University of Southampton

미래형 학생평가 역량 강화 배움자리

교사의 평가 판단은 어떻게 공정성과 신뢰를 확보할까?

잉글랜드의 평가 제도 사례

박원용 · 사우스햄튼 대학교 부교수(과학교육 전공)

2026년 11월 24일(화) 15:00–16:30 · 충청남도교육청교육과정평가정보원 · Zoom

목차

강의의 흐름

1

평가의 맥락

한국의 쟁점과 공정성
잉글랜드의 평가 체제

2

시험 채점의 신뢰

채점자 관리와 판단 차이
채점 검토의 원칙

3

교사 평가의 경험

2020·2021년 등급 산정
학교 밖 표본 점검

4

공정성과 AI

편향과 문항의 맥락
AI 사용과 본인 작업 확인

5

한국 교실의 선택

예시 공유와 외부 기준점
우리 학교의 선택과 토론

강사 소개

서울대학교 물리교육과 학사·석사
중등 과학·물리 교사 경력

영국 옥스퍼드대학교 박사(2021)
University of Oxford

사우스햄튼 대학교 조교수, 부교수
University of Southampton 과학교육

잉글랜드 교육과정 개정 과학과 검토위원
국가교육과정 개정 작업 참여

연구 분야

과학교육 재난교육 교실평가 시험문화

한국과 영국의 교실·시험을 모두 경험한 연구자의 시선으로, ‘교사 판단을 신뢰할 수 있게 만드는 조건’을 함께 탐색

Journal of Research in Science Teaching

WILEY

Check for updates

RESEARCH ARTICLE OPEN ACCESS

과학의 본성을 가르치는 과학교사의 형성평가 양상

Wonyong Park[1] | Sibel Erduran[2] | Judith Hillier[2]

[1]Southampton Education School, University of Southampton, Southampton, UK | [2]Department of Education, University of Oxford, Oxford, UK

Correspondence: Wonyong Park (w.park@soton.ac.uk)

Received: 17 February 2025 | **Revised:** 24 October 2025 | **Accepted:** 13 November 2025

Keywords: conversation analysis | formative assessment | nature of science | teacher practices

ABSTRACT

While understanding the nature of science (NOS) is recognized as a key component of science education because of its potential benefits in nurturing scientific literacy, there is currently limited knowledge on how NOS learning can be supported by teachers' ongoing, in-the-moment assessment in the classroom. This study explores the value of formative assessment as a vehicle to facilitate reflective instruction of the NOS, through building connections between NOS and other related components of science learning. We closely examined one experienced science teacher's lessons on the development of scientific knowledge in a 10th grade science classroom, along with the teacher's post-lesson reflections and classroom artifacts, with a focus on how the teacher planned and orchestrated whole-class discussions following student group presentations. We identify four distinct strategies for formative assessment the teacher enacted in an effort to create "teachable moments" for NOS learning: multicontextualizing NOS, mapping examples to NOS, focusing discussion on NOS, and exploring different aspects of NOS. Across the lessons, while some teacher moves prompted discursive exchanges, others revealed significant challenges and were often met with student silence. Our work uses formative assessment as a lens to reveal the complexities and in-the-moment challenges of NOS instruction, illustrating how even an experienced teacher can struggle to facilitate these conversations. We argue that NOS research should pay more attention to the potential of formative assessment and the need for professional development to support teachers with empirically grounded assessment tasks for use in NOS instruction.

1 | Introduction

The nature of science (NOS) has been advocated as a core component of the science curriculum to accomplish scientific literacy. Accordingly, there has been a great deal of interest among science educators in assessing learners' understanding of the NOS. It has been argued that for effective NOS teaching, it should be "intentionally planned for, taught and *assessed*" (Lederman et al. 2001, 137, italics added), suggesting that high-quality assessments should be considered an essential condition for the success of NOS instruction (Allchin 2011; Hanuscin et al. 2011).

The NOS refers to "the epistemology of science, science as a way of knowing, and the values and beliefs inherent to scientific knowledge and its development" (Lederman 2007, 833). Recent studies that analyzed standards, curricula and curriculum guidelines suggest that NOS is recognized as a valuable component of science curriculum and standards documents in many jurisdictions (Erduran and Dagher 2014; Ferreira and Morais 2013; Olson 2018; Park, Wu, and Erduran 2020; Park, Yang, and Song 2020; Reinisch and Fricke 2022). Despite the widespread curricular inclusion of the NOS as a goal of science education, reviews of NOS literature across decades have

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Journal of Research in Science Teaching* published by Wiley Periodicals LLC on behalf of National Association for Research in Science Teaching.

Research in Science Education
https://doi.org/10.1007/s11165-026-10332-x

Check for updates

A Level 물리학 수험생들의 틱톡 담화

Wonyong Park[1,2]

Received: 30 October 2025 / Revised: 9 February 2026 / Accepted: 26 March 2026
© The Author(s) 2026, modified publication 2026

Abstract

In recent years, social media has become an increasingly prominent space for students seeking study support and advice. This study examines how A-level Physics—a key university-entry qualification for 16-to-18-year-olds in England—is portrayed by student creators on TikTok, a short-form video platform widely used for peer-to-peer educational guidance, information sharing, and collective knowledge building. Using reflexive thematic analysis of 57 TikTok videos via #alevelphysics, this article explores how A-level Physics is characterised and what pathways to success are constructed within these informal, student-led discourses. The analysis shows that physics is depicted as both intellectually demanding and emotionally taxing, with the examination system perceived as a flawed gatekeeper. Student creators narrate their experiences through a blend of hyperbole, frustration, and resilience, offering sophisticated peer-generated pedagogies alongside cautionary tales. Such rants frequently pivot into "here's what worked for me", foregrounding strategic resource curation, sustained practice, systematic memorisation, and deliberate over-preparation. These portrayals of subject demands and students' agentic responses on TikTok narratives resonate with prior research on the cultural construction of physics as hard and exclusionary, while also illustrating how affinity spaces on social media enable students to reframe difficulty through shared feelings and tactics. The findings underscore the role of student-led online discourses in shaping expectations around high-stakes science examinations, with implications for encouraging more students to study physics and for supporting those currently studying the subject.

Keywords Physics exams · High-stakes assessment · A-levels · Social media · Affinity spaces

✉ Wonyong Park
w.park@soton.ac.uk

[1] Southampton Education School, University of Southampton, 1 University Road, Southampton SO17 1BJ, UK

[2] Department of Science Education, Ewha Womans University, 52 Ewhayeodae-gil, Seoul 03760, South Korea

Published online: 08 April 2026

Science Education

WILEY

Check for updates

Science Education

RESEARCH ARTICLE OPEN ACCESS

한국의 임용시험은 어떤 과학교사를 만드는가?

