

AI-Resilient Assessment: Evidence from 240 Classrooms

What happened to integrity flags, teacher marking time and student reasoning scores when 240 partner classrooms moved from written submissions to defence-based assessment.

240 classrooms · 3 countries · 2 academic terms · 2025 to 2026 · Published 2026-05-05 · Updated 2026-08-16 · CC BY 4.0

Key finding

Across 240 NASCA partner classrooms that replaced take-home written submissions with defence-based assessment, integrity flags fell from 23 percent of submissions to 4 percent, while measured reasoning scores rose 17 percent over two terms. Marking time per class fell by roughly a fifth once the rubric was in steady use.

The study covers two academic terms across primary, middle and senior classrooms in India, the UAE and the United States. The rubric used is published openly at </assessment/rubric>.

Headline numbers

23% → 4%

integrity flags per submission after redesign. Flags are teacher-raised, not detector-raised.

+17%

mean reasoning score across two terms. Scored against the published NASCA reasoning band.

-21%

teacher marking time per class. Measured from term two onward, once the rubric was familiar.

88%

of teachers would keep the redesigned format. End-of-study teacher poll, n = 268.

The data

Table 1. Before and after, by school stage

Integrity flags as a share of submissions. Reasoning score on a 0 to 4 band.

Stage	Classrooms	Flags before (%)	Flags after (%)	Reasoning before	Reasoning after
Primary (Grades 1-5)	74	11	3	1.9	2.3
Middle (Grades 6-8)	88	24	4	2.1	2.5
Senior (Grades 9-12)	78	33	6	2.4	2.8
All stages	240	23	4	2.1	2.5

Table 2. Which redesign moves did the most work

Teacher-ranked contribution to the change, top choice per classroom.

Redesign move	Share of classrooms ranking it first (%)
Live defence of the submitted work	34
Process evidence graded alongside the artefact	26
Task anchored to local, un-Googleable context	18
In-class checkpoint before submission	14
Peer critique round	8

What the data shows

Defence beats detection

Classrooms that added a short live defence saw the steepest fall in integrity flags. No AI-detection software was used anywhere in the study, and none was needed.

Redesign costs time once, then gives it back

Term one marking time rose slightly as teachers learned the rubric. From term two, marking time per class was about a fifth lower than the written-submission baseline.

Reasoning gains are largest where flags were worst

Senior classrooms started with the highest flag rate and posted the largest reasoning gain, suggesting the redesign converts avoidance behaviour into visible thinking.

Methodology

- Design: pre and post comparison across 240 classrooms in NASCA partner schools, no control group.

- Sites: India, the United Arab Emirates and the United States, across primary, middle and senior stages.
 - Period: two consecutive academic terms in the 2025 to 2026 year.
 - Measures: teacher-raised integrity flags per submission, reasoning score on the published NASCA 0 to 4 band, self-logged marking time, and an end-of-study teacher poll (n = 268).
 - Rubric: the open NASCA assessment rubric, published at /assessment/rubric, applied unchanged across all sites.
-

Limitations

- There is no control group, so term-on-term maturation cannot be fully separated from the redesign effect.
 - Integrity flags depend on teacher judgement and may be raised more readily once a school is discussing AI openly.
 - Marking time is self-logged.
-

Questions this report answers

Does AI-resilient assessment reduce cheating?

In the NASCA 240-classroom study, teacher-raised integrity flags fell from 23 percent of submissions to 4 percent after moving to defence-based assessment, without using any AI-detection software.

Does redesigning assessment cost teachers more time?

Only at first. Marking time rose slightly in term one while teachers learned the rubric, then settled about 21 percent below the written-submission baseline from term two onward.

What single change helps most?

A short live defence of the submitted work. 34 percent of classrooms ranked it the single biggest contributor, ahead of grading process evidence at 26 percent.

How to cite

NASCA Research Desk with the World STEM Federation (2026). AI-Resilient Assessment: Evidence from 240 Classrooms. NASCA. <https://www.nasca.edu.in/research/reports/ai-resilient-assessment-2026>

NASCA (North American STEAM & Coding Academy). Established 2017, headquartered in Mumbai, with offices in Delhi, Dubai and Orlando. Accredited by the World STEM Federation. info@nasca.edu.in · www.nasca.edu.in/research/reports