

교육과정평가연구

The Journal of Curriculum and Evaluation

2026, Vol. 29, No. 3, pp.1~30

DOI: <https://doi.org/10.29221/jce.2026.29.3.1>

고등학교 성취평가제 구현을 위한 성취기준 진술 방식의 재고: 2022 개정 교육과정 국어·수학·영어 교과를 중심으로

성민선 (건국대학교 박사수료)*

공미선 (건국대학교 박사수료)

김종훈 (건국대학교 부교수)**

요약

2025년 고교학점제의 전면 시행은 고등학교 평가 체제를 상대적 서열화로부터 준거참조적 성취평가제로 전환할 것을 요구한다. 성취평가제가 교육과정의 성취기준을 실질적 평가 준거로 삼는 체제임을 전제할 때, 제도의 실질적 작동은 점수 기준의 설정이 아니라 성취기준 진술의 질적 수준에 달려 있다. 이에 본 연구는 2022 개정 교육과정의 고등학교 공통 과목인 공통국어1·2, 공통수학1·2, 공통영어1·2의 성취기준을 대상으로, 이들이 평가 준거로서 갖추어야 할 요건을 어느 정도 충족하는지를 체계적으로 분석하였다. 이론적 검토를 통해 행동 동사의 인지 수준 명료성, 내용·수행의 식별 가능성, 평가 증거의 도출 가능성, 성취 조건·맥락의 명시 여부라는 네 가지 기준으로 분석 틀을 구성하였으며, 세 명의 공동 연구자의 독립 분석과 합의 과정을 거쳐 판정의 신뢰도를 확보하였다. 분석 결과, 성취기준 진술 방식은 태도·참여형, 과정·전략 포괄형, 복합 행위 혼재형, 수행 명료형의 네 유형으로 분류되었으며, 전체적으로 복합 행위 혼재형이 가장 큰 비중을 차지하였다. 교과별로는 수학이 복합 행위 혼재형에, 국어가 복합 행위 혼재형과 태도·참여형에, 영어가 과정·전략 포괄형과 수행 명료형으로 양분되는 양상을 보였으며, 특히 영어는 공통영어1과 공통영어2의 일부 성취기준이 동일하게 진술되어, 성취기준 본문만으로는 과목별 기대 수행 수준의 차이를 직접 확인하기 어려웠다. 결과를 바탕으로 본 연구는 핵심 수행의 명료성, 내용·수행 이원 구조, 평가 증거 가시성, 성취수준 분화 가능성, 과목 간 수준 명료화라는 진술 원리와 교과별 개선 예시를 제안하고, 차기 교육과정 개정 시 성취기준 개발 절차에 평가 가능성 검토 단계를 제도화할 것을 제안하였다.

주제어 : 고교학점제, 성취평가제, 성취기준, 성취수준, 2022 개정 교육과정

* 제1저자, minsunxz@konkuk.ac.kr

** 교신저자, ilovetem@konkuk.ac.kr

I. 서론

2025년 고교학점제가 전면 시행됨에 따라 한국 고등학교 교육은 구조적 전환을 맞이하게 되었다. 학생이 자신의 진로와 적성에 따라 과목을 선택하고, 일정 학점을 이수함으로써 졸업 요건을 충족하는 이 제도는 학교 교육과정의 설계와 운영, 나아가 학생 평가체제의 근본적인 재편을 요구한다. 특히, 고교학점제 하에서 학점 이수의 기준이 되는 최소 성취수준의 설정과 적용은 평가를 상대적 서열화의 도구에서 벗어나 교육적 도달 여부를 판단하는 준거 참조적 기제로 전환하는 것을 전제로 한다. 이러한 전환의 핵심 제도적 장치가 바로 성취평가제다.

고교학점제에서 과목 이수 여부는 학업성취율과 출석률 등 제도상 설정된 기준에 따라 판정되므로, 성취기준 진술의 언어적 명료성이 학생의 이수 여부를 직접 결정한다고 보기는 어렵다. 그러나 이수 판정의 근거가 되는 학업성취율은 학교가 설계한 평가 과제와 채점 기준을 통해 산출되며, 이러한 평가 설계는 교육과정에 제시된 성취기준과 성취수준을 근거로 이루어진다. 즉 성취기준은 이수 판정에 직접 적용되는 행정적 기준은 아니나, 이수 판정에 활용되는 평가 결과의 내용적 타당성과 설명 가능성을 뒷받침하는 기반이 된다. 과목 이수 자체는 정량적 기준에 따라 판정되지만, 그 판정에 사용되는 학업성취율의 타당성은 성취기준·성취수준·평가 도구 사이의 연계에 의해 뒷받침되며, 성취기준 진술의 질은 바로 이 연계가 성립하기 위한 조건이라 할 수 있다.

흔히 절대평가로 알려진 성취평가제는 국가 교육과정에 제시된 성취기준과 관련하여 구체적인 의미를 지닌다. 기존의 절대평가가 사전에 설정된 점수 기준에 따라 학생의 성취도를 분류하는 방식을 따른다면, 성취평가제는 교육과정에 명시된 성취기준(achievement standards)을 실질적 평가 준거로 삼아 학생이 그 기준에 어느 정도 도달하였는지를 판단하는 준거 참조 평가(criterion-referenced assessment) 체제이다(김순남, 이병환, 2018; 고영빈, 지은림, 2022). 이 점에서 성취평가제의 실질적 작동은 점수 기준의 설정이 아니라 성취기준 자체의 질적 수준에 달려 있다. 학생이 무엇을 알고 무엇을 할 수 있어야 하는가를 분명하게 기술하는 성취기준이 없다면, 평가는 그 준거를 잃고 다시 자의적이거나 상대적인 판단에 의존할 수밖에 없기 때문이다.

따라서 성취기준이 평가 준거로서 실질적으로 기능하기 위해서는 그 진술 방식이 특정 요건을 갖추어야 한다. 평가 준거로서의 성취기준은 교사가 어떤 수행을 어떤 증거로 판단해야 하는지를 합리적으로 추론할 수 있도록 진술되어야 하며, 나아가 그 수행이 어느 수준으로 이루어졌는지를 등급과 같은 방식으로 구분할 수 있도록 성취수준 기술과 논리적으로 연계되어야 한다(백남진, 2014; 서영진, 2013). 성취기준의 진술이 지나치

게 포괄적이거나 모호하면, 개별 교사는 동일한 성취기준으로부터 상이한 평가 과제와 채점 기준을 도출하게 되며, 이는 학교 간·교사 간 평가의 일관성을 저해하고 성취평가제의 공정성에 대한 신뢰를 훼손한다.

이러한 문제의식은 이미 여러 선행연구에서 제기되어 왔다. 백남진(2014)은 2009 개정 과학 교육과정을 대상으로 한국의 성취기준이 수행 기준(performance standard)이 아닌 내용 기준(content standard)의 성격을 띠고 있음을 지적하였고, 서영진(2013)은 역량 중심 교육과정의 취지를 구현하기 위해 성취기준의 기능 진술을 보다 구체화해야 함을 주장하였다. 조인영, 강완(2018)은 초등학교 수학과 교육과정 성취기준의 진술 방식을 분석하여, 성취기준이 ‘기능’을 충분히 담아내지 못한 채 내용 중심으로 진술되어 있다는 문제를 실증적으로 확인하였으며, 윤지영, 온정덕(2025)은 2022 개정 교육과정의 성취 수준 설정 방식이 성취기준의 내용 요소를 단순히 수준 구분의 기준으로 활용함으로써 역량 기반 교육의 특성을 반영하지 못하고 있다고 지적하였다. 이처럼 성취기준의 진술 방식과 성취수준과의 연계 문제는 교육과정 연구 분야에서 꾸준히 제기되어 왔으나, 고교학점제와 연계하여 성취평가제를 전면적으로 적용하기 위한 제도적·정책적 맥락 속에서 고등학교 주요 교과를 대상으로 체계적으로 분석한 연구는 아직 충분하지 않다.

성취기준의 진술 방식 문제는 교사의 평가 전문성 부족의 문제로만 환원되지 않는다. 아무리 평가 역량이 뛰어난 교사라도, 평가 준거로서의 요건을 갖추지 못한 성취기준으로부터 일관되고 타당한 평가 설계를 끌어내기는 구조적으로 어렵다. 따라서 이 문제는 교사연수나 교수·학습 지원의 차원이 아니라, 성취기준 개발의 원리와 절차의 차원에서 다루어져야 한다. 실제로 성취평가제 운영의 현장에서는 교사들이 국가 교육과정의 성취기준을 평가 준거로 직접 활용하기 어렵다는 점을 인식하여 평가준거 성취기준을 별도로 개발·운영하도록 하고 있는데(김영은 외, 2024; 이미경 외, 2017), 이는 역설적으로 현행 성취기준의 진술 방식이 평가 준거로서의 기능을 충분히 수행하지 못하고 있음을 인정하는 것이라 할 수 있다.

이에 본 연구는 2022 개정 교육과정의 고등학교 공통 교과의 과목, 구체적으로 공통 국어1·2, 공통영어1·2, 공통수학1·2의 성취기준을 분석 대상으로 삼아, 이들 성취기준이 평가 준거로서 갖추어야 할 요건을 어느 정도로 충족하고 있는지를 체계적으로 분석하고 그 한계를 진단한다. 세 교과를 분석 대상으로 선정한 것은, 국어·영어·수학이 고등학교 공통 교육과정의 핵심 교과로서 모든 학생이 이수하는 과목이라는 점, 그리고 인문·사회적 성격의 언어 교과와 수리·논리적 성격의 수학 교과를 함께 분석함으로써 교과의 특성에 따른 성취기준 진술 방식의 차이를 비교적 관점에서 검토할 수 있다는 점을 고려한 것이다. 분석의 이론적 토대로는 Bloom의 개정 교육목표분류체계(Anderson & Krathwohl, 2001), Mager의 행동목표 진술 이론(Mager, 1997), Wiggins & McTighe

의 백워드 설계(Wiggins & McTighe, 2005), 그리고 Webb의 지식의 깊이(Depth of Knowledge) 모형(Webb, 1997)을 절충적으로 활용한다. 아울러 미국의 Common Core State Standards(California Department of Education, 2013, 2023)와 IB(International Baccalaureate) 교육과정(International Baccalaureate Organization, 2026a, 2026b)의 기준 진술 사례를 보조적으로 검토함으로써 대안적 진술 방식의 방향을 탐색한다.

본 연구의 궁극적 목표는 비판적 진단에 그치지 않고, 성취평가제의 실질적 구현을 가능하게 하는 성취기준의 대안적 진술 방식을 원리 수준에서 제안하는 것이다. 이를 통해 차기 교육과정 개정 시 성취기준 개발 과정에서 평가 가능성 검토가 체계적으로 이루어질 수 있도록 하는 정책적 근거를 제공하고자 한다.

II. 이론적 배경

1. 성취기준의 이중적 기능과 평가 준거로서의 요건

교육과정에 제시된 ‘성취기준’은 본래 두 가지 목적을 동시에 수행하도록 설계된 개념이다. 하나는 교수·학습의 안내자로서, 교사가 무엇을 가르쳐야 하고 학생이 무엇을 배워야 하는지를 규정하는 내용 기준(content standard)의 역할이다. 다른 하나는 평가의 준거로서, 학생이 기대하는 수준에 도달했는지를 판단하는 기준의 역할이다. 한국의 국가 교육과정에 성취기준이 제7차 교육과정을 시작으로 도입된 이래 지금까지 이 두 역할은 하나의 진술문 안에 통합되어 왔다. 그러나 이 두 역할은 서로 다른, 때로는 상충하는 진술 방식을 요구한다는 점에서 근본적인 긴장 관계를 내포하고 있다(이승미, 박순경, 2014; 한수현, 김영실, 2025).