Wonyong Park[1] | Heesoo Ha[2]

[1]Southampton Education School, University of Southampton, Southampton, UK | [2]Department of Biology Education, Pusan National University, Busan, South Korea

Correspondence: Wonyong Park (w.park@soton.ac.uk) | Heesoo Ha (hsha@pusan.ac.kr)

Received: 17 January 2026 | **Revised:** 24 June 2026 | **Accepted:** 3 August 2026

Keywords: preservice science teachers | science teacher identity | structure/agency | teacher selection

ABSTRACT

While extensive research exists on science teacher shortages, less attention has been directed toward countries facing the opposite challenge: an oversupply of science teachers. This study examines the experiences of prospective science teachers in South Korea, where passing a high-stakes, annual competitive examination is a prerequisite for securing a public school position. Drawing on interviews, study group observations, and exam artifacts, we identify key structural conditions and illustrate how the structure of the teacher selection system interacts with participants' envisioned science teacher identities and shapes their agentive strategies. We show that candidates' experiences are shaped by structural features such as the exam's emphasis on hierarchical disciplinary knowledge, procedural opacity, and competitive intensity, which condition how the exam is interpreted as legitimate, irrelevant, or merely instrumental. In response, candidates adopt contrasting strategies—ranging from adaptive alignment and pragmatic distancing to strategic optimization—that either reinforce or selectively disengage from the system's normative expectations. These strategies, in turn, interact with candidates' envisioned professional identities, influencing whether the exam is integrated into, bracketed from, or decoupled from their understandings of what it means to become a science teacher. Our findings illustrate how individual agency can reproduce the structural conditions of science teacher selection while also questioning its legitimacy within candidates' informal preparation practices, offering new insights into science teacher agency and identity-in-the-making within an underexplored, exam-based selection context.

[Preparing for the exam] really didn't suit me at all. I think I experienced burnout here and there during the process. I often felt like I really didn't want to study anymore and thought, "This isn't for me" ... I constantly questioned whether I was on the right track. On top of that, the long and demanding exam preparation was really tough for me. Of course, I had taken the university entrance exam before, but I feel like these long exams just don't suit me. The preparation process was incredibly painful.

(Hyun)

1 | Introduction

The account from Hyun, a science teacher candidate, captures the self-doubt, exhaustion, and emotional strain that can accompany becoming a science teacher in an exam-oriented

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Science Education* published by Wiley Periodicals LLC.

출발점

한국의 평가 쟁점들

1 내신 5등급제

2025학년도 고1부터 A–E 성취도와 5등급 석차등급을 함께 기재(1등급은 상위 10%)

2 서·논술형 확대 요구

국가교육위원회 국민참여위원회 숙의(2026.9): 내신 서·논술형 확대 필요 72.8%, 수능 도입 필요 55.5%. 가장 큰 우려는 채점 공정성

3 AI 채점 확대

교육부: 2029년까지 17개 시·도교육청에 AI 기반 채점 체계 구축 계획. 충남은 10월 서울·경북·세종과 AI 서·논술형 평가·채점 시스템 공동 운영 협약

4 내신 신뢰도 관리

학교–교육청–외부점검단의 3단계 점검 체계(교육부, 2023). 시·도별 성취평가 모니터링과 컨설팅 확대(예: 경북, 2026)

오늘의 질문: 서·논술형을 늘리면서도 그 채점을 믿을 수 있게 하려면 무엇이 필요할까? 잉글랜드는 같은 문제를 다른 방식으로 풀어 온 사례

출처: 교육부 (2023.12.27). 2028 대학입시제도 개편 확정안; 국가교육위원회 (2026.9.17). 대입제도 국민참여위원회 숙의 결과 보고; 교육부 (2026.8.5). 2026년 하반기 업무계획; 교육부 (2023.6.21). 고교학점제 추진 계획 보완; 경북교육청 (2026.1.19; 2026.10.16) 보도자료.

Live poll

공정한 평가란 무엇일까요?

Mentimeter live poll

공정한 평가란 무엇일까요?

생각을 한두 문장으로 남겨 주세요.

남은 시간 **1:00**

타이머 시작 초기화

시작 전

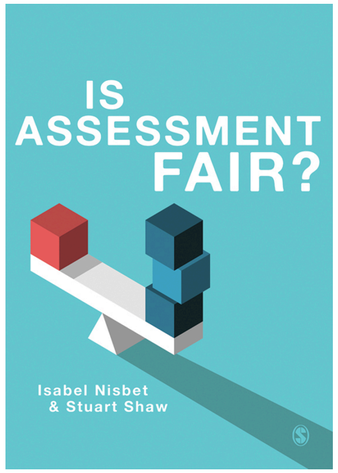
Mentimeter 새 창에서 열기 ↗

참여하기

QR코드 또는 아래
링크를 사용하여 간략히
답변을 제출해 주세요

menti.com/aloa5nrcjzcn

평가 공정성의 여섯 가지 의미



Nisbet & Shaw (2020)
Is Assessment Fair?

유형	핵심 내용	위반 예시
1 형식적 (formal)	정해진 규칙과 설계를 정확히 적용했는가? 규칙 자체의 타당성은 묻지 않는다.	교사가 공개된 채점 기준표 대신 '전체적인 인상'으로 점수를 매김
2 암묵적 계약 (implied contractual)	영향을 받는 사람들의 **정당한 기대**를 충족했는가? 명시적 약속이나 반복된 관행에서 기대가 생긴다.	'교과서에서만 출제한다'고 안내한 교사가 기말시험에 교과서 밖 문항을 출제함
3 관계적 (relational)	**관련 조건이 같은 사례**를 **같은 기준**으로 다뤘는가? 비교 가능성과 치우침 없는 판단을 포함한다.	같은 오류인데도 수업 태도가 좋은 학생은 감점하지 않고, 산만한 학생만 감점함
4 응보적 (retributive)	이전 행위에 걸맞은 보상이나 처벌을 받았는가? 즉, '**받을 만한 것을 받았는가**'를 묻는다.	학급에서 부정행위가 한 번 발생하자 교사가 반 전체의 수행평가 점수를 일괄 감점함
5 결과적 (consequential)	평가 결과가 **미래**의 부당한 조치를 정당화하는 근거로 쓰일 수 있는가?	초등 3학년 때 한 번 본 읽기 진단검사 점수로 수준별 반을 편성했는데, 이후 반 이동 기회가 거의 없고 낮은 반 교사는 기대 수준도 낮춤
6 회고적 (retrospective)	평가 결과가 **과거**의 부당한 사회적 조치에서 비롯된 것은 아닌가?	입국 후 한국어 지원을 받지 못한 이주배경 학생이 수학 서술형 문제를 풀지 못하자, 이를 수학 능력 부족으로 보고 보충반에 배정함

출처: Nisbet, I., & Shaw, S. (2020). Is Assessment Fair? 여섯 유형을 바탕으로 설명과 교실 사례를 재구성.

