

AI 시대, 무엇을 어떻게 가르치고 평가할 것인가

김선 (Ph. D.)

이 슬라이드 디자인은 “클로드 디자인”을 활용하여 제작하였음.

여는 말

요즘 우리가 느끼는 그 불편함

한 아이가 낸 독서감상문의 문장이 유난히 매끄럽다. 평소 교실에서 관찰된 결과와 어딘가 다름.

“이거... 정말 네가 쓴 거니?”

— 그리고 뒤따르는 고민들 “어떻게 가려내지?” “숙제를 내주지 말아야 하나?”

01

질문을 바꾸다

‘AI를 어떻게 잡아낼까’에서 ‘평가를 어떻게 다시 설계할까’로

74%

해외 한 대학 연구에서, 자신이 AI를 썼다는 사실을 제대로 신고하지 않은 학생의 비율

작동하지 않는 규칙

신고란은 신고를 만들지 않는다

대부분은 “썼다고 하면 불이익을 받을까 봐” 두려웠다고 답함.

규칙이 지켜지는지 확인할 방법이 없으면, 손해는 성실한 아이에게 감.

출처. Gonsalves(2024), King's College London 사례. Corbin, Dawson & Liu(2025), “Talk is cheap”, *Assessment & Evaluation in Higher Education* 50(7), 1090에서 재인용

만든 사람들의 고백

‘AI를 어디까지 써도 되는지’를 5단계로 나눈 유명한 도구가 있음. 그 도구를 만든 연구자들이 개정판에서 이렇게 인정함.

AI는 탐지가 사실상 불가능하기 때문에, 우리가 나눈 단계 구분조차 실제로 지켜지는지 확인할 길이 없다.

문제는 “답이 안 나온다”가 아님. **질문이 틀렸다는 것임.**

“AI를 어떻게 잡아낼까”가 아니라, “평가를 어떻게 다시 설계할까.”

경찰의 일

끊어진 것을 감시로 되돌리려는 일

설계자의 일

애초에 감시가 필요 없는 구조를 새로
짓는 일

02

무엇이 끊겼나

평가의 원래 구조와, AI가 벌려 놓은 틈

2 · 평가란 원래 무엇이었나

평가는 발자국을 읽는 일

우리는 아이의 머릿속을 직접 들여다볼 수 없음.
볼 수 있는 건 **아이가 남긴 발자국**뿐임.

25년 전의 정의

평가는 ‘시험’이 아니라
‘증거를 보고 추론하는 일’이다.

받아쓰기 공책, 일기, 독서감상문, 수학 풀이, 발표 — 이 발자국을 보고 “이만큼 할 수 있구나”를 짐작하는 것

출처. Pellegrino, Chudowsky & Glaser(eds.), NRC(2001),
Knowing What Students Know: The Science and Design of Educational Assessment, National Academies Press

평가 삼각형

인지

나는 이 아이의 무엇을 알고 싶은가

사진으로 치면,
무엇을 찍을지 정하기

관찰

그걸 알려면 아이가 무엇을 하는 걸 봐야 하나

어디서 어떻게 찍을지 정하기

해석

본 것을 어떻게 읽어서 판단할 것인가

나온 사진을 어떻게 읽을지 정하기

세 가지가 따로 놀면 사진은 아무것도 말해 주지 못함.

100년을 버텨 온 조용한 믿음

“아이가 낸 결과물은 그 아이 머릿속의 믿을 만한 흔적이다.”

예전에도 형이 대신 써 준 숙제는 있었음. 다만 품이 들었고, 흔적이 남았고, 들킬 수 있었음. 그래서 믿음은 대체로 지켜짐.

생성형 AI는 이 믿음을 어긴 게 아니라 폐기함.

이제 결과물은 사람과 기계가 섞인 데서 나옴. 그 안에서 **아이 몫이 얼마인지**는, 결과물만 보서는 원리적으로 알 수 없음.

누구나 남의 발자국 도장을 찍을 수 있게 된 것임.

오늘의 핵심 ①

AI는 튼튼했던 것을 부순 게 아니라, 원래 허술했던 것을 드러냈다.

우리가 ‘글쓰기 점수’라 불러온 것은 실은 ‘글이라는 물건의 완성도’였지,
‘그 아이의 생각하는 힘’ 그 자체는 아니었음.