교수·학습의 안내자로서 성취기준은 교과와 핵심 개념과 학습 방향을 포괄적으로 담아내야 한다. 교사가 교육과정에 대한 권한을 기반으로 자신의 수업을 다양하게 설계할 수 있는 여지를 남겨야 하고, 지식·이해·과정·기능·가치·태도의 여러 범주를 유기적으로 아우를 수 있어야 한다. 성취기준의 이러한 역할에서는 다소 포괄적이고 해석의 여지가 있는 진술이 오히려 교사의 자율성을 보장하는 장점이 된다. 반면, 평가의 준거로서 성취기준은 정반대의 요건을 갖추어야 한다. 학생이 무엇을 할 수 있어야 하는지가 관찰 가능하고 판단 가능한 형태로 진술되어야 하며, 서로 다른 교사가 동일한 성취기준을 읽고 유사한 평가 과제와 채점 기준에 도달할 수 있어야 한다. 이 역할에서는 모호성은 곧 결함을 의미한다.

한국 국가 교육과정 성취기준의 개발 역사를 보면, 이 두 역할 사이의 긴장이 제도적으로 해소되지 못한 채 이어져 왔음을 알 수 있다. 이승미, 박순경(2014)은 2009 개정 교육과정의 성취기준 실태를 분석하면서, ‘내용’과 ‘수행’의 기술 방식이 형식적으로 고정되어 있어 교과 특성을 반영하지 못하고 있으며, 평가 기준으로서의 성취기준은 교육과정 문서와 별도로 개발되는 이중 구조가 고착되어 있음을 지적하였다. 이 문제는 2015 개정 교육과정에서도 해소되지 않았다. 국가 교육과정의 성취기준이 실제 평가 상황에서 준거로 사용하기에 다소 포괄적이거나 추상적으로 진술되어 있다는 인식에 따라 별도의 평가준거 성취기준을 교과별로 개발·보급하는 체제가 유지되었던 것이 이를 방증한다(이미경 외, 2017). 현행 2022 개정 교육과정도 예외가 아니어서, 성취기준의 두 역할 사이의 긴장은 지금까지도 미결의 과제로 남아 있다.

그렇다면 성취평가제의 전면 시행이라는 맥락에서 이러한 긴장은 어떤 의미를 가지는가? 성취평가제가 단순한 절대평가가 아니라 교육과정의 성취기준을 실질적 평가 준거로 삼는 체제임을 전제할 때, 성취기준의 평가 준거 기능은 선택 사항이 아니라 제도의 논리적 기반이 된다. 즉, 준거 참조 평가의 본질은 미리 설정된 기준에 비추어 학생의 수행을 판단하는 것이다. 이때 기준이 불명확하면 학생의 수행에 대한 판단 자체가 성립하지 않는다. 성취기준이 무엇을 어느 수준에서 수행해야 하는지를 명시하지 못한다면, 교사는 결국 자신의 직관이나 경험에 의존하여 성취수준을 설정할 수밖에 없게 되고, 이는 학교 간·교사 간 평가 결과의 비교 가능성을 근본적으로 저해한다(김순남, 이병환, 2018).

이러한 사실에 비추어 볼 때, 평가 준거로서 기능할 수 있는 성취기준이 갖추어야 할 요건은 크게 네 가지로 정리된다. 가장 먼저 요구되는 것은 수행 동사의 명료성이다. Anderson & Krathwohl(2001)이 Bloom의 교육목표분류체계를 개정하면서 인지 과정의 여섯 수준(기억, 이해, 적용, 분석, 평가, 창안)을 체계화한 것은, 교육목표의 동사가 학생에게 기대하는 인지적 수행의 수준을 결정한다는 인식에 기반한다. 그러나 현행 교육과정의 성취기준에서 빈번하게 사용되는 ‘이해한다’, ‘탐구한다’, ‘참여한다’와 같은 동사들은 이 분류체계상의 어느 수준에 해당하는지 판단하기 어렵고, 결과적으로 어떤 수행으로 ‘이해했음’을 입증할 수 있는지가 불분명하다(김선영, 2016; 백남진, 2014). 동사가 명료해지더라도, 하나의 성취기준 안에서 ‘무엇을(내용)’ ‘어떻게 할 수 있는가(수행)’가 구조적으로 명확해야 비로소 평가의 대상이 특정된다는 점에서 내용과 수행의 식별 가능성은 그다음 요건이라고 할 수 있다. 다음으로, 수행이 실제 평가로 이어지기 위해서는 그 진술로부터 평가의 증거가 도출될 수 있어야 한다. 끝으로, 수행이 이루어지는 조건과 맥락이 제시되어야 한다. 동일한 수행 동사도 자료의 유형과 복잡도, 수행 범위 및 상황적 조건에 따라 요구되는 수행 수준이 달라지므로, 조건과 맥락은 평가 과제

의 범위와 난이도를 결정하는 독립적인 요소이다. Mager(1997)가 행동 목표 진술 요소로 조건, 수행, 준거를 제시한 것은 목표가 평가와 긴밀하게 연결되려면 구체적인 수행 맥락과 도달 기준이 명시되어야 한다는 원리를 체계화한 것이다.

결국, 성취기준의 이러한 이중 기능은 어느 한쪽을 포기함으로써 해소될 수 없다. 성취기준이 수행 동사를 명확히 하고, 식별 가능한 내용과 수행을 제시하며, 평가 증거를 자연스럽게 시사하고 수행의 조건과 맥락을 구체화할 때, 그 기준은 교사에게 수업 설계와 평가 설계 양쪽에서 모두 유용한 안내를 제공할 수 있다.

2. 성취기준-성취수준-평가의 정렬: 이론적 토대와 사례

성취기준이 평가 준거의 요건을 갖추었다고 하여 성취평가제가 곧바로 작동하는 것은 아니다. 진술의 명료성은 성취평가제를 위한 필요조건임에는 분명하나 충분조건은 되지 못하기 때문이다. 성취기준이 아무리 명료하게 진술되어도, 그것이 실제 평가 설계 및 성취수준 기술과 논리적으로 연결되지 않는다면 성취평가제는 형식적 제도에 머물 수밖에 없다. 교육과정·수업·평가의 정렬(alignment)은 성취평가제의 실질적 작동을 위한 구조적 전제이며, 이 정렬의 출발점은 성취기준의 진술 방식에 있다.

교육과정·수업·평가 간 정렬의 논리를 가장 체계적으로 제시한 이론은 Wiggins & McTighe(2005)의 백워드 설계(backward design)이다. 이들은 교육과정 설계의 순서를 관습적으로 해 오던 방식과 반대로 뒤집을 것을 제안한다. 즉, 먼저 학습을 통해 도달하기를 기대하는 결과를 명료하게 설정하고(identifying desired results), 다음으로 그 결과에 도달했음을 입증하는 수용 가능한 증거를 결정한 뒤(determining acceptable evidence), 마지막으로 수업 경험과 학습 활동을 계획하는 순서다(planning learning experience and instruction). 이 설계 원리에서 핵심은 성취기준이 ‘이해의 증거’로부터 역으로(backward) 규정될 수 있어야 한다는 것이다.

Webb(1997)의 지식의 깊이(Depth of Knowledge, DoK) 모형은 정렬의 문제를 인지적 복잡도의 측면에서 보완하는 관점을 제공한다. Webb은 교육 목표·수업·평가가 단순히 동일한 내용을 다루는 것만으로는 정렬되었다고 볼 수 없으며, 세 요소가 요구하는 인지적 복잡도의 수준이 일치해야 함을 강조한다. DoK는 수준 1(회상 및 재생), 수준 2(기능과 개념), 수준 3(단기 전략적 사고), 수준 4(확장적 사고)의 네 단계로 구성된다. 성취기준의 진술이 인지적 복잡도를 명시하지 않으면, 교사는 성취기준이 요구하는 인지 수준을 판단할 수 없고, 결과적으로 평가 문항이 성취기준이 겨냥하는 인지 수준에서 체계적으로 이탈할 위험이 발생한다(백남진, 2014; 윤지영, 온정덕, 2025).

정렬의 또 다른 핵심 고리는 성취기준과 성취수준 기술 사이의 논리적 연계이다. 성취수준이 성취기준과 논리적으로 연계되려면, 성취기준에서 평가의 핵심이 되는 수행을 식별할 수 있어야 한다. 수행이 명확하게 제시될 때 ‘이 수행을 자립적으로 완전하게 해낼 수 있다(A)’, ‘보조를 받아 부분적으로 할 수 있다(C)’, ‘기초적 수준에서만 접근할 수 있다(E)’와 같이 수행의 질적 차이를 위계적으로 기술하는 것이 가능하다. 윤지영, 온정덕(2025)은 2022 개정 교육과정에 따른 성취수준 설정이 성취기준의 내용 요소를 수준 구분의 기준으로 활용함으로써 인지적 엄격성을 반영하지 못하고 있다고 비판하는데, 이는 성취기준과 성취수준의 정렬이 구조적으로 이루어지지 않고 있음을 직접 지적한 것이다.

해외 교육과정 사례에서도 성취기준과 평가 사이의 정렬을 구체화하는 방식을 확인할 수 있다. 캘리포니아 수학 교육과정은 Common Core에 기반한 교과 내용 기준과 수학적 실천 기준을 함께 제시하며, 일부 성취기준에서는 학생에게 기대되는 수행과 그 판단 근거를 비교적 구체적으로 드러낸다(California Department of Education, 2013, 2023). 예를 들어 대수 기준 A-REI.1은 방정식 풀이의 각 단계가 앞선 단계에서 성립한 등식으로부터 어떻게 도출되는지를 설명하고, 풀이 방법을 정당화하는 타당한 논증을 구성하도록 요구한다(California Department of Education, 2013). 두 성취기준은 다루는 수학 내용이 다르지만 진술 구조의 측면에서 보면, A-REI.1은 ‘항등식의 성질과 나머지정리를 이해하고, 이를 활용하여 문제를 해결할 수 있다’고 진술한 [10공수 1-01-02]보다 학생이 산출해야 할 수행과 그 판단 근거를 직접적으로 제시한 사례이다. IB의 Middle Years Programme 언어 습득 교육과정은 학습을 여섯 단계로 조직하고, 교과 목표와 평가 기준을 연계하여 학생의 수행을 판단하도록 구성한다(International Baccalaureate Organization, 2026a, 2026b). 예를 들어 읽기 평가 기준은 학생이 글의 언어적·공간적·시각적 요소로부터 의미를 구성하고, 이러한 요소들이 결합하여 생각과 가치 및 태도를 전달하는 방식을 해석하도록 요구한다(International Baccalaureate Organization, 2020). 또한 읽기·쓰기·보기·말하기 수행을 평가 과제와 연계하고 그 결과를 숙달 수준으로 구분함으로써 기대 수행과 평가 판단 사이의 관계를 구체화한다(International Baccalaureate Organization, 2026b). 이러한 사례는 성취기준에 제시된 내용과 수행으로부터 평가 증거와 성취수준 기술을 얼마나 논리적이고 일관되게 도출할 수 있는지가 성취기준의 평가 준거 적합성을 판단하는 중요한 기준임을 보여준다.