영국의 교육: 한 나라, 네 개의 교육제도



교육은 네 지역 정부가 각자 관할

잉글랜드·스코틀랜드·웨일스·북아일랜드의 독립적 교육 운영(교육 분야에서는 사실상 '분리된' 왕국)

잉글랜드 교육부(DfE) 관할

오늘 강의는 인구가 가장 많은 잉글랜드 사례에 한정

지역별로 다른 국가교육과정

스코틀랜드는 GCSE·A level 대신 자체 자격시험 운영

유사한 평가 문화

역사적 배경을 공유해 시험 문화와 등급 체계에 공통점이 많음

입시에서 시험의 비중이 매우 높다는 점은 한국과 동일

잉글랜드 학교와 시험, 한눈에 보기

5–11세	11–14세	14–16세	16–18세
초등학교	**중등 저학년**	**GCSE 과정**	**A level 과정**
한국 초1–초6	한국 중학교	한국 중3–고1 나이	한국 고2–고3 나이
11세 국가시험	**국가시험 없음**	**16세 GCSE**	**18세 A level**

GCSE

16세 무렵 과목별로 치르는 중등 졸업 자격시험(보통 8–10과목). 등급은 9가 최고, 7이 예전의 A(한국과 방향이 반대)

한국으로 치면 · 고1 나이에 치르는 국가 단위 중등 수료 시험 격

A level

18세 무렵 응시하는 과목별 심화 시험(보통 3과목 집중). 대학 입학의 핵심 근거

한국으로 치면 · 수능에 해당하는 대학 입시의 핵심 요소

시험 주관기관

AQA, OCR, Pearson Edexcel 등. 출제와 채점을 전담하는 민간 평가기관이며 학교가 과목별로 기관을 고름

한국으로 치면 · 여러 기관이 수능 과목을 나눠 출제·채점하는 구조

평가 체제 이해

잉글랜드의 ‘시험’은 서·논술형 시험

한국 수능

장소 교육청이 지정한 시험장

방식 지필시험(종이 문제지·OMR 답안지)

일정 전국 같은 날, 하루에 영역별로 응시

문항 대부분 선다형(수학 일부 단답형)

채점 컴퓨터(OMR)

출제 한국교육과정평가원

잉글랜드 GCSE·A level

장소 보통 재학 중인 학교·칼리지(시험장)

방식 지필 중심(일부 편의 지원은 컴퓨터 사용)

일정 5–6월 여러 날에 나누어 응시.
같은 시험지는 공통 일자·회차에 응시

문항 선다형은 일부, 대부분은 단답형부터 6점 서술형·긴 논술형까지 학생이 직접 쓰는 답안

채점 사람(채점자)이 채점
2025년 여름에만 약 650만 개의 등급 산출

출제 시험 주관기관(exam boards)이 출제·채점

AQA Questions matter
Pearson | Edexcel
CAMBRIDGE OCR
eduqas Part of WJEC

출처: Ofqual (2026). How GCSE and A level marking works; Ofqual (2025). Reviews of marking and moderation for GCSE, AS and A level: summer 2025 exam series (grade counts). 운영 방식: KICE (2026), 2027학년도 수능 수험생 유의사항; Ofqual (2026–27), exam guides; JCQ (2026–27), Instructions for Conducting Examinations.

평가 체제 이해

진학을 좌우하는 성적, 누가 어떻게 매길까?

구분	잉글랜드	한국
외부 시험	GCSE(16세)·A level(18세). 시험 주관기관이 출제·채점	수능(대입). 한국교육과정평가원이 출제
문항 형식	대부분 단답·서술·논술형. 사람이 채점	수능은 선다형 중심. 컴퓨터로 채점
교사 채점이 진학에 반영되는 범위	일부 비시험평가(과학 실험, 미술, 디자인·기술 등)와 대학 지원용 예측 등급	고교 내신 전체: 지필고사(서·논술형 포함)와 수행평가(대개 40% 이상)
점검 주체	Ofqual(자격·시험감독청)이 시험 주관기관을 감독. 교사 채점 과제는 시험 주관기관이 표본 점검·조정	학교 학업성적관리위원회, 시·도교육청, 교육부의 지침과 점검

대학 지원과 '예측 등급': 고3(Year 13) 가을, 교사가 예상한 A level 등급으로 대학에 먼저 지원해 조건부 합격을 받은 뒤, 이듬해 8월 실제 시험 성적으로 입학 확정

	9–1월	1–5월	5–6월	8월
학생	지원 대학 선택 예측 등급으로 지원	조건부 합격 확인 대학 선택·시험 준비	A level 시험 응시	실제 등급 확인 조건 충족 시 입학 확정
교사	예측 등급 산정 추천서와 함께 제출	수업·모의시험 피드백 목표 등급 달성 지원	복습 지도 시험 준비 지원	결과·진학 상담 필요 시 추가모집 안내

출처: Ofqual (2026). How GCSE and A level marking works; 교육부 (2023.12.27). 2028 대학입시제도 개편 확정안. UCAS. Key dates timeline · Predicted grades; AQA. 2027 exam timetable.

핵심 주장

질문은 ‘교사를 믿느냐’가 아니라 ‘판단을 믿을 만하게 만드는 조건이 무엇이냐’

평상시
조건

채점자도 교사다

달라진 것은 사람이 아니라 조건: 이름 가리기, 공통 예시, 연수, 점검, 보수가 지급되는 채점 시간

2020–21
비상 조치

판단이 교사에게 돌아오다

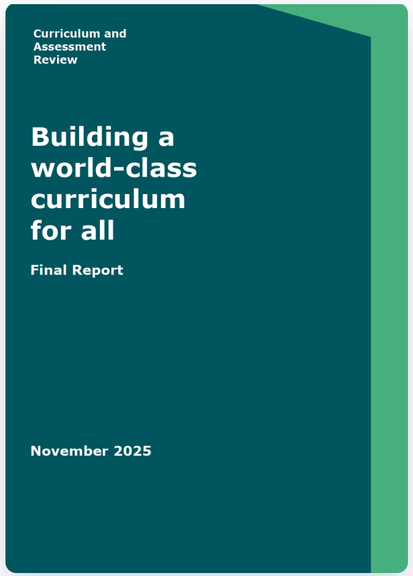
코로나19로 시험이 취소되자 조건을 갖출 틈 없이 교사가 등급을 매김 → 성적 부풀리기, 학교 간 격차, 학부모 압력

한국에의
시사점

잉글랜드로부터의 교훈

서·논술형 확대를 앞둔 한국 학교에서 이 조건들을 어떻게 갖출 수 있을까

잉글랜드는 지금 어디로 가고 있나: 2025년「교육과정·평가 검토」



시험 중심 기조 유지

정부 의뢰 검토 보고서(2025년 11월)에서 시험을 가장 공정하고 신뢰할 수 있는 방식으로 재확인하며 GCSE의 중심에 둘 것을 권고

AI 시대 과제 평가의 한계

수업 중 수행하는 과제(코스워크)는 생성형 AI로 인해 학생 본인 작성 여부를 검증하기 어려워졌다는 점을 근거로 제시

수험 부담 완화 병행

학생당 GCSE 총 시험 시간을 2.5–3시간 단축할 계획(개정 GCSE는 2029년 적용 예정)

잉글랜드가 지키려는 '시험'은 서·논술형 지필시험. 한국이 늘리려는 서·논술형 평가를, 잉글랜드는 외부 시험 안에서 이미 대규모로 운영하며 공정성·신뢰·AI의 고민을 함께 안고 있음

출처: Curriculum and Assessment Review (2025). Building a world-class curriculum for all: Final report. Department for Education.

이제부터 볼 여섯 가지 사례

1 평상시 시험

시험 체제는 채점자의 판단을 어떻게 관리할까?