03

무엇을 가르칠 것인가

‘AI 사용법’이 아니라 ‘위임의 판단력’, 그리고 기초라는 전제

아이들이 길러야 할 네 가지 힘

1 AI가 내놓은 답을 받아들일지·고칠지·
버릴지 판단하는 힘

2 내 생각을 지키면서 AI와 함께 만들어
보는 힘

3 어떤 일을 나와 AI가 어떻게 나눌지
정하는 힘

4 더 나은 AI를 상상하고 제안하는 힘

이 중 세 번째는 기존 교육과정에 자리가 없던 능력임.

5 · 무엇을 가르칠 것인가

위임의 판단력

무엇을 기계에 맡기고 무엇을 내가 끝까지 붙들 것인가를 매번 판단하는 힘.

그리고 윤리는 따로 떼어 한 차시로 가르치는 것이 아니라, 모든 수업에 끼고 보는 안경이어야 함.

“이건 무슨 정보로 배운 걸까?” “이 말이 항상 맞을까?”

학생의 말로 옮기면

이건 내가 스스로 해야
하는 것,
이건 도구의 도움을 받아도
되는 것

04

계산기와 구구단

초등에서 가장 중요한 이야기

메타인지적 게으름

메타인지란 “내가 지금 이해했나, 어디서 막혔나”를 스스로 알아채는 힘임. AI가 답을 척척 주면 그 점검 습관 자체가 사라짐.

AI를 쓴 아이가 **더 좋은 결과물을 더 빨리** 낼 수는 있음. 그러나 그 성과가 **오래가는 실력 향상**으로 이어지지 않는음.

국제 문서들이 서로 다른 각도에서 위와 같은 같은 결론에 도달함.

오늘의 핵심 ②

AI를 잘 쓰려면 AI 없이 쌓은 실력이
있어야 한다.
그런데 AI를 쓰면 그 실력이 안 쌓인다.

구구단을 안 배운 아이는 계산기도 잘 쓰지 못함

초등은 바로 그 ‘구구단’ — 읽고 쓰는 기초, 수 감각, 스스로 생각을 밀고 나가는 근육 — 을 처음 몸에 심는 시기임. 이때 도구가 그 기회를 가로채면, 나중에 어떤 도구로도 메우기 어려움.

$$47 \times 23 = 10$$

“어? 이상한데?” 하고 알아채려면, 구구단이 몸에 배어 있어야 함.

계산기를 잘 쓴다는 건 답이 대충 맞는 크기인지를 느낄 줄 안다는 뜻임.

두 개의 길

길 하나 : 도구 없이

교실에서, 아이가 스스로 해 보는 길

부정행위를 막으려고 있는 것이 아님. 무언가를 맡기기 전에, 그 아이 안에 있어야 할 것이 정말 생겼는지 확인하려고 있는 것임.

길 둘 : 도구와 함께

도구의 도움을 받아도 되는 길

여기서는 결과물의 완성도가 아니라, 무엇을 받아들이고 무엇을 버렸는지가 평가의 대상이 됨.

기준은 “AI로 하기 쉬운 것”이 아니라, “**도구에 맡기기 전에 이 아이 안에 꼭 있어야 하는 것**”임.

05

교실에서 바로 쓰는 것들

수업의 네 가지 원리, 평가의 다섯 가지 원칙

구조를 바꾸는 네 가지 방법

01 비교하게 만들기

혼자 생각을 적은 뒤, AI의 아이디어와 나란히 놓고 무엇을 받아들이고 무엇을 버릴지 말하게 함.

03 기준을 만들게 하기

“좋은 독서감상문이란 무엇일까”를 함께 정하고, 그 기준으로 자기 글과 친구 글을 살펴보게 함.

02 바뀐 곳에 이유 붙이기

초고와 고친 글을 나란히 두고, “이건 내가 고쳤고 이건 도움을 받아 고쳤다”고 표시하게 함.

04 즉석에서 말하게 하기_AI 시대 판단력 기르기

교실 네 모서리에 ‘내가 스스로 / 도구와 함께 / 사람만이 / 잘 모름’ 팻말을 붙이고, 활동마다 자기 자리로 이동해 이유를 설명하게 함.