이상의 논의를 종합하면, 성취기준·성취수준·평가의 정렬은 네 가지 요건이 충족될 때 구조적으로 가능해진다. 즉, 행동 동사가 요구하는 인지적 수행의 수준이 명료해야 하고, 하나의 성취기준 안에서 평가의 대상이 되는 내용과 수행이 구조적으로 식별 가능해야 하며, 진술로부터 평가 증거를 합리적으로 도출할 수 있어야 한다. 또한, 수행의 범위

와 요구 수준을 판단할 수 있도록 성취 조건과 맥락이 제시되어야 한다. 이 요건들은 다음 장에서 현행 성취기준을 분석하기 위한 틀의 이론적 근거를 구성한다.

III. 연구 방법

1. 분석 대상

본 연구는 2022 개정 교육과정의 고등학교 공통 과목 중 공통국어1·2, 공통영어1·2, 공통수학1·2에 수록된 성취기준 전체를 분석의 대상으로 삼는다. 세 교과는 고등학교 공통 교육과정의 핵심 교과로서 대부분의 학생이 이수하며, 따라서 성취평가제의 실질적 작동에서 가장 광범위하게 영향을 미치는 교과라고 할 수 있다. 또한, 국어·영어는 언어 기반의 교과, 수학은 논리·수리적 교과로서 성격이 상이하여 성취기준 진술 방식의 교과 맥락 의존성을 비교적 관점에서 검토하기에 적합하다.

분석 대상 성취기준의 수는 다음과 같다. 국어는 공통국어1의 14개, 공통국어2의 15개로 총 29개이며, 수학은 공통수학1의 19개, 공통수학2의 20개로 총 39개이다. 영어는 공통영어1의 16개, 공통영어2의 17개로 총 33개이다. 세 교과를 합산하면 분석 대상인 성취기준은 총 101개이다. 분석은 교육과정에 진술된 성취기준 본문만을 직접 대상으로 하며, 성취기준 해설 및 성취기준 적용 시 고려사항은 진술 방식 분석의 보조 자료로 참조하되 분석의 직접 대상에는 포함하지 않는다. 이는 성취기준 해설이나 고려사항이 교사에게 참고 정보를 제공하더라도 교육과정과 평가의 준거는 어디까지나 성취기준 본문이기 때문이다.

2. 분석의 틀

본 연구의 분석 틀은 II장의 이론적 논의로부터 도출된 네 가지 기준으로 구성된다. 아래 <표 1>은 각 기준의 이론적 근거, 핵심 판단 질문, 적합·부분 적합·부적합 판단 기준을 보여준다.

<표 1> 성취기준 평가 준거 적합성 분석 틀

분석 기준	이론적 근거	핵심 판단 질문	적합/부분 적합/부적합 판단 기준
기준 1: 행동 동사의 인지 수준 명료성	Anderson & Krathwohl(2001)의 Bloom 개정 교육목표분류체계, Webb(1997)의 지식의 깊이(DoK)	행동 동사가 인지적 수행에서 어느 수준에 해당하는지 식별 가능한가? 동사가 수행의 인지적 수준과 사고의 깊이를 명확히 지시하는가?	•적합: 분석한다, 평가한다, 증명한다, 구성한다 등 인지적 수행의 성격과 사고 수준이 명확한 동사 •부분 적합: 설명한다, 해석한다, 적용한다, 활용한다, 표현한다 등 수행 내용·맥락에 따라 판단이 달라질 수 있는 동사 •부적합: 이해한다, 탐구한다, 참여한다, 인식한다 등 인지 수준 불명확한 동사
기준 2: 내용·수행의 식별 가능성	Mager(1997)의 행동목표 진술론, 이승미, 박순경(2014)의 내용·수행 기술 방식 분석	단일 기준 내에서 '무엇을(내용)'과 '어떻게 할 수 있는가(수행)'가 구조적으로 식별 가능한가? 복수의 이질적 수행이 혼재하지 않는가?	•적합: 내용 요소와 수행 요소가 명확히 식별되어 단일 수행이 진술된 경우 •부분 적합: 내용 요소와 수행 요소가 식별되나 복수의 내용 또는 수행이 포함된 경우 •부적합: 복수의 이질적 행위가 접속사로 나열되거나 인지적 수행과 태도가 혼재된 경우
기준 3: 평가 증거의 도출 가능성	Wiggins & McTighe(2005)의 백워드 설계, 백남진(2014)의 수행 기준 진술 방향 탐색	진술만으로 교사가 어떤 평가 과제와 산출물을 설계해야 하는지 합리적으로 추론할 수 있는가?	•적합: 진술로부터 평가 증거(응답, 설명, 산출물, 수행 결과 등)가 자연스럽게 시사되는 경우 •부분 적합: 진술로부터 평가 과제 방향은 추론 가능하나 평가 증거의 유형이나 판단 초점이 모호한 경우 •부적합: 어떤 수행으로 도달을 입증하는지 불분명한 경우
기준 4: 성취 조건· 맥락의 명시 여부	Mager(1997)의 조건·수행·준거의 3요소 모형, 김선영(2016)의 역량 기반 기능 진술 탐색	수행이 이루어지는 상황적 조건이나 맥락(자료 유형, 인지적 맥락, 수행 범위 등)이 진술에 포함되어 있는가?	•적합: 수행의 조건이나 범위가 진술에 명시된 경우 •부분 적합: 수행의 조건이나 범위가 진술에 일부 제시된 경우 •부적합: 조건 없이 수행 동사만 제시되어 어떤 상황에서의 수행인지 불명확한 경우

위 네 가지 기준은 독립적이면서도 상호보완적인 관계를 형성한다. 기준 1(행동 동사의 인지 수준 명료성)은 성취기준이 평가 준거로 기능하기 위한 가장 기초적인 요소이다. 기준 2(내용·수행의 식별 가능성)는 단일 성취기준 내에서 평가해야 할 대상이 특정될 수 있는지를 판단한다. 기준 3(평가 증거의 도출 가능성)은 성취기준과 평가 설계 사

이의 정렬 가능성을 직접 겨냥하며, 기준 4(성취 조건·맥락의 명시 여부)는 수행의 상황적 맥락이 얼마나 구체화되어 있는가를 검토한다. 각 성취기준은 네 기준에 따라 ‘적합’, ‘부분 적합’, ‘부적합’의 세 단계로 판정하며, 인지적 수행 수준이 식별 가능한 행동 동사가 제시되어 기준 1이 적합하더라도, 자료의 유형이나 수행 범위가 제시되지 않았다면 기준 4는 부분 적합 또는 부적합으로 판정할 수 있다.

네 가지 기준의 판정 결과는 다음과 같은 방식으로 유형 판정에 반영되었다. 기준 1과 기준 3이 함께 부적합으로 판정되어 종결 동사가 태도나 참여로 수렴하는 경우는 태도·참여형(A)으로, 기준 1이 적합 또는 부분 적합이나 기준 4(성취 조건·맥락)가 부적합 또는 부분 적합에 그쳐 수행의 범위나 조건이 특정되지 않는 경우는 과정·전략 포괄형(B)으로 분류하였다. 기준 1은 충족되더라도 기준 2(내용·수행의 식별 가능성)가 부분 적합 또는 부적합으로 판정되어 복수의 이질적 수행이나 내용이 하나의 진술에 결합된 경우는 복합 행위 혼재형(C)으로, 기준 1·2·3이 모두 적합에 해당하는 경우는 수행 명료형(D)으로 분류하였다. 하나의 성취기준이 둘 이상의 유형적 특성을 동시에 보이는 경우에는 표장 3절에서 서술한 바와 같이 평가 증거의 도출을 가장 직접적으로 제약하는 기준의 판정 결과를 중심으로 대표 유형을 결정하였다.

3. 분석 절차

분석은 세 명의 공동 연구자가 동일한 분석 기준을 각자 독립적으로 수행한 후, 판정의 일치도와 불일치 양상을 상호 비교·검토하여 합의에 이르는 조정 과정을 거쳐 진행되었다. 세 연구자 중 1명은 교육과정 전공 교수이며, 2명은 교육과정 전공 박사과정을 수료한 연구자이다. 세 연구자는 모두 국가 교육과정 및 교육과정 분야의 연구를 수행해 온 전문성과 경험을 바탕으로 분석에 참여하였다.

분석에 앞서 연구자 간 <표 1>의 분석 틀과 유형 분류 기준, 그리고 ‘적합·부분 적합·부적합’의 판정 준거를 공유하고, 사전 협의를 통해 판정 기준의 적용 방식과 유형 간 경계를 구체화하였다. 이를 토대로 분석은 다음 세 단계로 진행되었다. 우선 세 연구자는 각자 다른 연구자의 판정을 참조하지 않은 상태에서 101개의 성취기준 각각에 대해 네 가지 기준을 순서대로 적용하여 독립적으로 판정하고, 판정의 근거를 진술의 언어적 특징과 연결하여 기술하였다. 이 단계에서는 성취기준 해설 및 적용 시 고려사항을 병행 검토하되, 해설이 진술의 모호성을 보완하는 방식으로 작성된 경우, 이를 진술 자체의 한계에 대한 반증 자료로 활용하였다. 이어서, 개별 분석 결과를 바탕으로 성취기준 진술 방식의 유형을 귀납적으로 도출하였다. 유형 분류는 네 가지 분석 기준에서 나타나는

결함의 패턴을 기준으로 이루어졌으며, 유형별로 대표 사례를 제시하여 분석의 구체성을 확보하였다. 마지막으로 각 연구자는 국어·영어·수학 세 교과 사이에서 나타나는 진술 방식의 경향성 차이를 비교적 관점에서 기술하였다.

개별 분석이 완료된 후에는 세 연구자의 독립 판정 결과에 대한 연구자 간 일치도(inter-rater agreement)를 산출하였다. 각 성취기준에 부여된 유형을 명목 변수로 보아 연구자 쌍별 단순 일치율과 우연에 의한 일치를 보정한 Cohen's Kappa(κ)를 계산하고, 세 연구자 전체의 일치도는 Fleiss' Kappa(κ)로 함께 산출하였다. 그 결과 세 연구자 쌍의 단순 일치율은 86.1~87.1%, Cohen's Kappa(κ)는 .798~.810으로 나타났으며, 세 연구자 전체의 Fleiss' Kappa(κ)는 .801로 Landis & Koch(1977)의 기준상 '상당한 일치(substantial agreement)'에 해당하였다.

<표 2> 연구자 쌍별 일치도

연구자 쌍	전체 성취기준 수	일치	일치율	Cohen's κ
A-B	101	87	86.1%	.798
A-C		88	87.1%	.810
B-C		87	86.1%	.798
3인 전체 (Fleiss' κ)	101	-	-	.801

다만 연구자 간 일치도는 교과에 따라 차이를 보였다. 아래 <표 3>에서 보듯, 수학에서 일치도가 가장 높았던 반면(Fleiss' κ = .885), 영어는 다른 두 교과보다 유형 판정의 차이가 상대적으로 크게 나타났다(국어 .802, 영어 .661). 이는 유형 판정의 난도가 성취기준 진술 방식과 직접 관련됨을 시사한다. 다만 교과별 일치도의 차이만으로 성취기준 진술의 모호성을 직접 입증할 수는 없으므로, 불일치 항목에 대해 연구자별 판정 근거를 비교하고 네 가지 분석 기준과 유형별 조작적 정의를 다시 적용하여 최종 판정을 조정하였다.