- A 채점자 관리
- B 채점은 얼마나 일치할까
- C 채점 검토의 원칙

2 코로나19 시기 교사 평가

판단이 교사에게 돌아오면 무슨 일이 생길까?

- D 2020년 알고리즘 사태
- E 2021년 교사 산정 등급

3 대안과 쟁점

교사 판단을 믿을 만하게 만드는 다른 길은?

- F 외부 평가 조정
- 공정성: 편향, 문항 맥락, AI

6점짜리 과학 서술형, 어떻게 채점할까?

AQA GCSE 통합과학(2023년 6월) · 잔디 잎에서 색소를 추출하고 분리하는 방법을 서술하시오. (6점)

수준	점수	수준별 기술문(요약)
3	5–6	타당한 결과를 얻을 수 있는 방법. 핵심 단계를 모두 짚고 논리적인 순서로 제시
2	3–4	타당한 결과를 얻지 못할 수도 있는 방법. 대부분의 단계를 짚었지만 순서가 완전하지 않음
1	1–2	타당한 결과를 얻을 수 없는 방법. 일부 단계만 있고 단계 간 연결이 분명하지 않음
0	0	관련 내용 없음

예시 내용(indicative content)

- 용매를 넣고 잎을 으깬 뒤 거르기
- 거름종이에 연필로 선을 긋고 추출액 점찍기
- 용매 높이를 그 선보다 낮게 유지하기

정답표 대신 '수준'으로 판단하는 구조. Pearson Edexcel 등 다른 출제기관도 같은 방식 사용

출처: AQA (2023). GCSE Combined Science: Synergy 8465/1H, Mark scheme, June 2023; Pearson Edexcel. GCSE Sciences: Extended open response guidance.

채점자 지침: 체크리스트가 아니라 ‘가장 잘 맞는 수준’

1

수준부터 결정

가장 낮은 수준부터 기술문을 사다리처럼 올라가며 검토. 여러 수준의 특징이 섞여 있으면 ‘가장 잘 부합하는 수준(best fit)’ 선택

2

해당 수준 안에서 점수 확정

기술문과 책임 채점자가 채점한 예시 답안을 참고한 뒤, 학생 답안을 다시 읽으며 점수 부여

3

예시 내용은 체크리스트가 아님

예시 목록에 없더라도 과학적으로 타당하면 점수 부여. 모든 요소를 다 쓰지 않아도 최고 수준 도달 가능하나, 만점을 받으려면 모순되거나 틀린 내용이 없어야 함

왜 이렇게까지 할까?

채점자 간 판단 차이가 가장 크게 벌어지는 문항이 바로 6점 서술형. 따라서 기술문만 제공하지 않고, 점수와 채점 근거를 명시한 예시 답안을 함께 제공

출처: AQA (2023). GCSE Combined Science: Synergy 8465/1H, Mark scheme, June 2023 (guidance for examiners).

채점자 지침(Information to Examiners)

3.2 기호 / 화학식의 사용	학생이 요구된 화학명 대신 화학 기호 / 화학식을 쓰거나 물리 공식에서 양을 나타내기 위해 기호를 사용하는 경우, 그 기호 / 화학식이 정확하고 해당 문항의 맥락에서 그러한 사용이 적절하다면 만점을 부여할 수 있음.
3.3 계산 문항의 채점 절차	학생들에게 풀이 과정을 제시하도록 지시하므로, 올바르게 수행한 계산의 각 단계에 점수를 부여해야 함. 학생은 계산의 어느 지점에서든 풀이 단계를 생략할 수 있음. 이후 단계가 올바르게 제시되면 해당 점수를 부여할 수 있음. 잘못된 풀이 과정에서 나온 최종 답이 옳더라도 만점을 부여하지 않음.
3.5 이전 단계 오류의 처리	오류는 한 소문항에서 다음 소문항으로 이어질 수 있음. 이를 약어 ‘ecf’로 표시함. 하나의 소문항 안에서 계산의 한 단계에 잘못된 값이 있더라도 이후의 모든 점수를 부여할 수 없게 되는 것은 아님.
3.6 발음에 따른 철자	철자가 정확하지 않더라도 의도가 분명하면 점수를 부여해야 함. 다만 다른 전문 용어와 혼동할 가능성이 있는 경우는 제외함.
3.8 허용(Allow)	채점 기준의 추가 정보에서 ‘allow’는 점수를 인정할 수 있는 대안 답안을 나타내는 데 사용됨.

출처: AQA (2023), 8465/1H 채점 기준, pp. 4–5, Information to Examiners, §§3.2, 3.3, 3.5, 3.6, 3.8. 발췌 직역.

사례 A · 시험 채점자는 누구이고, 어떻게 채점할까?

약 20년

채점자의 평균 교직 경력. 대부분 전·현직 교사이며 시험 주관기관이 모집·보수 지급

약 10년

평균 채점 경력. 매년 같은 채점자가 참여하며 경험 축적

0명

채점자가 아는 학생 수. 모든 답안은 이름과 학교를 가린 채 채점

채점자도 교사. 달라진 것은 사람이 아니라 조건: 이름 가리기, 공통 예시, 채점 기준 맞추기 연수, 점검, 보수가 지급되는 채점 시간

채점자 12,681명에게 물었더니 (Ofqual 조사)

대부분 온라인 채점 연수에 만족했으나, 예전 대면 모임에서 나누던 '채점 기준표에 대한 깊은 토론'을 아쉬워하는 응답도 많음. 판단을 맞추는 데는 결국 대화가 중요하다는 시사점

출처: Ofqual (2026). How GCSE and A level marking works; Ofqual (2018). Online standardisation.

사례 A · 기준표만 믿지 않는다: 판단을 둘러싼 다섯 겹의 장치

1

채점 기준표

점수 부여 방식 명시(서술형은 수준별 기술문과 예시 내용 병기)

2

채점 기준 맞추기 연수

본채점 전 실제 답안으로 연습하고, 기준에 도달했음을 입증해야 채점 자격 부여

3

점검용 답안 투입

선임 채점자가 미리 '기준 점수'를 확정한 답안을 채점 대기열에 몰래 섞어 넣음(채점자는 식별 불가)

4

오차 초과 시 중단

허용 오차를 벗어나면 즉시 채점 중단 및 재연수. 이미 채점한 답안도 재점검 가능

5

결과 발표 후 검토

학교가 채점 검토를 신청할 수 있으나, 채점 오류가 있을 때만 점수 변경(사례 C)

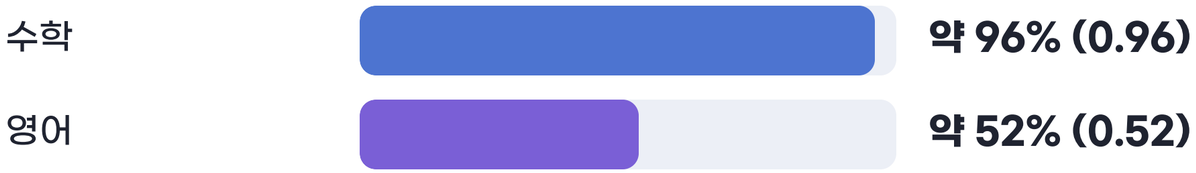
한국에 중요한 이유

채점의 일치는 기준표 문구가 아니라 '함께 본 예시와 피드백'에서 나옴. 한국 학교의 채점 협의회가 하는 일을 잉글랜드는 제도로 의무화하고 효과까지 점검

출처: Ofqual (2026). How GCSE and A level marking works.

