

THE REALITY OF ASSESSMENT IN THE FACE OF AI, CHALLENGES AND OPPORTUNITIES

JET Conference-July 31, 2026

Daniel M. Settlage

Professor of Economics

College of Business and Industry

University of Arkansas – Fort Smith

Jim R. Wollscheid

Professor of Economics

College of Business and Industry

University of Arkansas – Fort Smith



UNIVERSITY of ARKANSAS
FORT SMITH®

OUTLINE



Motivation of
Research



Strategies That
Do Not Work



Strategies That
May Work



Strategies That
Do Work



Example Strategy
In Action



Conclusions

MOTIVATION OF RESEARCH

- The rapid proliferation of AI raises numerous pedagogical issues.
- Paramount among these is ability to conduct honest assessment of student learning outcomes in the face of AI.
- This is especially true for assessments conducted outside of a fully proctored environment.
 - Even proctored assessment environments are not fully immune from the effect of student AI usage.

MOTIVATION OF RESEARCH

- Authentic assessment of student learning is required by all major accrediting bodies (e.g. HLC) as well as specialized business school accrediting bodies (e.g. AACSB, ACBSP, etc.)
- Failure to conduct authentic assessment of student learning is a major impediment to accreditation efforts.
- In addition, assessment is the foundation of effective teaching.
- The landscape of assessment of student learning is lagging the development student use of AI tools.

RISE OF AI

- From a recent working paper that examines ChatGPT (Chatterji, 2025).
- ChatGPT:
 - Launched November 2022.
 - By July 2025: ChatGPT used weekly by more than 700 million users (9% of world pop.)
 - Processed more than 2.5 billion messages per day.
- 10% of messages were classified as requests for tutoring and/or teaching.
- Roughly half of all messages were sent by adults under 26.

WHY IS AI DIFFERENT?

- AI is substantially broader and more intellectual compared to prior examples. It requires little effort on the user part to ‘acquire information’.
- From a study by Bastani, et. al (2025):
 - Students who used AI to help with math did significantly *better* on their homework than students who did not use AI.
 - They performed 17% *worse* on the overall test (where AI was prohibited).
 - AI users *believed* that they learned just as much as non-AI users.

ECON CHEATING SCANDAL AT BROWN UNIVERSITY

- Dr. Roberto Serrano was teaching principles of econ and gave (for the first time ever) a take home midterm and final (Chronicle of Higher Education-July 6, 2026).
 - Typical in-person midterm exams were between 65%-80%.
 - Take home average was 96% with nearly half earning 100%.
 - Strong evidence of widespread Ai use was detected.
- Dr. Serrano announced that the final was in person and if the distribution of grades were similar to the midterm, the midterm would stand. Otherwise, it would not count.
 - Of the 86 students that took the midterm, only 59 took the final exam.
 - The final exam average was 48%.
 - 22 of the students that dropped scored a perfect on the midterm.

MOTIVATION OF RESEARCH

- Given that it is becoming more difficult to detect student AI usage and enforce AI policies, there is a need to design 'AI resistant' assessments of student learning.
- Some countermeasure that instructors employ actually push students to use *more* AI.
- Many countermeasures create a moral dilemma for students:
 - Do I 'cheat' and use AI, knowing that many other students use it?
 - Do I forego the use of AI knowing that my grade will suffer relative to my peers?

PURPOSE

- The purpose of this presentation is to explore AI resistant strategies that:
 - Enhance student learning.
 - Are easy to implement.
 - Are cost and time effective.
 - Allow genuine assessment of student learning for both the student and instructor.
 - Do not create an ethical dilemma for students.

STRATEGIES THAT DO NOT WORK

- Assessment strategies that tend to fail in the face of advances in AI:
 - Take home/online assignments.
 - Lockdown browsers.
 - ‘Unique’ question sets.
 - Timed tests with short time limits per question.
 - High stakes assignments (especially take home).
 - Assignments focused on answer accuracy and recall.
 - AI detection and plagiarism detection software.

STRATEGIES THAT MAY WORK

- Assessment strategies that may work (imperfectly) in the face of advances in AI:
 - Online proctoring.
 - AI proctoring.
 - Live proctoring.
 - Multiple small-stakes assignments.
 - Reflective assignments focusing on personalized details.
 - Handwritten assignments showing work.
 - Post exam verification.
 - Suspected cheating/AI usage must retake assignment.