<표 3> 교과별 연구자 간 일치도

교과	성취기준 수	3인 만장일치	일치율	Fleiss' κ
국어	29	23	79.3%	.802
수학	39	36	92.3%	.885
영어	33	22	66.7%	.661
전체	101	81	80.2%	.801

연구자 간 판정이 불일치한 항목의 조정은 독립 판정과 연구자 간 일치도 산출을 완료한 후에 실시하였다. 먼저 각 연구자는 해당 성취기준에 부여한 유형과 그 근거가 된 수행 동사, 내용과 수행의 결합 방식, 평가 증거 및 수행 조건을 제시하였다. 이후 세 연구자는 <표 1>의 네 가지 분석 기준과 유형별 조작적 정의를 항목별로 다시 적용하여 판정 근거를 비교하였다. 하나의 성취기준이 둘 이상의 유형적 특성을 동시에 보이는 경우에는 평가 증거의 도출을 가장 직접적으로 제약하는 진술 특성을 중심으로 대표 유형을 결정하였다. 최종 판정은 세 연구자가 모두 동의한 유형으로 확정하였으며, 합의에 이르지 못한 경우에는 성취기준 해설과 적용 시 고려사항을 보조적으로 재검토한 후 다시 논의하였다. 예를 들어 [10공영1-01-02] ‘말이나 글의 주제나 요지를 파악한다’에 대해서는 과정·전략 포괄형과 수행 명료형으로 판정이 갈렸다. 일부 연구자는 텍스트의 유형과 난이도, 수행 조건 및 평가 산출물이 제시되지 않아 평가 증거를 구체화하기 어렵다는 점을 중심으로 과정·전략 포괄형으로 판정하였다. 다른 연구자는 ‘파악한다’가 단일한 인지적 수행을 제시하고 내용과 수행도 식별 가능하다는 점에서 수행 명료형으로 판정하였다. 조정 과정에서는 수행 조건의 부족을 기준 4의 문제로 구분하고, 유형 분류에서는 단일 수행이 명료하게 제시되어 있다는 진술의 주된 특성을 반영하여 수행 명료형으로 최종 합의하였다. IV장에 제시한 유형 분류와 빈도는 이러한 전원 합의를 거쳐 확정된 최종 판정에 따른 것이다.

IV. 연구 결과

1. 성취기준 진술 방식의 유형 분류

2022 개정 교육과정 공통국어1·2, 공통영어1·2, 공통수학1·2의 성취기준 101개를 Ⅲ장의 분석 틀에 따라 검토한 결과, 진술 방식의 특성에 따라 네 가지 유형이 귀납적으로 도출되었다. 세 연구자의 독립 판정을 합의로 확인한 유형 분포는 <표 4>와 같다. 각 유형은 상호 배타적이지 않으며, 하나의 성취기준이 두 유형의 특성을 동시에 보이는 경우도 있다. 그러나 각 유형은 평가 준거로서의 기능을 저해하는 방식에서 구별되는 고유한 패턴을 지니므로, 이를 유형화함으로써 진술 방식의 문제를 체계적으로 진단할 수 있다.

<표 4> 교과별 성취기준 진술 유형 분포(3인 합의 기준)

유형	국어 (n=29)	수학 (n=39)	영어 (n=33)	전체 (n=101)
A. 태도·참여형	5 (17.2%)	1 (2.6%)	3 (9.1%)	9 (8.9%)
B. 과정·전략 포괄형	5 (17.2%)	1 (2.6%)	13 (39.4%)	19 (18.8%)
C. 복합 행위 혼재형	13 (44.8%)	28 (71.8%)	4 (12.1%)	45 (44.6%)
D. 수행 명료형	6 (20.7%)	9 (23.1%)	13 (39.4%)	28 (27.7%)

첫 번째 유형은 행동 동사가 인지적 수행이 아니라 내면적 태도나 참여 행위를 기술하는 경우로, 이를 ‘태도·참여형’이라 명명한다. ‘참여한다’, ‘태도를 지닌다’, ‘생활화한다’, ‘인식한다’ 등의 동사가 이 유형의 전형적 표지이다. 대표 사례는 “[10공국1-05-01] 문학 소통의 특성을 고려하며 문학 소통에 참여한다”이다. 여기서 ‘참여한다’는 학생이 문학 소통에 능동적으로 관여하기를 기대하는 교육적 지향을 표현하지만, 어떤 수행으로 그 참여가 충분한 것인지, A 수준의 참여와 E 수준의 참여가 어떻게 다른지를 진술로부터 추론하기란 사실상 불가능하다. “[10공국2-05-02] 주체적인 관점에서 작품을 해석하고 평가하며 문학을 생활화하는 태도를 지닌다”는 ‘해석하고 평가하며’라는 인지적 행위를 품고 있어 얼핏 구체적으로 보이지만, 종결 동사 ‘태도를 지닌다’가 성취기준 전체를 태도 목표로 수렴시키며 평가의 준거가 될 수행을 모호하게 한다. 이 유형은 분석 기준 1과 3을 동시에 충족하지 못하며, 네 유형 중 평가 준거로서의 적합성이 가장 낮다.

두 번째 유형은 행동 동사는 비교적 명확하나 수행의 대상이나 조건이 지나치게 포괄적으로 설정되어 교사 간 해석의 편차를 구조적으로 허용하는 경우로, 이를 ‘과정·전략 포괄형’이라 명명한다. “[10공영1-02-07] 적절한 전략과 다양한 매체를 활용하여 상황과 목적에 맞게 말하거나 쓴다”가 그 전형이다. ‘말하거나 쓴다’는 수행의 방향을 지시하지만, 어떤 전략을, 어떤 매체로, 어떤 상황과 목적에서 동원해야 하는지가 모두 열려 있어 평가 과제를 특정하기 어렵다. 수학에서도 “[10공수2-01-02] 두 직선의 평행 조건과 수직 조건을 탐구하고 이해한다”처럼 ‘탐구하고 이해한다’로 끝나 산출물이 드러나지 않는 진술이 이 유형에 속한다.

세 번째 유형은 단일 성취기준 안에 인지적으로 이질적인 복수의 수행이 접속사로 연결되어 제시되는 경우로, 이를 ‘복합 행위 혼재형’이라 명명한다. “[10공국1-01-01] 대화의 원리를 고려하여 대화하고 자신의 듣기·말하기 과정과 공동체의 담화 관습을 성찰

한다”는 ‘대화하고’와 ‘성찰한다’를 한 기준에 묶는다. 전자는 언어적 수행이고 후자는 메타인지적 성찰로서, Bloom 분류체계상 상이한 인지 수준에 해당하며 각각 독립적인 평가 설계를 요구한다. 수학에서도 “[10공수1-01-02] 항등식의 성질과 나머지정리를 이해하고, 이를 활용하여 문제를 해결할 수 있다”가 ‘이해’와 ‘해결’을 결합하는데, 교사는 어느 쪽을 성취의 핵심 준거로 삼아야 하는지를 진술만으로는 판단하기 어렵다.

네 번째 유형은 행동 동사가 명확하고, 내용과 수행이 비교적 명확히 식별 가능하며, 진술로부터 평가 과제를 합리적으로 추론할 수 있는 경우로, 이를 ‘수행 명료형’이라 명명한다. “[10공수1-02-09] 미지수가 1개인 연립일차부등식을 풀 수 있다”는 수행의 대상과 행위가 단일하게 특정되어, 교사가 평가 과제를 설계할 때 개입해야 하는 해석의 여지가 최소화된다. 다만, 수행 명료형으로 분류된 성취기준도 기준 4(성취 조건·맥락의 명시 여부)에서는 여전히 부분 적합에 머무는 경우가 많다는 점은 해석에 유의할 필요가 있다.

이상의 분류에서 드러나는 핵심은 두 가지다. 첫째, 전체 101개 중 복합 행위 혼재형(C)이 45개(44.6%)로 가장 많다. 현행 성취기준의 지배적 진술 방식이 단일 수행으로 분해되지 않는 복합 진술임을 보여준다. 둘째, 유형 분포가 교과에 따라 뚜렷이 갈린다. 수학은 복합 행위 혼재형에, 국어는 복합 행위 혼재형과 태도·참여형에, 영어는 과정·전략 포괄형과 수행 명료형의 양분에 각각 편중된다. 어느 교과에서도 수행 명료형(D)이 과반을 점하지 못한다는 사실은, 평가 준거로 곧바로 활용 가능한 진술이 교육과정 전반에서 예외에 가깝다는 점을 시사한다.

2. 교과별 분석 결과

가. 국어

국어 교과의 성취기준은 여섯 개 영역(듣기·말하기, 읽기, 쓰기, 문법, 문학, 매체)에 걸쳐 공통국어1·2 합산 29개가 분포한다. 분석 결과, 국어 성취기준은 복합 행위 혼재형(C)이 13개로 가장 많고, 태도·참여형(A)과 과정·전략 포괄형(B)이 각 5개, 수행 명료형(D)이 6개로 나타나 네 유형이 비교적 고르게 분포되어 있다.

태도·참여형 성취기준이 집중된 곳은 문학과 문법, 그리고 읽기 영역이다. “[10공국1-05-01] 문학 소통의 특성을 고려하며 문학 소통에 참여한다”, “[10공국2-04-01] 과거 및 현재의 국어생활에 나타나는 국어의 변화를 이해하고 국어문화 발전에 참여한다”, “[10공국2-02-03] 의미 있는 사회적 독서 활동에 참여함으로써 …이해하는 태도를 지닌다”는 모두 종결 동사가 ‘참여한다’ 또는 ‘태도를 지닌다’로 수렴하여, 학생의 수

행이 어떤 산출물로 드러나야 하는지를 성취기준으로부터 추론하기 어렵다.

복합 행위 혼재형 성취기준은 듣기·말하기, 읽기, 문법 영역에 폭넓게 걸쳐 있다. “[10공국1-01-02] 논제의 필수 쟁점별로 논증을 구성하고 논증이 타당한지 평가하며 토론한다”는 ‘논증 구성’, ‘논증의 타당성 평가’, ‘토론’이라는 복수의 수행을 하나의 성취기준에 포함하고 있다. 논증을 구성하는 수행은 쟁점별 주장과 근거를 조직하는 능력을, 타당성을 평가하는 수행은 논증의 신뢰성·타당성·공정성을 판단하는 능력을, 토론은 이러한 논증을 바탕으로 상대방과 상호작용하는 수행을 요구한다. 이들 수행은 하나의 토론 과정에서 연계될 수 있으나, 평가 증거와 판단 준거의 측면에서는 구분되므로 성취기준만으로 평가의 핵심 단위를 특정하기 어렵다. “[10공국1-02-01] 다양한 글이나 자료를 읽으며 논증의 타당성을 평가하고 자신의 관점을 바탕으로 논증을 재구성한다”는 ‘논증의 타당성 평가’(비판적 읽기)와 ‘논증의 재구성’(창안에 가까운 수준)을 하나의 기준으로 묶어, 교사가 어느 수행을 평가의 핵심으로 삼아야 하는지를 판단하기 어렵게 한다. 한편, 문법 영역은 흥미로운 사례를 제공한다. “[10공국1-04-02] 음운 변동을 탐구하여 발음과 표기에 올바르게 적용한다”는 종결 동사 ‘적용한다’가 Bloom 분류체계의 적용 수준을 명확히 지시함에도, 선행하는 ‘탐구하여’가 별도의 인지적 수행을 끌어들이므로써 결국 복합 행위 혼재형으로 분류되었다. 동사의 명료성만으로는 평가 준거의 요건이 충족되지 않음을 보여주는 대목이다.