사례 B · 선임 채점자가 채점해도 같은 등급일까?

실제 등급이 그대로 유지될 확률(추정)



점수 차이가 등급을 바꾸는가?

가상 예: A등급 경계가 70점일 때
실제 채점 68점(B), 선임 기준 72점(A)이면 불일치.
선임 기준이 69점(B)이면 점수는 달라도 등급은 일치.

같은 학생의 같은 답안으로 비교

실제로 받은 등급

실제 채점자들이 부여한 점수를 합산해
성적표에 기록한 최종 과목 등급

선임 채점자가 채점했다면 받을 등급

같은 답안을 선임 채점자의 판단대로
채점했다고 가정할 때 나올 최종 과목 등급

불일치를 모두 오채점으로 볼 수는 없음.
타당하게 인정할 수 있는 점수가 여럿일 수 있음.

출처: Ofqual (2018). Marking consistency metrics: An update, pp. 6–16, 21, 35 (2017 summer data).

사례 B · 과목별 차이의 원인과 해석

문항 유형의 차이

수학은 짧고 정답이 명확한 문항이 많으나, 영어·역사는 긴 논술형 비중이 커서 채점자 간 판단이 정당하게 갈릴 여지가 큼

등급 구간의 폭

한 등급에 해당하는 점수 폭이 좁을수록 작은 채점 차이에도 등급 변동(채점만큼 시험 설계가 중요)

엄격한 비교 기준

선임 채점자의 단일 기준 점수와만 비교한 결과. 정당하게 인정 가능한 인접 점수까지 포함하면 일치 확률은 더 높아짐

대부분의 차이는 한 등급 이내

모든 시험에서 '동일 등급 또는 한 등급 이내'일 확률은 95% 이상. 5년간 안정적이었으며, 2인 채점을 도입해도 개선 폭은 크지 않았음

출처: Ofqual (2018). Marking consistency metrics: An update.

사례 C · 오류는 고치고, 정당한 점수는 그대로 둔다

잉글랜드의 채점 검토: 학교가 시험 주관기관에 건당 수수료(수만 원 수준)를 내고 신청, 등급이 바뀌면 환불. 학교 안에서 무료로 처리하는 한국의 이의신청과 다른 구조

2016년 이전의 문제점

- 재채점 시 '합리적인 점수'를 또 다른 '합리적인 점수'(대개 더 높은 점수)로 바꾸는 관행 만연
- GCSE 수학 검토 신청자의 86%는 등급 변동 없음. 등급이 바뀐 경우의 87.5%도 3점 이하의 미세한 차이
- 수수료를 감당할 수 있는 학생만 검토를 신청해 유리해지는 형평성 논란

Ofqual의 개정 원칙(2016)

- 정답·오답이 명확한 문항: 잘못 부여된 점수는 즉시 바로잡음
- 전문적 판단이 필요한 문항: 최초 채점 판단이 명백히 비합리적으로 적용된 경우에만 점수 변경

"검토자가 다른 점수를 주었을 것이라는 이유만으로, 합리적인 점수를 또 다른 합리적인 점수로 바꿔서는 안 된다."

Ofqual 지침, 2016

출처: Ofqual (2016). Marking mistakes must be corrected, legitimate marks should stand; Ofqual (2016). Guidance on considering marking errors on a review or appeal; Ofqual (2016). Regulatory impact analysis.

사례 C · 2025년 여름, 이의 제기와 등급 변경의 실제 규모

자격시험	부여된 등급	이의 제기 등급	실제 변경 등급	전체 대비 변경률
GCSE	5,661,970	243,150 (4.3%)	58,725 (이의의 24.2%)	1.0%
AS·A level	864,810	58,830 (6.8%)	14,150 (이의의 24.1%)	1.6%

어떤 과목이 많이 바뀌었나

- 변경된 등급은 거의 모두 상향 조정
- GCSE: 경제 3.4%, 체육 2.4%, 고전 2.1%
- A level: 음악 3.1%, 경제 3.0%

출처: Ofqual (2025). Reviews of marking and moderation for GCSE, AS and A level: summer 2025 exam series. Official statistics.

사례 D · 2020년 코로나19로 인한 시험 취소 사태

2020년 3월 코로나19로 GCSE·A level 시험 전면 취소. 시험 없이 진학용 등급을 정해야 했고, 정부는 교사 판단에 통계 모형을 더하는 방식을 선택

1단계 · 봄

학교가 등급과 석차 제출

모든 학생의 학교 산정 등급과 과목별 석차 제출

2단계 · 여름

통계 모형으로 조정

학교별 과거 시험 성적 분포에 맞춰 등급 조정(단, 수강생이 적은 소규모 반은 교사 등급 유지)

출처: Schools Week (2020). A-level results 2020; Tes (2020). U-turn will mean A-level and GCSE teacher grades stand; Royal Statistical Society (2020). Significance, October issue.

사례 D · 결과 발표 나흘 만에 철회된 통계 조정

8월 13일

A level 결과 발표

교사가 부여한 등급의 39.1%(약 28만 건)가 하향 조정되어 발표

8월 17일

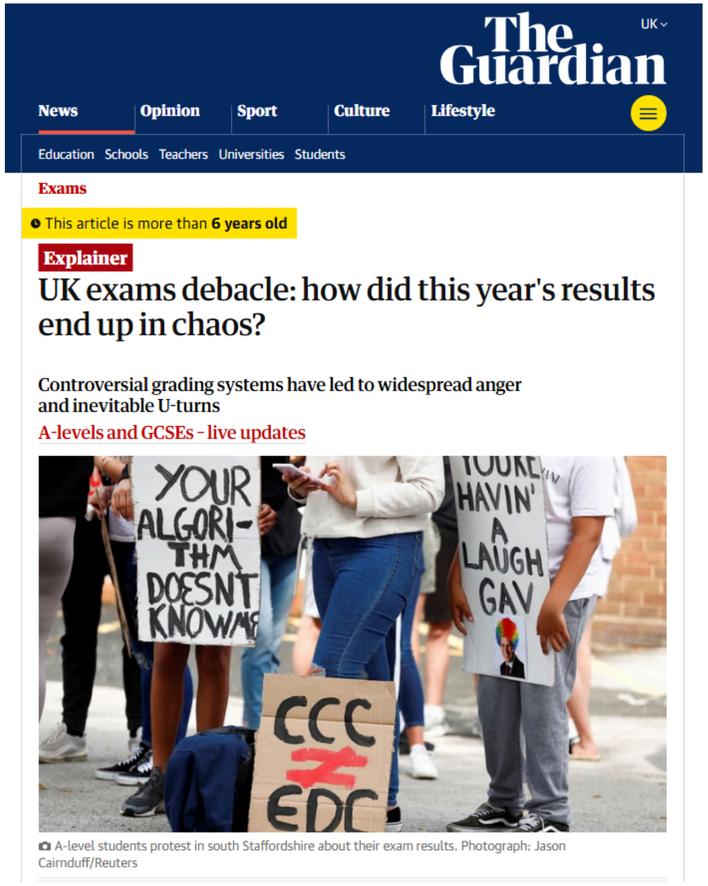
나흘 만의 전격 철회

거센 사회적 항의 끝에 알고리즘 결과를 철회하고, 교사 산정 등급과 조정 등급 중 높은 쪽을 최종 인정

왜 불공정하다고 느꼈을까?