STRATEGIES THAT DO WORK

- Assessment strategies that tend to work in the face of advances in AI:
 - In-class proctored assessments.
 - Live or recorded video oral presentations.
 - Assignments that focus on the process/reflection, rather than the outcome.
 - Post-assignment oral reflection.
 - Multiple step submission process-oriented assignments.
 - Outline, draft, final, critique/explanation of process.
- Assignments that remove incentive to cheat.
 - No point assignments (practice).
 - Completion-based (as opposed to performance-based grading) on out of class assignments.

THE BOTTOM LINE

- If AI can answer a question generically, students will be able to use it to create polished submissions.
- We cannot stop AI use outside of a fully proctored environment.
- Failure to acknowledge this reality leads to poor assessment and poor student learning.
- We can design assessment to be AI resistant by:
 - Excluding AI use through in person proctoring.
 - Focusing on things that are 'harder' for AI to do (oral presentations, process-oriented assignments, reflective assignments).
 - Removing the incentive for students to use AI.

EXAMPLE STRATEGY IN ACTION

- We examine the effect of employing one of these strategies in principles of microeconomics using multiple-choice quizzes and tests in an online setting.
- Control Group
 - Class is run normally, without anti-AI intervention for all three units/tests.
- Treatment Group
 - Class is run normally as a baseline control through test 1 material as a control.
 - After test 1, we announce (and reinforced every class period) that all remaining quizzes taken by a student will be bumped up to 100%.
 - Students are asked to **NOT** use any AI/outside material/apps on the quizzes.
 - Tests 2 and 3 are given as before, in an online proctored environment.

EXAMPLE STRATEGY IN ACTION

- For the treatment group, the first 4 quizzes = pre-intervention control.
 - Same testing conditions for both treatment and control groups.
- Last 10 quizzes for treatment group are given post-intervention.
- 80% would be considered a 'high' score based on prior test results.
 - Performance on first 4 quizzes is very similar between control and treatment groups.
- Treatment was group markedly different than control on the last 10 quizzes.
 - Far fewer high scores ➔ evidence that AI use stopped post-intervention.
 - Far fewer 0% scores ➔ evidence of more class engagement post-intervention.

Student Score on Quiz	<u>Control Group</u>		<u>Treatment Group</u>	
	First 4 Quizzes	Last 10 Quizzes	First 4 Quizzes	Last 10 Quizzes
≥ 80%	70.83%	60.64%	75.69%	37.03%
0%	6.41%	10.77%	4.73%	5.14%

EXAMPLE STRATEGY IN ACTION

Statistical test results-pre intervention

- Statistical evidence to suggest treatment group was using AI **more** than control group (2 of 4 quizzes had statistically significantly higher averages).
 - AI use more prevalent as AI advances and student use increases.
- No statistically significant difference between Test 1 scores for control and treatment group.
 - Baseline performance on exams for control vs. treated group was similar.

EXAMPLE STRATEGY IN ACTION

Statistical test results-post intervention

- All post-treatment quizzes were statistically significantly lower in the treatment group vs. the control group.
 - Statistical evidence that ***the intervention reduced AI use.***
- Post treatment test averages were statistically significantly higher for treatment group (nearly one letter grade higher).
 - Statistical evidence that ***the intervention increased student learning.***
- Correlation between quiz performance and test performance higher post-treatment (0.28 pre-treatment vs. 0.54 post-treatment).
 - Statistical evidence that ***the intervention led the quizzes to become more 'honest' assessments of student learning.***

CONCLUSIONS

- Our results reveal that student AI use is prevalent, difficult to detect, and hinders student learning.
- We propose several strategies that can be employed to mitigate the effect of AI use on student learning and assessment.
- As the power and adoption of AI rapidly advances, it creates an ever-evolving assessment landscape.
- This has the effect of forcing instructors to be on the forefront of AI developments and necessitates potential modifications of how traditional and longstanding assessment is conducted.

THANK YOU FOR YOUR TIME AND
PARTICIPATION!

QUESTIONS?



UNIVERSITY of ARKANSAS
FORT SMITH®

Daniel M. Settlage

Professor of Economics

College of Business and Industry

University of Arkansas – Fort Smith

Jim R. Wollscheid

Professor of Economics

College of Business and Industry

University of Arkansas – Fort Smith