수행 명료형은 쓰기와 매체 영역에 주로 분포한다. “[10공국2-03-02] 논증 요소에 따른 분석을 바탕으로 효과적으로 내용을 조직하여 논증하는 글을 쓴다”, “[10공국1-06-01] 사회적 의제를 다룬 매체 자료를 비판적으로 분석한다”는 ‘쓴다’, ‘분석한다’라는 산출물이 분명한 동사를 종결로 삼아 평가 과제의 윤곽을 비교적 뚜렷하게 보여준다. 국어에서 평가 준거에 가장 근접한 진술이 문법이 아니라 쓰기·매체 영역에 있다는 점은, 산출물이 구체적인 영역일수록 진술이 명료해진다는 일반적 경향과도 부합한다.

결론적으로 국어 교과 성취기준의 전반적 특성은, 교과가 지향하는 총체적·과정 중심적 언어 능력이 진술 방식에 그대로 투영된 결과라는 점에서 이해할 만하다. 그러나 교과의 특성만으로 진술의 포괄성을 정당화하기는 어렵다. 이 점에서 총체적 언어 능력을 지향하면서도 평가 준거로 기능할 수 있는 진술 방식을 모색하는 것은 국어 교과가 성취평가제 전환 과정에서 마주한 핵심 과제라고 할 수 있다.

나. 수학

수학 교과의 성취기준은 공통수학1·2 합산 39개가 일곱 개 내용 영역에 분포한다. 분석 결과, 수학 성취기준은 복합 행위 혼재형(C)이 28개(71.8%)로 압도적이며, 수행 명

료형(D)은 9개(23.1%)에 그친다. 태도·참여형과 과정·전략 포괄형은 각 1개에 불과하다. 이 점에 비추어 볼 때, 수학 교과가 세 교과 가운데 평가 준거로서 상대적으로 양호하다는 통념은 부분적으로만 적용된다. 그 양호함은 수행 명료형의 비중이 아니라 수행 동사의 명료성에서 비롯되기 때문이다.

‘계산할 수 있다’, ‘구할 수 있다’, ‘증명할 수 있다’, ‘판별할 수 있다’, ‘풀 수 있다’처럼 산출물이 토렷한 동사가 광범위하게 사용되어 기준 1을 안정적으로 충족한다는 점에서 수학과 교육과정의 성취기준 진술이 지닌 강점은 분명하다. 그러나 이 동사들은 대개 선행하는 개념 이해와 한 문장 안에 묶인다. “[10공수1-01-02] 항등식의 성질과 나머지 지정리를 이해하고, 이를 활용하여 문제를 해결할 수 있다”, “[10공수2-03-01] 함수의 개념을 설명하고, 그 그래프를 이해한다”에서 보듯, ‘이해하고/설명하고 …할 수 있다’는 정형화된 결합 패턴이 반복된다. 그 결과 동사는 명료하되 기준 2(내용·수행의 식별)에서 대거 부분 적합에 머물러, 다수가 복합 행위 혼재형으로 분류되었다. 이는 수학의 복합 행위 혼재형이 동사 자체가 불명확한 국어의 태도·참여형과 성격을 달리하는 지점이다. 전자가 ‘무엇을 평가할지는 분명하나 둘 중 어느 것을 평가할지가 모호한’ 경우라면, 후자는 ‘무엇을 평가할지부터 불분명한’ 경우다.

수행 명료형은 주로 방정식과 부등식 영역에 집중된다. “[10공수1-02-09] 미지수가 1개인 연립일차부등식을 풀 수 있다”, “[10공수1-01-03] 다항식의 인수분해를 할 수 있다”처럼 단일 내용에 단일 수행이 결합된 기준은 평가 과제를 비교적 안정적으로 도출할 수 있다. 다만 ‘이해한다’, ‘탐구한다’와 같은 동사도 잔존한다. “[10공수2-03-01] 함수의 개념을 설명하고, 그 그래프를 이해한다”의 ‘이해한다’는 그래프와 관련하여 어떤 수행이 기대되는지를 모호하게 하며, “[10공수2-01-02] 두 직선의 평행 조건과 수직 조건을 탐구하고 이해한다”는 산출물이 드러나지 않아 과정·전략 포괄형으로 분류되었다.

다. 영어

영어 교과의 성취기준은 이해(reception)와 표현(production)의 두 영역으로 구성되며, 공통영어1·2 합산 33개가 분포한다. 분석 결과, 영어 성취기준은 과정·전략 포괄형(B)과 수행 명료형(D)이 각각 13개(39.4%)로 양분되고, 복합 행위 혼재형(C) 4개, 태도·참여형(A) 3개가 뒤를 잇는다.

영어 교과에 나타난 양분은 영어과 교육과정 성취기준 진술의 극단적 압축성에서 비롯된다. 구체적으로, 이해 영역의 단일동사 진술은 명징한 동사를 포함하고 있다는 점에서 수행 명료형으로 분류된다. “[10공영1-01-01] 말이나 글에 포함된 세부 정보를 파

악한다”, “[10공영1-01-02] 말이나 글의 주제나 요지를 파악한다”가 그 예다. 그러나 이 진술들은 어떤 유형과 난이도의 텍스트에서, 어떤 방식으로 파악해야 하는지에 관한 조건을 전혀 담지 않아, 기준 4에서는 부분 적합에 그친다. 이와 달리, 전략·표현 영역의 진술은 수행의 범위가 매우 광범위하게 열려 있어 과정·전략 포괄형으로 분류된다. “[10공영1-02-07] 적절한 전략과 다양한 매체를 활용하여 상황과 목적에 맞게 말하거나 쓴다”는 전략·매체·상황·목적이 모두 개방된 채 제시되어, 교사마다 전혀 다른 평가 과제를 설계할 수 있다. 분석자 간 일치도가 영어에서 가장 낮았던 것(III장, Fleiss $\kappa=.66$) 역시 관찰 가능한 단일 수행이 제시되었다는 점과 수행 조건 및 평가 증거의 구체성이 부족하다는 점을 중시한 판정이 엇갈린 해석으로 볼 수 있다. 이는 성취기준 진술의 ‘명료하지만 구체적이지 않은’ 진술과 ‘열려 있어 포괄적인’ 진술 사이의 경계가 모호한 데서 비롯된다.

공통영어1과 공통영어2의 일부 성취기준은 동일한 문장으로 제시되어 있어, 성취기준 본문만으로는 두 과목의 기대 수행 수준의 차이를 직접 식별하기 어렵다. 예컨대 “[10공영1-01-01] 말이나 글에 포함된 세부 정보를 파악한다”와 “[10공영2-01-01] 말이나 글에 포함된 세부 정보를 파악한다”는 진술이 동일하다. 이해 영역의 8개 기준 중 6개가 이처럼 두 과목에서 같은 문장을 공유하며, 표현 영역에서도 유사한 양상이 확인된다.

이러한 진술의 동일성은 두 과목의 교수·학습 내용이나 실제 수행 수준에 차이가 없음을 의미하지는 않으며, 영어 교과 특성 반영된 것으로 볼 수 있다. 영어 교과에서는 동일한 언어 기능도 주제의 친숙성, 텍스트의 길이와 복잡도, 실제 사용되는 어휘의 범위와 난도, 언어 형식 및 과제의 인지적 요구에 따라 서로 다른 수준의 수행으로 구체화될 수 있다. 또한 성취기준은 특정한 수업 활동이나 교수 방법을 지시하는 기준이 아니므로, 교과서와 교사는 동일한 성취기준을 바탕으로 과목의 특성과 학습자 수준에 적합한 수업 및 평가를 설계할 수 있다.

다만 본 연구가 분석한 성취기준 진술문을 기준으로 볼 때, 두 과목의 기대 수행 수준을 구분하는 정보가 일부 성취기준 본문에는 직접 제시되지 않고 내용 체계, 성취기준 해설, 언어 자료 및 교수·학습 과제를 통해 보완되도록 구성되어 있다. 동일한 성취기준을 사용하더라도 과목별 자료와 과제 조건을 달리하여 성취수준을 구체화할 수 있으므로, 분석의 쟁점은 서로 다른 성취수준의 설정 가능성보다 그 차이를 일관되게 설정하기 위한 정보가 교육과정 문서와 평가 지원 자료에 어떻게 제공되는가에 있다. 이에 영어과에서는 성취기준 본문과 내용 체계, 성취기준 해설 및 관련 자료의 연계를 명료하게 제시하여 교사가 과목별 평가 요구 수준을 일관되게 해석할 수 있도록 지원할 필요가 있다.

3. 교과 간 비교 및 성취기준-성취수준 연계의 구조적 문제

앞서 제시한 교과별 분석 내용을 종합하면, 성취기준 진술 방식의 문제는 교과 보편적으로 나타나되 그 양상과 강도에서 교과별로 뚜렷한 차이를 보인다. <표 5>는 세 교과의 분석 결과를 주요 차원에서 대조한 것이다.

<표 5> 국어·수학·영어 성취기준 진술 방식 비교

분석 차원	국어	수학	영어
지배적 유형	C 우세, A·B·D 분산	C 압도(71.8%), D 일부	B·D 양립 (각 39.4%)
주요 진술 구조	수행 및 태도 요소 복수 연결	개념 이해와 절차적 수행 결합	단일 수행 동사 중심 압축
과목 간 수준 구분	공통국어1·2 간 부분 차별화	공통수학1·2 간 내용 영역으로 구분	공통영어1·2 간 진술 거의 동일
가치·태도 혼입	빈번(17.2%)	드물(2.6%)	일부(9.1%)
동사 명료성(기준 1)	낮음~중간	높음	중간
평가 증거 도출 가능성(기준 3)	낮음	중간	낮음~중간

<표 5>에서 확인되는 교과 간 차이는 단순한 진술 방식이나 성격의 차이가 아니라, 각 교과가 성취기준을 어떤 목적으로 설계하는지에 대한 암묵적 전제의 차이를 반영한다. 수학은 성취기준을 학생이 수행해야 할 절차적·논리적 과제의 목록에 가깝게 간주하며, 이 전제 아래 수행 동사의 명료성은 자연스럽게 높아진다. 다만 그 과제가 개념 이해와 한 문장에 묶이면서 복합 행위 혼재형으로 귀결된다. 국어는 성취기준을 언어 능력의 총체적 성장을 지향하는 교육 목표에 가깝게 설계하는 경향이 있고, 영어는 의사소통 기능의 목록을 압축적으로 제시함으로써 진술의 포괄성이 불가피하게 높아진다.

이번 분석에서 확인된 중요한 사실 가운데 하나는, 연구자 간 판정이 갈린 지점이 무작위가 아니라 특정 진술 패턴에 집중된다는 점이다. 불일치는 영어 이해 영역의 압축적 단일동사 진술(수행 명료형으로 볼 것인가, 과정·전략 포괄형으로 볼 것인가)과 수학·국어의 복합동사 결합형(‘이해하고 …할 수 있다’를 어디에 둘 것인가)에 몰렸으며, 세 연구자가 완전히 엇갈린 유일한 항목 또한 영어의 전략 적용 기준이었다. 이는 진술이 평가 준거로서 모호할수록 전문가 사이의 판정 근거 차이도 커질 수 있음을 시사한다. 다만 표장에서 밝힌 바와 같이 일치도 수치만으로 진술 모호성을 직접 입증하기는 어려우며, 불일치 항목에 대한 재검토 결과 영어 교과에서는 관찰 가능한 단일 수행의 제시 여

부와 수행 조건·평가 증거의 구체성에 대한 판단 기준의 차이가 유형 판정을 가르는 주된 요인으로 확인되었다.