- 알고리즘이 개별 학생이 아닌 '학교의 과거 성적'을 평가(과거 성적이 낮았던 학교의 우수 학생이 본인 실력과 무관하게 하향 조정되는 불이익)
- 소규모 반 비중이 높은 사립학교(정부 지원 없이 높은 학비로 운영, 학생의 약 7%) 학생들이 통계 조정을 피해 갔다는 형평성 비판

결과: 연도 간 비교 가능성보다 개인에 대한 공정성을 우선



출처: Schools Week (2020); Tes (2020); RSS, Significance (Oct 2020); The Guardian (17 Aug 2020).

사례 D · 교사가 매긴 등급에서 보인 경향과 두 가지 교훈

2020년 학교 산정 등급 분석 결과 (Ofqual)

- 예년보다 평균 약 반 등급 상승
- 과목 수강생이 적을수록 상승 폭이 큼(A level 소규모 0.77등급, 대규모 0.47등급)
- 교사 채점 과제 비중이 큰 예체능·응용 과목의 상승 폭이 가장 큼
- 사립학교와 기존 성적이 낮던 학교의 상승 폭이 큼
- 특정 학생 집단에 대한 뚜렷한 편향은 발견되지 않음

① 외부 기준점이 없으면 판단은 위로 움직인다

함께 참고할 객관적 외부 기준이 없으면 선의의 판단도 관대해지며, 관대화 정도 역시 학교마다 달라짐

② 통계 보정은 개인의 공정성을 해친다

과거 성적 분포에 맞추면 연도 간 비교 가능성은 유지되지만, 개별 학생에게는 명백히 부당한 결과 초래

한국에 중요한 이유

상대평가·절대평가 논쟁의 압축판. 절대평가는 성적 부풀리기로, 과거 분포에 맞춘 규준 보정은 '소속 학교에 따른 불이익' 문제로 직결

출처: Ofqual (2021). An evaluation of centre assessment grades from summer 2020.

사례 E · 2021년, 교사가 진학 등급을 직접 매기다

2021년 시험 재취소. 이번에는 알고리즘 없이 교사가 등급을 직접 부여. 한국 교사들이 매년 내신으로 하는 일을 잉글랜드 교사들이 한 해 동안 경험한 셈이며, 오늘 강의의 핵심 사례

1

학교 평가 방침 수립

근거 자료 활용법, 교내 채점 기준 맞추기 절차, 차별 방지 방안을 문서화. 시험 주관기관이 모든 학교의 요약본 검토

2

비대면 방문 점검

평가 방침에 우려 사항이 있으면 시험 주관기관이 학교의 교내 채점 협의 절차를 직접 확인

3

표본 점검

등급 제출 후 표본 학교(무작위 및 예년 대비 성적 차이가 큰 학교)는 학생 5명 이상의 채점 근거 자료를 48시간 내 제출

학교 안에서는: 런던의 공립 중등학교 Mulberry Stepney Green의 방침

관리자·교과 부장의 교과별 품질 관리 지원 · 담당 교사가 한 명뿐인 과목은 별도 점검 · 외부 검토 대비 근거 자료 전부 보관 · 예년과 다른 결과도 이유를 기록하면 인정

출처: Ofqual (2021). Quality assurance for GCSE, AS and A level: information for schools and colleges [blog]; Mulberry Stepney Green (2021). Centre appeal policy: teacher assessed grades 2021.

사례 E · 절차는 갖췄지만, 성적 부풀리기는 막지 못했다

지표	수치
제출된 교사 산정 등급	570만 건
점검 뒤 바뀐 등급	195건 (내림 179, 올림 16)
근거 자료를 점검받은 학교	5곳 중 약 1곳
A level A 이상 비율	44.3% (2019년 25.2%, 2020년 38.1%)
GCSE 상위 등급(7 이상, 9가 최고) 비율	30.0% (2018–2019년 21.8%)

교사들은

교직원 1,785명 조사: 거의 모두 자신이 매긴 등급이 정확하다고 응답. 그러나 대부분 스트레스와 업무 부담이 커서 다시 하고 싶지 않다고 응답

학생들은

학교 간 기준 차이에 대한 우려. '우리 선생님이 다른 학교보다 깐깐했다면?'

단서: 2021년은 사전 준비 없이 몇 달 만에 시행한 비상 조치. 교사 평가의 공정한 시험대라기보다, 조건을 갖추지 못한 채 교사에게 맡긴 판단의 사례

출처: Ofqual (2021). GCSE, AS and A level summer report 2021; Ofqual (2022). Teacher-assessed grades in 2021: student and teacher experiences; FE Week; Schools Week; Tes (2021).

사례 E · 교사 판단에 대한 압력

절반

교직원 약 1,500명 중 절반이 관리자·학부모·학생의 ‘부당한 압력’ 경험(2020년 31%)

23% vs 11%

학부모 압력을 받은 교사 비율: 사립학교 vs 저소득층 학생이 많은 공립학교

등급은 전문적 판단의 문제로, 학부모와 협상할 대상이 아님

Ofqual 의장의 공개 발언 요지

한국에 중요한 이유

- 교사 판단이 진학을 직접 좌우하게 되자, 잉글랜드 역시 한국과 유사한 보호 장치(서면 방침, 교내 채점 협의, 근거 보관, 외부 표본 점검)를 즉각 도입
- 평가에 대한 외부 압력은 특정 국가의 문화가 아니라 ‘평가 결과가 진로에 미치는 중대성(고부담 여부)’에서 비롯됨

출처: FE Week. ‘Stressful and exhausting’: Ofqual report exposes teacher grade woes; IE Today. Private schools facing most pressure from parents over grades (Teacher Tapp); Tes (2021). The 7% private-school figure is a commonly cited estimate.

사례 F · 교사평가에 대한 외부 감사 체계

① 비시험평가(GCSE·A level)

- 학교 내 채점 협의 후 시험 주관기관이 점수대별 표본을 추출해 검토
- 학교 점수가 허용 오차를 벗어나면 학교 전체 점수를 일괄 상·하향 조정

② 초등 6학년 쓰기 평가

- 국가 기준에 따른 교사 평가 후, 지역 교육청이 매년 공립학교의 25% 이상을 방문·외부 점검
- 해설이 달린 국가 예시 자료를 공통 기준점으로 활용

③ A level 실험 역량 인증

- 교사가 수업 전반에서 공통 기준으로 실험 역량을 평가하고, 등급과 별도로 '통과/미분류' 표기
- 시험 주관기관이 학교를 방문해 이행 여부 점검

약 70%

초등 6학년 쓰기 349편: 교사와 선임 점검자의 판단 일치율(2020)

한국에 중요한 이유

한국의 수행평가는 잉글랜드의 비시험평가에 해당. 한국에도 학교–교육청–외부점검단의 점검 체계와 시·도별 모니터링·컨설팅이 있음. 다만 답안 표본을 다시 채점해 학교 점수를 조정하는 방식은 아님. 이 차이를 어떻게 메울지가 과제

출처: Ofqual (2022). Marking and moderation; 교육부 (2023.6.21). 고교학점제 추진 계획 보완; STA (2025). Teacher assessment moderation, standardisation and training process; Tes (2020). 1 in 3 primary statutory writing grades are wrong; AQA. A-level Physics 7408 specification.