과제 카드

카드를 한 장씩 뽑아 친구와 이야기해 봅시다

★ 표시가 있는 카드는 함께 이야기할 카드

01

구구단 외우기

02



그림일기에
넣을 그림
그리기

03

넘어져 우는
친구 달래 주기

04

모르는 낱말 뜻
알아보기

05



두 자리 덧셈
답이 맞는지
확인하기

06



할머니께 드릴
편지에 쓸 말
정하기

07

받아쓰기 연습
하기

08

짝 바꿀 때 누구
옆에 앉을지

09

색칠할 때 무슨
색이 어울릴지



빈 카드에는 우리
반이 정한 과제를
적어요

네 가지의 공통점

“성찰하세요”라고 당부하는 대신,
성찰하지 않고는 과제를 끝낼 수 없는 구조를
만들.

초등 교실을 위한 다섯 가지 원칙

- 01 무엇을 재려는지부터 정한다** 목표가 흐릿한 평가는 어떤 규칙도, 어떤 프로그램도 구제하지 못함.
- 02 한 과제가 아니라 여러 접점을 본다** 선생님들이 이미 가장 잘하시는 일임.
- 03 결과물이 아니라 과정을 본다** 짧은 대화, 중간 확인, 친구와의 피드백을 평가 안에 넣음.
- 04 채점 기준을 아이와 함께 만든다** 채점을 편하게 하는 도구가 아니라, 성찰이 곧 과제가 되는 구조임.
- 05 공정성을 설계에 넣는다** 도구를 전제한 평가는 지갑의 차이를 성적의 차이로 옮겨 놓음.

8 · 어떻게 평가할 것인가

한 장의 사진이 아니라 한 학기의 앨범

사진 한 장은 조작할 수 있어도, 한 학기치 앨범 전체를 앞뒤 맞게 조작하기는 훨씬 어려움.

개별 과제 하나를 'AI로도 못 뚫는' 과제로 만들려는 시도는 실패하기 마련임.

3주 전 대화에서 이 개념을 설명하지 못
하던 아이가 오늘 그 개념을 능란하게 다
룬 글을 냈다면 —

그 불일치 자체가 어떤 탐지 프
로그램보다 믿을 만한 정보임.

매일 아이를 보고, 매일 대화하는 선생님만이 가
진 무기임.

8 · 어떻게 평가할 것인가

명시적 기준으로 스스로를 점검한 집단에서, 항상 효과는 원래 성적이 낮았던 아이에게 가장 크게 나타남.

AI 시대의 가장 큰 위험이 ‘메타인지적 게으름_스스로 점검하는 습관의 상실’이라면, 자기평가·동료평가는 가장 잘 검증된 처방임.

출처. White & Frederiksen(2000), ThinkerTools 반성적 평가 연구 — NRC(2001), *Knowing What Students Know*, Box 6-3(p.238–239)에서 재인용

AI가 없던 25년 전의 보고서

우리에게 필요한 것은 새로운 규칙이 아니라,
오래된 원리를 새 조건에서 다시 적용할
용기와 자원임.

출처. NRC(2001), *Knowing What Students Know*, 결론 및 권고

참 고 문 헌

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (eds.), National Research Council (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. National Academies Press.

Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A Framework for Ethical Integration of Generative AI in Educational Assessment. *JUTLP*, 21(6).

OECD (2026). *Navigating an Evolving Digital World: First Draft of the PISA 2029 Media and Artificial Intelligence Literacy (MAIL) Assessment Framework*.

Corbin, T., Dawson, P., & Liu, D. (2025). Talk is cheap: why structural assessment changes are needed for a time of GenAI. *Assessment & Evaluation in Higher Education*, 50(7), 1087–1097.

OECD / European Union (2026). *Empowering Learners for the Age of AI: An AI Literacy Framework for Primary and Secondary Education*. OECD Publishing, Paris.

White, B. Y., & Frederiksen, J. R. (2000). ThinkerTools 반성적 평가 연구 — NRC(2001) Box 6-3에 수록.

아이를 오래 지켜보고, 대화하고, 스스로 점검하도록
돕는 일 — 그것이 바로 그 오래된 원리입니다.

감사합니다.