교과 간 비교에서 도출된 또 하나의 중요한 발견은, 진술 방식의 문제가 개별 성취기준 수준에 머무르지 않고 성취기준과 성취수준 사이의 연계 구조 전반으로 확장된다는 점이다. 성취기준에서 평가의 핵심이 되는 수행이 명료하게 제시되어야 그 수행의 질적 차이를 A에서 E에 걸쳐 일관되게 기술할 수 있다. 현행 성취수준은 ‘매우 적절하게’, ‘대체로’, ‘부분적으로’ 등 형용사나 부사를 활용하여 수준 간 차이를 구분하는 경우가 많다. 이러한 정도 표현은 평가 대상이 되는 수행이 명확할 때에는 수행의 질적 차이를 간결하게 나타낼 수 있으나, 하나의 성취기준에 복수의 수행이 결합되어 있거나 평가 증거가 불분명한 경우에는 무엇의 수준 차이를 의미하는지 모호해질 수 있다. 따라서 형용사나 부사를 활용한 수준 구분이 평가자의 종합적 인상에 의존하지 않도록, 성취기준에는 해당 표현이 수식하는 핵심 수행과 평가 증거가 식별 가능하게 제시될 필요가 있다. 그러나 태도·참여형의 경우 ‘참여’라는 행위 자체에 질적 위계를 부여할 근거가 성취기준 진술에 담겨 있지 않으므로, 성취수준 기술은 성취기준과 독립적인 별도의 판단 기준에 의존할 수밖에 없다. 과정·전략 포괄형에서도 마찬가지다. “세부 정보를 파악한다”에서 A 수준과 E 수준을 가르는 기준이 텍스트의 복잡도인지, 파악의 정확도인지, 아니면 둘 다인지를 진술로부터 확정할 수 없다. 복합 행위 혼재형에서는 문제가 한층 복잡해진다. 결합된 두 수행 가운데 어느 것을 위계화의 축으로 삼느냐에 따라 성취수준 기술이 달라지기 때문이다.

이 문제는 윤지영, 온정덕(2025)이 지적한, 2022 개정 교육과정의 단위학교용 성취수준이 성취기준의 내용 요소를 수준 구분의 기준으로 활용함으로써 인지적 엄격성을 반영하지 못하는 현상과 직결된다. 성취기준의 수행 요소가 불명확하면, 성취수준 기술은 수행의 질적 차이를 위계화하는 대신 내용 요소의 포함 범위나 양으로 수준을 가르는 손쉬운 대안에 의존하게 된다. 결국, 현행 성취기준 진술 방식의 문제는 개별 기준의 언어적 불완전성에 그치지 않는다. 그것은 성취기준-성취수준-평가 설계로 이어지는 정렬의 연쇄를 출발점에서부터 흐트리는 구조적 결함이다. 이 결함이 진술 차원의 문제인 이상, 그 해결 또한 성취기준 개발 단계로 거슬러 올라가야 하며 사후의 어떤 보완과 지원으로도 대체되기 어렵다.

V. 대안적 성취기준 진술 방식의 제안

1. 진술 원리

IV장의 분석 결과는 현행 성취기준 진술 방식의 문제가 개별 기준의 언어적 부주의에서 비롯된 것이 아니라, 성취기준 개발 과정에서 평가 준거 기능에 대한 체계적 고려가 결여된 데서 구조적으로 발생함을 보여주었다. 더욱이 연구자 간 판정이 갈린 지점이 특정 진술 패턴에 집중되고 교과별 일치도가 진술 명료성의 순위와 일치했다는 사실은, 진술의 모호성이 분석자에게뿐 아니라 평가 현장의 교사에게도 동일한 해석 편차를 유발할 것임을 시사한다. 따라서 대안적 진술 방식의 모색은 개별 기준의 수정에 앞서, 성취기준이 평가 준거로 기능하기 위해 갖추어야 할 진술의 원리를 먼저 확립하는 데서 출발해야 한다. 본 연구는 III장의 분석 틀과 IV장의 유형 분석 결과를 바탕으로 다음 다섯 가지 원리를 제안한다. 이 원리들은 IV장에서 확인된 진술 유형의 문제와 과목 간 기대 수행 수준의 구분 문제를 종합하여 구성한 것이다.

가장 먼저 요구되는 것은 ‘핵심 수행의 명료성 원칙’이다. 다시 말해, 하나의 성취기준은 하나의 핵심 수행을 분명히 드러내도록 진술되어야 한다. 이때 핵심 수행을 구체화하는 내용, 조건, 방법, 근거 및 산출 형식은 보조 요소로 함께 제시될 수 있다. 다만 둘 이상의 수행이 각각 독립적인 평가 대상이 되는 경우에는 이를 하나의 성취기준 안에 병렬적으로 나열하지 않아야 한다. IV장에서 전체의 44.6%를 차지하며 가장 지배적이었던 복합 행위 혼재형(유형 C)은, 인지적으로 이질적인 복수의 행위를 한 문장에 묶음으로써 평가의 단위를 불명확하게 만들고 성취수준 기술의 논리적 분화를 저해한다. 특히 수학에서 반복적으로 나타난 ‘이해하고 ...할 수 있다’는 결합 패턴은 그 전형적인 사례이다. 또한, 수행 동사는 개정 분류체계상의 인지 수준을 명확하게 지시해야 하며, ‘이해한다’, ‘탐구한다’, ‘참여한다’와 같이 그 인지 수준을 구체적으로 식별하기 어려운 동사는 가급적 지양하거나, 불가피하게 사용할 경우 수행의 산출물을 함께 명시함으로써 보완해야 한다.

수행 동사가 단일해지더라도 평가 대상이 특정되기 위해서는, ‘무엇을(내용)’과 ‘어떻게 할 수 있는가(수행)’가 구조적으로 식별 가능해야 한다. 이는 ‘내용·수행 이원 구조의 원칙’이라고 할 수 있다. 내용과 수행 간 이원 구조가 명확할 때 교사는 무엇을 가르쳐야 하는지와 그 결과를 어떻게 평가해야 하는지를 성취기준으로부터 동시에 읽어낼 수 있다. IV장에서 태도·참여형(유형 A)이 평가 준거로서 가장 부적합했던 까닭은, 종결 동사가 ‘참여한다’·‘태도를 지닌다’로 수렴하면서 내용과 수행이 모두 태도 목표로 흡수되

어 버린 데 있다. 따라서 태도나 가치는 성취기준의 종결 동사로 사용하는 것을 지양하고, 수업의 지향이나 맥락을 안내하는 방식으로 성취기준 해설이나 교수·학습 유의사항에서 다루는 것이 바람직하다.

이렇게 식별된 수행이 실제 평가로 이어지려면, 진술로부터 평가의 단서가 드러나야 한다. ‘평가 증거 가시성의 원칙’은 성취기준이 그 진술만으로 어떤 유형의 평가 과제와 산출물이 적합한지를 자연스럽게 시사하도록 작성될 것을 요구한다. IV장의 과정·전략 포괄형(유형 B)은 동사는 명료하나 수행의 대상·조건·범위가 모두 열려 있어, ‘명료하지만 속이 비어 있는’ 진술로 귀결되었다. 이는 Wiggins & McTighe(2005)의 백워드 설계 원리를 성취기준 진술에 적용한 것으로, 성취기준을 작성하는 단계에서 이미 ‘이 기준에 도달했음을 어떤 수행으로 입증할 수 있는가’를 사유하는 것을 의미한다.

나아가 성취평가제가 등급 판정 체제인 이상, 성취기준은 수준 간 차이를 담아낼 수 있어야 한다. ‘성취수준 분화 가능성의 원칙’은 성취기준에서 평가의 핵심이 되는 수행과 평가 증거를 식별할 수 있도록 함으로써, 그 수행의 질적 차이를 A에서 E에 걸쳐 일관되게 기술할 수 있는 구조를 갖출 것을 요구한다. IV장에서 확인되었듯, 수행 요소가 불명확하면 성취수준 기술은 수행의 질적 차이보다 내용 요소의 양이나 ‘매우’, ‘대체로’, ‘부분적으로’와 같은 정도 표현에 의존할 가능성이 커진다. 따라서 작성된 성취기준 진술을 놓고 “이 진술로부터 A 수준과 E 수준의 수행을 논리적으로 구분할 수 있는가”를 검토하는 것을 성취기준 개발의 필수 절차로 제도화할 것을 제안한다.

끝으로, 연계하여 편성되는 과목 사이에서는 기대 수행 수준의 차이를 식별할 수 있어야 한다. 이것이 본 연구가 IV장에서 확인한 문제에 대응하는 ‘과목 간 수준 명료화’이다. 이러한 문제는 공통영어1과 공통영어2에서 주로 확인되었으며, 두 과목의 일부 성취기준이 동일한 문장으로 제시된 것은 영어과의 교과적 특성을 고려하더라도, 과목별 기대 수행 수준을 구분하는 정보가 보다 명료하게 제공될 필요가 있음을 보여준다. 성취기준 본문에는 핵심 수행을 제시하고, 성취기준 해설에서 텍스트의 복잡도, 어휘와 언어 형식, 수행 조건 및 산출물의 범위를 보다 구체적으로 제시하는 방식으로 과목별 기대 수행 수준의 차이를 보완할 수 있다. 따라서 성취기준 본문과 해설을 함께 참조했을 때 교사가 연계 과목 사이의 기대 수행 수준을 일관되게 식별할 수 있도록 관련 정보가 명료하게 작성되어야 한다.

이상 다섯 가지 원칙은 성취기준이 교수·학습의 안내자로서의 기능보다 평가에 적합한 도구나 기준으로만 기능해야 한다는 주장이 아니다. 오히려 진술이 명료할수록 교사는 수업에서 무엇을 중점적으로 다루어야 하는지를 더 분명히 인식하게 되고, 평가는 수업에서 다룬 것을 자연스럽게 확인하는 방식으로 설계될 수 있다.

2. 교과별 개선 진술 예시

위의 다섯 원칙을 적용하여 본 연구는 현행 성취기준을 보완·개선할 수 있는 대안적 진술 방식을 제안한다. 아래 <표 6>은 각 교과별 개선 사례를 현행 진술, 개선 진술(안), 주요 문제 유형과 함께 정리한 것이다.