사례 F · 교사 채점 과제가 학교 책무성 평가와 얽히면: 코스워크가 줄어든 역사

연도	결정	이유
2013	말하기·듣기를 GCSE 영어 등급에서 제외	학교 간 운영·채점의 일관성을 보장하기 어려웠음
2013	2009년 도입한 '통제된 과제 평가'도 문제가 있다는 검토 결과	의도한 능력을 늘 평가하지는 못했음
2013–2015	새 GCSE는 시험이 기본, 시험으로 평가할 수 없는 능력만 교사 채점 과제로	교사 채점 과제의 신뢰도, 학생 본인이 했는지에 대한 우려
2018	GCSE 컴퓨터과학 프로그래밍 과제를 등급에서 제외	규정 위반(부정행위) 사례 확인

코스워크 축소의 배경

잉글랜드는 학교별 성적을 공개하고 이를 학교 책무성 평가에 활용. 교사가 채점한 점수가 여기에 반영되면 점수를 후하게 줄 압력 발생

반론

과제를 줄이면 평가 근거도 줄어 신뢰도가 오히려 낮아진다는 비판(의회). 타당도·신뢰도·신뢰 사이의 균형 문제

출처: Ofqual (2013). Review of controlled assessment in GCSEs; Ofqual (2013). Our announcement on speaking and listening assessments [blog]; Ofqual (2018). Decisions on GCSE computer science assessments; Hansard (11 June 2013). GCSEs.

2018년 컴퓨터과학: 누구의 공정성을 지킬 것인가?



SCHOOLS WEEK

20 November 2026

News Jobs Opinion Features The Leader Weekly Edition Other

Botched computing test won't count towards final GCSE grades, Ofqual confirms

A 20-hour non-examined assessment in computer science will not count towards pupils' final GCSE grades in 2018 or 2019 after tasks from the test were leaked online, Ofqual has confirmed. The regulator proposed changes to the assessment, currently worth 20 per cent of the overall grade, after it discovered that tasks and detailed solutions were posted on [...]

Alix Robertson

2 min read | 8 January 2018, 2:00 pm

과제·해답 유출 후 20% 과제 점수를 등급에서 제외(2018·2019).
학교별 유불리를 줄이려고 모든 학교의 과제 수행 20시간은 유지.

> "students whose breach of the rules is not detected will be advantaged …"
> 적발되지 않은 규칙 위반 학생이 유리해진다.
> Ofqual (2018), p. 8

형식적	과제 유출·외부 도움 금지 규칙이 모두에게 지켜졌는가?
암묵적 계약	과제 20% 반영을 믿고 준비한 학생의 기대를 바꿔도 되는가?
관계적	유출 자료 접근 여부, 과제 수행 시점이 다른 학생을 공평하게 비교할 수 있는가?
응보적	규칙을 지켜 완성한 과제의 성취를 등급에서 인정하지 않아도 되는가?
결과적	부정행위 방치와 규칙 변경이 학습 의욕·진학 기회에 미칠 영향은?

출처: Ofqual (2018), Decisions on GCSE computer science assessments regulated by Ofqual, pp. 7–9; Nisbet & Shaw (2020), Is Assessment Fair?

교사의 평가 권한: 신뢰와 우려

교사의 판단을 신뢰하는 근거

2020 · 지역사회의 반발

런던 Newham 구: 통계 조정은 성실하게 평가한 교사들의 정직성을 훼손했다는 비판.

2021 · 정부의 전환

교육부는 알고리즘 대신 교사를 신뢰하겠다고 선언. 시험이 불가능한 상황의 대안으로 채택.

평상시 · 전문성의 활용

외부 시험도 대부분 전·현직 교사가 익명으로 채점. 공통 기준과 점검 아래 판단을 활용.

전문성과 등급 판단에 대한 신뢰

Ofqual·YouGov, 2021년 9월 · 3,565명 · 동의 비율

정확히 평가할 전문성이 있다



64%

지나치게 엄격하거나 관대하지 않게 등급을 매겼다고 신뢰한다



46%

7개 이해관계자 집단 합산(교사·학생·학부모 등).

외부 검토는 공정성 확인과 교사 보호를 함께 겨냥

2021 · 학생·학부모의 요구

등급 이의를 학교 밖 제3자가 검토해야 한다는 요구.
판단한 학교와 최종 검토 기관을 분리.

2021 · 교사 판단의 보호

성적은 협상이 아닌 전문적 판단의 문제라는 입장.
외부 점검은 근거 없는 상향 압력을 거절할 근거.

출처: Newham Council (2020); HC Deb (6 Jan 2021); Ofqual/YouGov, Public Sentiment 2021 (2022), fig. 38; Ofqual, ASCL speech (2021), How marking works (2026).

공정성 ① 채점이 일치해도 불공정할 수 있다: 교사 판단의 편향

학생 특성	교사 평가와 시험을 비교한 연구들
성별	결과가 엇갈림. 여학생에게 약간 유리하다는 보고가 많음
인종·민족 배경	결과가 엇갈림. 유리·불리·차이 없음이 모두 보고됨
사회경제적 배경	형편이 어려운 학생에게 불리하다는 보고가 많음
특수교육 대상	특수교육 대상 학생에게 불리하다는 보고가 많음

Ofqual 문헌 검토(2021). 차이가 곧 교사의 잘못은 아니나, 교사 평가는 더 주관적이어서 편향에 더 취약

일곱 살에 벌써 고정관념? (Campbell, 2015)

7세 약 5,000명 분석. 수학 시험 점수가 같아도 여학생은 남학생보다 '평균 이상' 평가를 덜 받음

익명 시험과의 비교 (Burgess & Greaves, 2013)

초등 국가시험과 교사 평가 비교: 일부 집단은 일관되게 낮게, 일부는 높게 평가됨

2021년에 요구한 보호 장치

판단마다 동료와 논의, 필요하면 특수교육 전문가 참여, 등급마다 교사 두 명 이상 승인

출처: Lee & Newton (2021). Systematic divergence between teacher and test-based assessment. Ofqual; Campbell, T. (2015). Journal of Social Policy, 44(3); Burgess, S. & Greaves, E. (2013). Journal of Labor Economics, 31(3).

