 입력에 포함시킴으로써 시점별 분포 차이를 모형이 학습할 수 있도록 설계하였다. 다만 코호트 간 이질성에 대한 명시적 통제는 수행하지 않았으므로, 결과 해석 시 이 점을 고려할 필요가 있다. 후속 연구에서는 코호트 지표를 통제변수로 포함한 민감도 분석을 통해 이를 보완할 수 있을 것이다.

이에 따라 후속 연구에서는 방법론적 확장이 필요하다. 모의실험 연구를 통해 가속도 지표를 보다 정밀하게 검증할 수 있으며, 다른 교과, 지역, 국가의 자료를 적용함으로써 외적 타당도를 확장할 수 있을 것이다. 또한 교육 개입 연구와 결합하여, 가속도 변화에 따른 실제 교육적 개입 효과를 검증하는 연구도 유의미한 후속 과제가 될 것이다.

결론적으로 본 연구는 성장 참조 평가의 실현을 위한 하나의 도구로서 가속도 개념을 제안하고, 그 평가적 타당성을 검증하였다. 맞춤형 기반 교육을 구현하기 위해서는 학생 성취를 보다 입체적으로 포착할 수 있는 지표의 확립이 필수적이다. 본 연구는 성취 가속도라는 새로운 성장 참조 평가 지표가 그러한 시도의 출발점이 될 수 있을 것이라고 본다.

## 참고문헌

- 교육부(2022). **초·중등학교 교육과정 총론**. 교육부 고시 제2022-33호 [별책 1].
- 김소영, 윤기봉(2016). 잠재성장모형을 이용한 청소년의 삶의 만족도 변화추정: 변화에 있어서의 개인차에 대한 요인 검증. **한국청소년연구**, 27(2), 271-296.  
<https://doi.org/10.14816/sky.2016.27.2.271>
- 김수정, 이윤수, 송미령, 송지훈(2020). 학교에 대한 학생인식의 종단적 변화 연구: 잠재성장 모형의 접근. **한국산학기술학회 논문지**, 21(6), 275-285.
- 노언경, 이현정, 이은수, 홍세희(2017). 중학생의 학교폭력 피해경험 및 가해경험의 변화 양상: 다변량 다층 잠재성장모형을 적용한 개인 및 학교효과 분석. **한국청소년연구**, 28(1), 37-65.
- 대구광역시교육청 대구미래교육연구원(2024). **대구교육종단연구(Daegu Educational Longitudinal Study, DELS) 데이터 [Data set]**. 대구: 대구광역시교육청 대구미래교육연구원.
- 박희진, 남궁지영(2019). 학업성취 변동 종단 분석: 교육격차 완화 요인 탐색을 중심으로. **교육사회학연구**, 29(4), 65-88.
- 백수진, 박현정, 신중휘(2023). 청소년기 학업 스트레스 종단적 변화유형 및 예측변인 분석: 성장혼합모형과 SEM Tree 기법의 비교를 중심으로. **아시아교육연구**, 24(4), 641-666.
- 서미정(2009). 초기 청소년의 외현적 공격성 변화와 비행, 우울/불안 및 학업성취감: 잠재성장분석. **한국청소년연구**, 20(2), 141-167.
- 심재권(2025). 학업부진 예측모델 개발 연구: 대구교육종단연구의 국어와 수학을 중심으로. **정보교육학회논문지**, 29(1), 53-61.
- Alzubi, J., Nayyar, A., & Kumar, A. (2018). Machine learning from theory to algorithms: An overview. *Journal of Physics: Conference Series*, 1142, 012012.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Baker, R. S. J. D., & Inventado, P. S. (2014). Educational data mining and learning analytics. In J. A. Larusson, & B. White (Eds.), *Learning Analytics: From Research to Practice* (pp. 61-75). New York: Springer.

- Balfanz, R., Herzog, L., & Mac Iver, D. J. (2007). Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools: Early identification and effective interventions. *Educational Psychologist, 42*(4), 223–235.
- Betebenner, D. W., & Linn, R. L. (2010). *Growth in student achievement: Issues of measurement, longitudinal data analysis, and accountability*. Princeton, NJ: K–12 Center at Educational Testing Service.
- Bloom, B. S. (1976). *Human Characteristics and School Learning*. New York: McGraw–Hill.
- Bollen, K. A., & Curran, P. J. (2006). *Latent Curve Models: A Structural Equation Perspective*. Hoboken, NJ: Wiley.
- Caspi, A., Wright, B. R. E., Moffitt, T. E., & Silva, P. A. (1998). Early failure in the labor market: Childhood and adolescent predictors of unemployment in the transition to adulthood. *American Sociological Review, 63*(3), 424–451.
- DiPrete, T. A., & Eirich, G. M. (2006). Cumulative advantage as a mechanism for inequality: A review of theoretical and empirical developments. *Annual Review of Sociology, 32*(1), 271–297.
- Downey, D. B. (2020). *How Schools Really Matter: Why Our Assumption about Schools and Inequality Is Mostly Wrong*. Chicago: University of Chicago Press.
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks–Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428–1446.
- Henry, K. L., Knight, K. E., & Thornberry, T. P. (2012). School disengagement as a predictor of dropout, delinquency, and problem substance use during adolescence and early adulthood. *Journal of Youth and Adolescence, 41*(2), 156–166.
- Lohman, D. F., & Korb, K. A. (2006). Gifted today but not tomorrow? Longitudinal changes in ability and achievement during elementary school. *Journal for the Education of the Gifted, 29*(4), 451–484.
- McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology, 60*(1), 577–605.

- Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, 55(1), 107–122.
- Muthén, B. O. (2001). Latent variable mixture modeling. In G. A. Marcoulides, & R. E. Schumacker (Eds.), *New Developments and Techniques in Structural Equation Modeling* (pp. 1–33). Mahwah, NJ: Lawrence Erlbaum Associates.
- Muthén, B., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research*, 24(6), 882–891.
- Reimers, N., & Gurevych, I. (2017). Optimal hyperparameters for deep lstm-networks for sequence labeling tasks. arXiv preprint arXiv:1707.06799.
- Reschly, D. J., & Bergstrom, M. K. (2009). Response to intervention. In T. B. Gutkin, & C. R. Reynolds (Eds.), *The Handbook of School Psychology* (4th ed., pp. 434–460). Hoboken, NJ: Wiley.
- Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(3), e1355.
- Scammacca, N., Fall, A. M., Capin, P., Roberts, G., & Swanson, E. (2020). Examining factors affecting reading and math growth and achievement gaps in grades 1–5: A cohort-sequential longitudinal approach. *Journal of Educational Psychology*, 112(4), 718–734.
- Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York: Oxford University Press.
- Stanovich, K. E., Nathan, R. G., & Vala-Rossi, M. (1986). Developmental changes in the cognitive correlates of reading ability and the developmental lag hypothesis. *Reading Research Quarterly*, 21(3), 267–283.
- TNTP (2018). *The Opportunity Myth: What Students Can Show Us About How School Is Letting Them Down—and How to Fix It*.  
[https://tntp.org/tntp\\_the-opportunity-myth\\_web](https://tntp.org/tntp_the-opportunity-myth_web) (검색일: 2026. 04. 23.)
- Tzelgov, J., & Henik, A. (1991). Suppression situations in psychological research: Definitions, implications, and applications. *Psychological Bulletin*, 109(3), 524–536.

ABSTRACT

## Validating the Concept of Achievement Acceleration Using Machine Learning

**Jeongmin Ji**

Associate Research Fellow, Korea Institute for Curriculum and Evaluation

This study seeks to validate the concept of achievement acceleration as an indicator for capturing students' learning growth with greater precision. Conventional growth-referenced assessment has primarily focused on changes in achievement level, and thus has been limited in its capacity to adequately reflect subtle shifts in the direction or dynamic properties of student growth. Achievement acceleration is a higher-order indicator characterized by nonlinear and long-term dependent structures, and its valid measurement requires specialized models capable of capturing these properties.

As a measurement tool suited to these characteristics, this study applies a Deep-LSTM-based time-series model that simultaneously accounts for nonlinearity, long-term dependency, and structured missingness patterns. Vertically scaled achievement scores from the Daegu Education Longitudinal Study (DELS), covering the years 2017 through 2024, were used as the analytic data.

The results confirm that the Deep-LSTM model is equipped with a predictive structure consistent with the conceptual definition of achievement acceleration. Long-term cumulative growth context information was also found to contribute substantively to acceleration prediction. Furthermore, in the criterion-related validity analysis, the achievement acceleration indicator generated by the Deep-LSTM provided additional explanatory power for subsequent rates of achievement change, even after controlling for prior achievement level and rate of change. This indicates that achievement acceleration converges with existing indicators while remaining distinct as an independent construct. Alongside these quantitative findings, multiple illustrative cases demonstrate that achievement acceleration can be interpreted within the framework of growth-referenced assessment and utilized as educationally

meaningful evaluative information.

Taken together, this study suggests that an acceleration-based indicator—one that reflects the fine-grained structural changes in growth trajectories—can function as meaningful evaluative information in long-term educational longitudinal data, extending beyond achievement level or simple growth magnitude.

*Key Words: Growth-referenced assessment, Achievement acceleration, Machine learning, Daegu Education Longitudinal Study (DELS), Validation*