<표 6> 교과별 성취기준 개선 진술 예시

교과	코드	현행 진술	개선 진술(안)	주요 문제 유형
국어	[10공국1-05-01]	문학 소통의 특성을 고려하며 문학 소통에 참여한다.	작가 맥락, 독자 맥락, 사회·문화적 맥락 등 문학 소통의 다층적 특성을 고려하여 작품에 대한 해석을 그 근거와 함께 글이나 말로 설명한다.	태도·참여형 (유형 A)
국어	[10공국1-03-01]	내용 전개에 대한 일반적 원리를 고려하여 사회적 쟁점에 대한 자신의 견해를 정교하게 표현하는 글을 쓴다.	사회적 쟁점에 대한 자신의 견해를 논증 요소(주장, 근거, 반론 고려)와 예상 독자를 고려한 어휘와 문체로 표현하는 글을 쓴다.	복합 행위 혼재형 (유형 C)
국어	[10공국1-02-01]	다양한 글이나 자료를 읽으며 논증의 타당성을 평가하고 자신의 관점을 바탕으로 논증을 재구성한다.	다양한 분야의 글이나 자료에서 논증의 주장, 근거, 논증 방법을 분석하여, 논리적 오류나 근거의 부적절성을 들어 타당성을 평가한다.	복합 행위 혼재형 (유형 C)
영어	[10공영1-01-02]	말이나 글의 주제나 요지를 파악한다.	친숙한 일반적 주제를 다룬 다양한 유형의 말이나 글에서 핵심어와 반복되는 정보를 근거로 주제나 요지를 한두 문장으로 요약하여 진술한다.	과정·전략 포괄형 (유형 B)
영어	[10공영2-02-09]	다른 사람과 의견을 조율하며 문제 해결을 위해 협력한다.	상대방의 의견을 경청하고 자신의 의견 사이의 공통점 및 차이점을 바탕으로 합의점을 도출하기 위한 제안이나 타협의 표현을 상황에 맞게 사용한다.	태도·참여형 (유형 A)
수학	[10공수2-03-01]	함수의 개념을 설명하고, 그 그래프를 이해한다.	두 집합 사이의 대응 관계로서 함수의 개념을 설명하고, 주어진 함수의 그래프에서 정의역·치역·공역의 관계를 파악한다.	'이해한다'류 동사 잔존 (유형 C)
수학	[10공수1-01-02]	항등식의 성질과 나머지정리를 이해하고, 이를 활용하여 문제를 해결할 수 있다.	항등식의 성질과 나머지정리를 이용하여 다항식의 계수를 구하거나 나머지를 계산하는 문제를 해결한다.	복합 행위 혼재형 (유형 C)

성취기준의 구체화는 평가의 핵심이 되는 수행과 평가 증거를 분명히 함으로써 성취 수준 간 차이를 구분하고 평가자 간 판단의 일관성을 높이는 데 기여할 수 있다. 특히

수행 대상과 필요한 조건이 명료하게 제시되면, 교사는 이를 토대로 평가 과제와 채점 기준을 설계하고 A에서 E에 이르는 성취수준을 보다 구체적으로 진술할 수 있다. 그러나 성취기준이 특정 자료, 활동 절차 또는 산출물의 형식까지 지나치게 규정할 경우, 교과서와 교수·학습 활동이 유사한 형태로 수렴하여 교육과정 구현의 다양성이 제한될 수 있다. 따라서 성취기준의 구체화는 특정한 교수·학습 활동을 지시하는 수준으로 확대되기보다, 평가의 핵심 수행과 수행 대상 및 필요한 조건을 식별할 수 있는 범위에서 이루어질 필요가 있다. 구체적인 자료와 활동 및 산출물의 예시는 성취기준 해설이나 평가 지원 자료를 통해 보완할 수 있으며, 이를 통해 평가 준거의 명료성과 교과서 및 수업 설계의 다양성을 함께 확보할 수 있다.

국어 문학 영역의 [10공국1-05-01]은 태도·참여형의 전형적 사례다. 현행 기준의 종결 동사 ‘참여한다’는 어떤 인지적 수행이 기대되는지를 전혀 특정하지 못한다. 개선 진술에서 ‘설명한다’가 종결 수행 동사로서 평가의 핵심을 지시하고, ‘문학 소통의 다층적 특성을 고려하여’가 수행의 조건을 명시함으로써, A 수준의 학생은 복수의 맥락을 정교하게 통합하여 타당한 해석을 구성하고 이를 체계적으로 설명할 수 있는 반면, E 수준의 학생은 단일 맥락에 의존한 단순한 해석을 부분적으로 설명하는 수준에 머문다는 방식으로 성취수준이 논리적으로 분화될 수 있다.

영어 이해 영역의 [10공영1-01-02]는 과정·전략 포괄형의 대표 사례다. 현행 기준은 어떤 조건과 방법으로 파악해야 하는지에 관한 정보를 전혀 담지 않는다. 개선 진술에서 텍스트의 조건, 수행의 방법, 수행의 산출물이 모두 포함됨으로써 교사가 평가 과제를 설계하기 위한 실질적 근거가 마련된다. 나아가 공통영어2의 동일 기준은 텍스트의 복잡도를 높이거나 수행의 산출물을 확장하는 방식으로 진술함으로써, 성취기준 본문만으로는 직접 확인하기 어려운 두 과목의 기대 수행 수준 차이를 보다 명료하게 제시할 수 있다. 이는 앞서 제시한 과목 간 수준 명료화의 원칙을 영어 이해 영역에 구체화한 것이다.

수학의 [10공수1-01-02]는 “...를 이해하고, 이를 활용하여 문제를 해결할 수 있다”의 정형화된 패턴을 보여준다. 개선 진술에서 ‘이해’를 수행의 전제로 남겨 두는 대신 ‘이용하여 해결한다’는 단일한 수행 구조로 통합함으로써 평가의 핵심 준거가 명확해진다. 이 진술로부터 A 수준은 복잡한 조건의 문제를 다양한 전략으로 해결하고, E 수준은 단순한 형태의 문제를 주어진 절차에 따라 부분적으로 해결하는 방식으로 성취수준이 논리적으로 분화될 수 있다.

이상의 성취기준 개선 진술 예시들은 공통적으로 두 가지 방향을 지향한다. 하나는 수행의 산출물을 진술 안에 가시화하는 것이고, 다른 하나는 수행의 조건이나 맥락을 최소한으로라도 명시하는 것이다. 이 두 방향이 실현될 때 성취기준은 교사에게 무엇을 가르치고 어떻게 평가해야 하는지를 동시에 안내하는 실질적 준거로 기능할 수 있다.

VI. 결론 및 제언

이 연구는 고교학점제의 전면 시행과 함께 성취평가제의 실질적 구현이 요청되는 시점에서, 그 제도적 기반이 되는 성취기준의 진술 방식에 주목하였다. 성취평가제가 교육과정의 성취기준을 실질적 평가 준거로 삼는 준거 참조 평가 체제임을 전제할 때, 성취기준의 진술 방식은 제도의 논리적 기반이자 실행의 출발점이 된다. 이러한 문제의식에 따라 이 연구는 2022 개정 교육과정 공통국어1·2, 공통영어1·2, 공통수학1·2의 성취기준 101개를 대상으로, 행동 동사의 인지 수준 명료성, 내용·수행의 식별 가능성, 평가 증거의 도출 가능성, 성취 조건·맥락의 명시 여부라는 네 가지 기준에 따라 분석을 수행하고, 그 결과를 바탕으로 대안적 진술 방식의 원리와 예시를 제안하였다.

분석 결과를 종합하면 다음과 같다. 첫째, 현행 성취기준의 진술 방식은 태도·참여형, 과정·전략 포괄형, 복합 행위 혼재형, 수행 명료형의 네 가지 유형으로 분류되며, 이 중 평가 준거로서의 적합성이 높은 수행 명료형은 영어 교과가 가장 높은 비율로 나타났으나, 과정·전략 포괄형도 동일한 비율로 나타나 수행 명료형이 우세하다고 보기는 어렵다. 수학 교과에서는 복합 행위 혼재형이 압도적으로 큰 비중을 차지하였으며, 국어 교과에서는 네 유형이 광범위하게 분포하고 있다. 어느 교과에서도 수행 명료형이 과반을 차지하지 않았다는 점은 현행 성취기준 중 평가 준거로 직접 활용할 수 있는 진술이 충분히 확보되지 않았음을 보여준다. 둘째, 교과별 분석에서 국어는 복합 행위 혼재형이 가장 높은 가운데, 태도·참여형, 과정·전략 포괄형 및 수행 명료형이 혼재하여 교과가 지향하는 총체적 언어 능력이 평가 준거의 명료성을 저해하는 방식으로 진술에 투영되어 있음이 확인되었다. 영어는 과정·전략 포괄형과 수행 명료형이 동일한 비율로 나타났으며, 공통영어1과 공통영어2의 일부 성취기준에서는 진술 본문만으로 과목별 기대 수행 수준의 차이를 직접 식별하기 어렵다는 문제가 추가로 확인되었다. 수학은 개념 이해와 절차적 수행을 하나의 성취기준에 결합하는 진술 방식으로, ‘이해한다’와 같은 동사의 간존과 정형화된 복합 행위 패턴이 일관성을 저해한다. 셋째, 세 교과에서 공통적으로 성취기준의 진술 방식이 성취수준 기술과 평가 증거를 논리적으로 연계되기 어려운 구조적 제약이 확인되었다. 나아가 연구자 간 판정이 갈린 지점이 특정 진술 패턴에 집중되고 교과별 일치도가 진술 명료성의 순위와 일치했다는 사실은, 진술의 모호성이 분석자뿐 아니라 평가 현장의 교사에게도 동일한 해석 편차를 일으킬 것임을 방증한다.

이러한 분석 결과는 다음 세 가지 차원의 정책적 제언으로 이어진다. 첫째, 차기 교육과정 개정 과정에서 성취기준 개발 절차에 평가 가능성 검토 단계를 공식적으로 포함할 필요가 있다. 현행 교육과정 개발 과정에서는 교과별 전문성을 고려하여 구성된 정책연

구진이 공통의 각론 개발 지침을 토대로 교과 교육과정과 성취기준 시안을 개발한다. 그러나 2022 개정 교육과정의 각론 개발 지침은 성취기준을 수행기준으로 개발하기 위한 방향과 구체적인 안내가 충분하지 않다는 지적도 제기된다(한혜정, 2023). 이에 교과별 내용과 수행의 특성을 반영하면서도, 개별 성취기준으로부터 평가 증거와 성취수준을 일관되게 도출할 수 있는지를 검토하는 절차가 필요하다. 본 연구가 III장에서 제시한 네 가지 분석 기준은 이러한 검토 절차의 구체적 준거로 활용될 수 있다. 특히 ‘이 성취기준으로부터 성취수준 A와 E를 논리적으로 구분할 수 있는가’라는 질문을 성취기준 개발의 최종 점검 기준으로 제도화하는 것이 필요하다. 이 점검은 교과 내용 전문가만이 아니라 교육과정 평가 전문가와 현장 교사가 함께 참여하는 방식으로 이루어질 때 실효성이 높아질 수 있다.

둘째, 교육과정 성취기준과 평가준거 성취기준의 이원적 체제에 대한 재검토가 요구된다. II장에서 살펴본 두 역할의 긴장, 즉 국가 교육과정의 성취기준이 평가 준거로 직접 기능하기 어렵다는 인식하에 평가준거 성취기준을 별도로 개발·보급하는 방식은, 성취기준의 두 역할 사이의 긴장을 제도적으로 해소하지 않고 우회하는 것에 해당한다. 이 이원적 체제는 교육과정 성취기준의 진술 방식 개선에 대한 압력을 약화시키는 구조적 유인으로 작용한다. 성취기준이 교수·학습의 안내자이면서 동시에 실질적 평가 준거로 기능하도록 진술 방식 자체를 개선하는 것이 장기적으로 더 근본적인 해결책이며, 평가준거 성취기준은 그 과도기적 보완 장치로서의 성격을 명확히 하면서 단계적으로 그 필요성을 줄여나가는 방향을 지향해야 한다.