공정성 ② 배경지식을 재는 문항, 그리고 정치적 중립

낯선 맥락은 지식을 평가하기 어렵게 만듦

Ofqual 지침(2022): 문항 맥락은 측정하려는 능력의 발휘를 도와야 함

- 독일어 시험(2019): 스키 여행의 장단점 → 스키를 타 본 적 없는 학생은 불리
- 수학 문항(2017): 극장의 위층·아래층 좌석 → 극장에 가 본 경험이 필요
- ‘Skittles’: 영국 전통 펍 게임이지만 십 대 대부분은 같은 이름의 사탕으로 읽음
- 기니피그 우리 설계 과제: 기니피그를 모르는 학생은 문제 자체를 이해하지 못함

GCSE 문항은 사전 검증을 거치지 않아 편향이 시험 후에야 드러남

논쟁적 내용과 정치적 중립

- 잉글랜드 교육부 지침(2022): 당파적 견해 주입 금지, 상반된 견해의 균형 있는 제시
- 평가 기준은 특정 입장에 대한 동의 여부가 아니라 ‘논증과 추론의 질’을 측정해야 함

한국에 중요한 이유

- 다문화·이주배경 학생 증가로, 특정 계층의 생활 경험을 전제한 서술형 문항과 학생 배경에 따른 교사의 무의식적 기대가 공정성을 위협
- 익명 채점과 출제·검토 단계의 문항 맥락 점검은 저비용으로 즉시 도입 가능한 보호 장치

출처: Ofqual (2022). Guidance on designing and developing accessible assessments; AQA. Accessible for all; OCR. When context is king; PA Media (2019); DfE (2022). Political impartiality in schools.

AI 시대의 평가 공정성: 학습의 도구와 평가의 증거 구분하기

잉글랜드의 선택: 새 AI 규정을 만들기보다, **기존 평가의 '진위(authenticity)' 원칙**을 다듬는 방향

기관별 역할

- 교육부: AI 활용 범위는 학교가 판단, 평가에서는 부정행위 방지 의무 준수
- Ofqual: 새 규정 대신 기존 부정행위 관리 의무 적용. 2026년 3월 시험 주관기관에 진위 확인 강화 요구
- 시험 주관기관 협의체: 제출물은 **학생 본인의 독립적 결과물**이어야 함(2025년 4월 개정 지침)

허용과 금지의 경계

- 허용 예: 자료 조사, 글의 기본 구조 계획, 노트 요약(OCR 영어 과제 지침)
- 금지 예: 분석 내용 생성, 학생 글의 재작성·덧붙이기
- AI 사용을 밝혀도 학생이 **스스로 하지 않은 부분은 점수 없음**(시험 주관기관 협의체)
- 초등 6학년 쓰기: AI로 쓰거나 AI 피드백으로 고친 글은 **교사 평가 증거에서 제외**(국가시험 담당 기관)

기록으로 진위 증명

- AI 사용 시 도구명·날짜, **질문과 출력의 사본(캡처)**, 활용 방식 설명 제출
- 학생과 교사 모두 **진위 확인서에 서명**
- 서명 전 의심: 학교 자체 처리 / 서명 후: 시험 주관기관에 공식 보고

'AI 사용 금지'라는 방향 대신, 과제별로 '허용되는 보조'와 '평가 대상의 대행'을 구분해 명시하고 과정 기록을 남기는 장치 필요

출처: DfE (2025). Generative artificial intelligence (AI) in education; Tes (2026, 3 March). Ofqual calls for tougher action on phones and AI cheating; JCQ (2025). AI Use in Assessments: Your role in protecting the integrity of qualifications (Revision 2); OCR (2024). Guidance and support for the use of AI in English; STA (2026). Key stage 2 teacher assessment guidance 2027.

시험 주관기관 협의체가 공개한 실제 사례: 번역기, 탐지기, 그리고 학교의 책임

사례	상황과 처분	드러난 쟁점
A level 역사 (비시험평가, AQA)	학생 A: ChatGPT가 만든 작성 지침을 실수로 제출 → 자격 실격. 학생 B: 내용 작성에 ChatGPT 사용 인정 → 해당 과제 0점	처분의 비례성: 위반 정도와 징계 수위의 균형
논문형 확장 과제 (A level 수준, Pearson)	출처 표시 없는 AI 생성 부분 확인, 학생은 소명 거부 → 자격 실격	점수 불인정(0점)과 부정행위 징계는 별개의 판단
A level 미술 (OCR)	스케치북 조사 내용을 DeepL로 번역. 허용되지 않음을 알고 있었음. 학교 추정 98% AI 영향 → 해당 영역 0점	번역기도 AI: 언어 지원과 평가 대상의 경계
GCSE 종교학 (워드프로세서 시험, AQA)	인터넷 차단이 안 된 컴퓨터로 AI 사용 → 해당 영역 0점, 감독관 재교육	학생 개인뿐 아니라 시험 환경을 관리할 학교의 책임

출처: JCQ (2025). AI Use in Assessments: Your role in protecting the integrity of qualifications (Revision 2), Appendix A: AI misuse case studies.

AI 시대의 평가 설계: ‘적발’에서 ‘과정과 본인 작업 확인’으로

잉글랜드가 강조하는 원칙

규칙의 ‘공지’가 아닌 ‘이해’:

AI 규정을 따로 안내받지 못한 학생들이 적발된 사례도 있음. 서명 전 구체적 사례로 교육

과정으로 진위 확인:

일부 과제는 교사의 직접 감독 아래 작성, 작업 기록, 학생과의 짧은 구술 대화로 본인 작업인지 확인

언어 지원의 경계:

번역 도구도 AI. 다른 언어로 쓴 뒤 번역해 제출하는 것은 불허 (Pearson)

사람의 판단은 대체 불가:

“AI는 인간 전문가의 판단과 깊은 교과 지식을 대신할 수 없다”(교육부)

출처: JCQ (2025). AI Use in Assessments (Revision 2); Pearson. A level English Literature: Guidance on AI and coursework; DfE (2023). Generative artificial intelligence (AI) in education.

정리

가져올 것은 장치가 아니라 원칙

잉글랜드의 장치는 유급 채점자와 국가 단위 조직을 전제로 작동.
교과 교사가 한두 명뿐인 한국 학교에 그대로 이식할 수는 없음.

원칙 1
예시

예시의 공유

기준 문구보다 채점된 예시 답안과 그 근거가 판단을 일치시킴

원칙 2
구분

오류와 차이의 구분

잘못된 점수는 바로잡고, 정당한 판단 차이는 인정

원칙 3
기준점

학교 밖 기준점

외부 기준점이 없으면 선의의 판단도 관대해지고 학교마다 달라짐

토론

함께 이야기해 볼 세 가지 질문

1. 잉글랜드는 수업을 담당한 교사와 진학을 위한 시험의 채점자를 점점 구분하는 방향으로 평가제도가 발전해 왔습니다. 한국에서도 가능할까요? 바람직할까요?

2. 더 자세한 채점 기준표와 더 많은 채점 예시, 우리 학교 내신의 신뢰를 높이는 데 무엇이 더 도움이 될까요?

3. 공정성의 다양한 측면을 고려할 때,
수행평가에서 학생의 AI 사용은 어떤 방식으로, 어느 정도까지 허용되어야 할까요?





감사합니다.

질문과 의견을 자유롭게 나눠 주세요