셋째, 차기 국가 교육과정 개발 시 총론 차원에서 성취평가를 위한 성취기준 진술의 공통 틀을 제시할 필요가 있다. 현행 교육과정 개발 체제에서 각론의 성취기준은 교과별로 독립적으로 개발되며, 총론은 성취기준의 기능과 역할에 대한 원론적 정의를 제시하는 데 그칠 뿐 진술 방식에 관한 구체적 지침을 제공하지 않는다. 그 결과 교과에 따라 성취기준의 진술 방식이 현저히 상이하고, 평가 준거 기능의 충족 여부도 교과별로 편차가 크다. 이러한 문제를 구조적으로 해소하기 위해서는 총론이 각론 편성의 원리를 제시하는 단계에서, 본 연구가 V장에서 제안한 성취기준 진술의 공통 원리, 즉 핵심 수행의 명료성, 내용·수행 이원 구조, 평가 증거 가시성, 성취수준 분화 가능성, 과목 간 수준 명료화를 공통 틀로 명시적으로 제안해야 한다. 다만 이러한 공통 틀이 교과별 성취기준 진술을 획일화하는 방향으로 작용해서는 안 된다. 국어·영어처럼 총체적 언어 능력을 지향하는 교과와 수학처럼 절차적·논리적 수행이 중심이 되는 교과는 성취기준이 담아야 하는 수행의 성격 자체가 다르며, 이 차이는 진술 방식에도 반영되어야 한다. 따라서 총론의 공통 지침과 각론의 교과별 구체화라는 역할 구분은 유지하되, 현행 개발 체제에 성취기준의 평가 준거 적합성을 검증하는 절차를 명시적으로 보완할 필요가 있다. 구체

적으로 총론 차원에서는 행동 동사의 인지 수준 명료성, 내용·수행의 식별 가능성, 평가 증거의 도출 가능성, 성취 조건·맥락의 명시 여부를 공통 점검 기준으로 제시해야 한다. 각론 개발 과정에서는 이러한 기준을 교과별 지식 구조와 수행 특성에 맞게 적용하고, 작성된 성취기준으로부터 성취수준과 평가 과제를 시범적으로 구성하여 진술의 적합성을 검토할 필요가 있다. 또한 검토 과정에서 성취수준의 구분이나 평가 증거의 도출이 어려운 것으로 확인된 성취기준은 그 결과를 진술 수정에 다시 반영해야 한다. 이러한 과정을 통해 성취기준의 교과 간 일관성과 교과 내 적합성이 균형 있게 확보될 수 있다.

본 연구는 다음과 같은 한계를 지닌다. 우선 분석 대상을 고등학교 주요 세 교과의 공통과목으로 한정하였으므로, 분석 결과를 고등학교 전체 교과 또는 다른 학교급으로 일반화하는 데에는 신중함이 요구된다. 또한, 분석에 참여한 연구자들이 모두 교육과정 전 공자이자 연구 문제와 분석 관점을 공유한 공동 연구자였다는 점에서, 공통된 이론적 전제가 개별 판정에 영향을 미쳤을 가능성을 완전히 배제하기 어렵다. 본 연구는 독립 판정과 연구자 간 일치도 산출을 통해 이러한 가능성을 줄이고자 하였으나, 후속 연구에서는 교과교육 및 교육평가 분야의 외부 전문가 검토를 통해 분석 기준의 내용타당성과 유형 판정의 교과 적합성을 추가로 확인할 필요가 있다. 아울러 본 연구가 제안한 개선 진술 예시들은 실제 교육과정 개발 현장에서의 타당성 검토와 현장 교사 및 교과 전문가의 검증을 거치지 않은 한계가 있다. 또한 복합 행위 혼재형으로 분류된 성취기준 중 일부는 2022 개정 교육과정 총론이 요구하는 내용 체계상 복수 학습 요소의 결합 방식에서 비롯되었을 가능성이 있다. 본 연구는 이러한 총론 차원의 진술 방식에 따른 복수 범주의 결합과, 평가 준거로서 결합에 해당하는 이질적 행위의 혼재를 구분하지 않고 분석하였으므로, 이 두 성격을 구분하여 재검토하는 것은 후속 연구의 과제로 남는다. 이와 같은 한계를 보완하기 위한 후속 연구의 방향은 다음과 같다. 첫째, 분석 대상을 고등학교 선택 교과 및 다른 학교급으로 확장하여 성취기준 진술 방식의 문제가 교육과정 전반에 걸쳐 어떻게 분포하는지를 체계적으로 파악하는 연구가 필요하다. 둘째, 본 연구가 제안한 분석 틀과 개선 원리를 실제 교육과정 개발 과정에 적용하여 그 실행 가능성과 효과를 검증하는 실행 연구가 요청된다. 셋째, 성취기준 진술 방식의 개선이 실제 교사의 평가 설계와 성취수준 기술에 어떤 영향을 미치는지를 추적하는 종단적 연구가 이루어진다면, 본 연구의 이론적 제안이 실천적 근거를 확보하는 데 기여할 수 있을 것이다.

성취기준은 교육과정과 수업, 평가를 잇는 핵심 고리다. 이 고리가 튼튼하지 않으면 아무리 정교한 평가 제도도 그 논리적 기반을 확보하기 어렵다. 이 연구가 일관되게 확인한 것은, 현행 2022 개정 교육과정의 성취기준 진술 방식이 평가 준거로서 갖추어야 할 요건을 충분히 충족하지 못하며, 이러한 상황이 성취기준에서 성취수준으로, 다시 평가 설계로 이어지는 정렬의 연쇄를 그 출발점에서부터 흔든다는 사실이다. 성취평가제

의 실질적 구현은 평가체제의 개편만으로 이루어질 수 없으며, 그 제도의 기반이 되는 성취기준 자체의 질적 개선을 동반해야 한다. 성취평가제가 ‘경쟁에서 성장으로의 전환’이라는 본래의 취지를 실현하려면, 그 전환은 학생의 도달 여부를 판단하는 평가의 타당성을 뒷받침하는 성취기준의 진술에서부터 시작되어야 한다.

참고문헌

- 고영빈, 지은림(2022). 고교학점제와 성취평가 도입에 따른 진로선택과목 운영 및 이수에 대한 고교생의 인식. **학습자중심교과교육연구**, 22(22), 1085-1102.
- 김선영(2016). 역량중심 교육과정 구현을 위한 교과 교육과정 성취기준의 진술방식 탐색. **교육연구**, 66, 113-140.
- 김순남, 이병환(2018). 고등학교 성취평가제의諸문제와 정착 방안. **열린교육연구**, 26(3), 109-129.
- 김영은, 김경희, 김인숙, 김현미, 노은희, 허예지, 김병연, 김은하, 김현종, 노혜정, 박현진, 윤재룡(2024). 2022 개정 교육과정에 따른 고등학교 성취수준 개발 연구(총론). 한국교육과정평가원 연구보고 CRC 2024-2.
- 백남진(2014). 교과 교육과정 성취기준 진술의 개선 방향 탐색: 한국과 미국 과학 교육과정 검토를 중심으로. **교육과정연구**, 32(2), 101-131.
- 서영진(2013). 국어과 교육과정 '내용 성취 기준'의 진술 방식에 대한 비판적 고찰. **국어교육학연구**, 46, 415-450.
- 윤지영, 온정덕(2025). 2022 개정 교육과정에 따른 성취평가에 대한 비판적 고찰: 성취수준 설정을 중심으로. **교육과정연구**, 43(1), 81-108.
- 이미경, 박순경, 권점례, 백경선, 이근호, 김성혜, 김희경, 한혜정, 이주연, 김현정, 김현수, 박은아, 박주현, 변희현, 김현경, 김현정, 박영수, 서지영, 이경언, 최성희, 장경숙, 이용백, 김영춘, 김종윤(2017). 2015 개정 교육과정에 따른 고등학교 교과별 평가기준 개발 연구(총론). 한국교육과정평가원 연구보고 CRC 2017-5-1.
- 이승미, 박순경(2014). 교과 교육과정의 성취기준 개발 실태와 개선 방안 탐색. **교육학연구**, 52(2), 53-79.
- 조인영, 강완(2018). 초등학교 수학과 교육과정 성취기준 진술 방식 분석. **한국초등수학교육학회지**, 22(1), 69-92.
- 한수현, 김영실(2025). 사용자 관점에 근거한 성취기준의 구조적 특성 탐색. **교육연구**, 92, 335-357.
- 한혜정(2023). 총론과 각론의 연계 관점에서 2015 개정과 2022 개정 각론 개발 지침 비교·분석 및 시사점 고찰. **학습자중심교과교육연구**, 23(5), 367-385.

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning*,

- teaching, and assessing: A revision of Bloom's taxonomy of educational objectives* (Complete ed.). New York, NY: Longman.
- California Department of Education. (2013). *California Common Core State Standards: Mathematics*. Sacramento, CA: California Department of Education.
- California Department of Education. (2023). *Mathematics framework for California public schools: Kindergarten through grade twelve*. Sacramento, CA: California Department of Education.
- International Baccalaureate Organization. (2020). *Middle Years Programme subject brief: Language acquisition*.
- International Baccalaureate Organization. (2026a). *Assessment and exams*. <https://www.ibo.org/programmes/middle-years-programme/assessment-and-exams/> (검색일: 2026. 6. 13.)
- International Baccalaureate Organization. (2026b). *Language acquisition*. <https://www.ibo.org/programmes/middle-years-programme/curriculum/language-acquisition/> (검색일: 2026. 6. 13.)
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.
- Mager, R. F. (1997). *Preparing instructional objectives: A critical tool in the development of effective instruction* (3rd ed.). Center for Effective Performance.
- Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education* (Research Monograph No. 6). Madison, WI: National Institute for Science Education, University of Wisconsin-Madison.
- Wiggins, G., & McTighe, J. (2005). *Understanding by design* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

논문접수 : 2026.7.2. / 수정본접수 : 2026.7.30. / 게재승인 : 2026.8.11.

ABSTRACT

Rethinking the statement of achievement standards for implementing achievement-based evaluation in high schools: Focusing on the Korean, Mathematics, and English subjects of the 2022 Revised Curriculum

Minsun Sung

Ph.D. Candidate, Konkuk University

Misun Kong

Ph.D. Candidate, Konkuk University

Jonghun Kim

Associate Professor, Konkuk University

As the high school credit system takes full effect in 2025, the success of criterion-referenced achievement-based evaluation hinges less on score cutoffs than on how clearly achievement standards are stated. This study analyzed the achievement standards of three common subjects in the 2022 Revised National Curriculum—Common Korean 1·2, Common Mathematics 1·2, and Common English 1·2—against four criteria: the clarity of the action verb’s cognitive level, the separability of content and performance, the derivability of assessment evidence, and the explicitness of performance conditions and contexts. The statements were classified into four types, with the mixed-action type accounting for the largest share overall. The analysis of English further showed that the achievement-standard statements alone did not clearly distinguish the expected performance levels between Common English 1 and Common English 2. These deficiencies extend beyond individual statements to undermine the alignment among achievement standards, achievement level descriptors, and assessment design from the outset. The study therefore proposes five principles for stating achievement standards, along with subject-specific examples, and recommends institutionalizing a review of assessability within the standard-development procedure in the next curriculum revision.

Key Words: high school credit system, achievement-based evaluation, achievement standards, achievement level descriptors, 2022 Revised National Curriculum