

Completion Is the Wrong Metric

Measuring transfer and performance in AI-enabled adult learning

In executive and professional education, completion rates, attendance, satisfaction scores, and short quizzes are operational signals. They can help leaders understand participation and learner experience. They do not reliably show whether adults can apply learning in consequential settings, make better decisions, or perform more effectively in the flow of work.

That weakness is not new, but AI makes it harder to ignore. When drafting, summarization, translation, and basic ideation can be accelerated by tools, surface indicators become even less trustworthy as evidence of meaningful learning. Institutions need a stronger measurement model, one that follows learning past content consumption and into role-based transfer, judgment, and performance.

Emma Sanchez, Ed.M.

Education Strategy and Learning Systems Advisor

CORE CLAIM

Completion is an activity metric, not an outcome.

In adult learning, value is created when people can apply, adapt, and sustain new practice.

WHY NOW

AI increases the gap between visible output and actual capability. Leaders need evidence that survives automation and assistance.

WHAT TO MEASURE

Transfer to role context, quality of decisions, observable performance change, and durability over time.

AUDIENCE

Executive education leaders, chief learning officers, program heads, academic leaders, and teams responsible for impact measurement.

ES

A PRACTICAL POSITION

AI-era learning systems need stronger evidence than completion, satisfaction, or content output.

The central question is no longer whether learners finished a course or produced an acceptable artifact. It is whether they can use new knowledge and tools to improve real work: solve cases more effectively, make better judgments, communicate more clearly, and perform with higher consistency under real conditions.

Leaders should treat measurement as part of learning design. If programs are built only to optimize participation metrics, they will miss the outcomes that actually justify investment.

THE SHIFT

The old question was, Did learners complete the experience? The better question is, What changed in practice, judgment, and performance after the learning experience, and how do we know?

STRATEGIC PREMISE

AI lowers friction for production. It does not automatically increase competence.

Programs create durable value when they measure what matters after the course ends: transfer into work, improved decisions, and sustained performance in context.

AI exposes how thin most learning metrics really are

Many organizations still judge learning programs through the signals that are easiest to collect: enrollments, completion rates, participation time, satisfaction surveys, and end-of-course quizzes. Those measures are convenient and operationally useful, but they are weak proxies for performance. They tell leaders that something happened. They do not tell leaders whether the learning experience changed what people can actually do.

In AI-enabled environments, this weakness becomes more serious. Learners can produce polished drafts more quickly, retrieve answers more efficiently, and generate apparently competent work with limited underlying understanding. That does not make AI a problem. It makes measurement design more important. If leaders keep relying on surface metrics, they risk mistaking assisted output for durable capability.

WHAT AI CHANGES

- Drafting quality becomes easier to fake as evidence of mastery.
- Short-answer correctness tells us less about reasoning and transfer.
- Time spent in a platform becomes a weaker signal of productive effort.
- Learner confidence can rise even when independent capability has not.
- Programs need measures that remain meaningful when assistance is available.

A BETTER VALUE QUESTION

What evidence would convince a skeptical manager, dean, or client that participants can now perform better in the settings that matter most?

COMMON METRIC	WHAT IT CAPTURES	WHY IT UNDERPERFORMS IN AI-ENABLED LEARNING	STRONGER EVIDENCE
Completion	Exposure to the program and persistence through required steps.	Finishing a course says little about whether learners can apply or adapt the material.	Role-based task performance and post-program application evidence.
Satisfaction	Learner reaction, perceived relevance, and delivery quality.	Useful experiences are not always rigorous, and pleasant experiences do not guarantee transfer.	Follow-up evidence of changed decisions, behaviors, or work quality.
Quiz scores	Recall, recognition, and limited conceptual understanding.	AI assistance and memorization both inflate performance on low-complexity tasks.	Scenario explanations, case judgments, and authentic performance tasks.
Time on platform	Activity volume and course navigation.	More time can indicate friction rather than progress, and less time may reflect tool use.	Quality of practice, task completion under standards, and manager-observed use.
Content output	Quantity of writing, slides, notes, or drafts generated.	AI makes output faster, but it does not guarantee discernment, accuracy, or fit.	Rubric-based review of reasoning, judgment, error detection, and contextual fit.

Bottom line: weak evidence has not become harmless because it is easy to collect. In the AI era, weak evidence becomes more misleading because assisted performance can look convincing before real transfer has occurred.

Measure the quality of transfer, not just the volume of activity

Adult learning matters when it changes performance in context. That means leaders need evidence that participants can use ideas, tools, and methods in situations that resemble real work. The goal is not to collect more data for its own sake. It is to gather the few forms of evidence that show whether learners can perform with greater independence, better judgment, and stronger consistency after the program.

That evidence should be role-relevant, interpretable by non-specialists, and timed across the learning journey. Immediate course reactions still have value, but they should sit at the beginning of the evidence chain, not at the end of it.

A USEFUL TEST

Can the program show that participants **used what they learned in their own role, improved the quality of a consequential task, and sustained that improvement beyond the course itself?** If not, the measurement model is still too shallow.

01

Demonstrated application

Can learners use the concept or method on a realistic case, simulation, or deliverable rather than simply describe it?

02

Judgment under ambiguity

Can learners explain tradeoffs, identify risks, and make sound decisions when the answer is not obvious or fully specified?

03

Transfer to role context

Can participants adapt the learning to their own work, tools, stakeholders, and constraints rather than repeat a classroom example?

04

Observable performance change

Do managers, peers, clients, or reviewers see improvement in quality, speed, consistency, communication, or decision quality?

05

Durability over time

Does the new practice persist after the course, or does performance decay once the structured environment disappears?

06

System conditions for use

Are there manager reinforcement, workflow supports, and norms that make transfer likely, or is the program asking individuals to sustain change alone?

The practical consequence: strong measurement in adult learning is multi-layered. It combines what learners can demonstrate now, what they do later in context, and what changes others can actually observe.

Build an evidence chain, not a single success score

Leaders often want one number that proves impact. That instinct is understandable, but adult learning rarely works that way. A better model follows the learner from participation through practice, transfer, and performance. Each layer answers a different question, and weaker layers should never be mistaken for stronger ones.

DO NOT LET WEAK SIGNALS STAND IN FOR STRONG ONES

- Completion is not transfer.
- Satisfaction is not capability.
- Fluent output is not judgment.
- Fast production is not better performance.
- Self-report is not enough when stakes are meaningful.

01

Participation

Did the learner start, attend, and complete the required experience? Useful, but only as a baseline.

02

Practice quality

Did the learner engage with meaningful cases, feedback cycles, and standards-aligned exercises during the program?

A USEFUL DESIGN RULE

For each claimed outcome, specify the strongest available evidence, who can observe it, and when it should appear. Then collect weaker proxy data only as supporting context, not as proof.

03

Application

Can the learner perform on authentic tasks that reveal reasoning, adaptation, and contextual fit?

04

Transfer

Has the learner used the new capability in actual work, with evidence from real outputs, observation, or self-documentation?

05

Performance

Is there visible change in work quality, decision quality, speed to proficiency, client outcomes, or reduction in preventable errors?

06

Durability

Does the improvement hold after coaching ends, workload increases, or new cases introduce complexity?

Design principles for AI-era measurement

Measure authentic work. Use cases, decisions, deliverables, and tasks that mirror the demands of the learner's role.

Keep evidence human-legible. Decision makers should be able to understand what changed without reading a technical dashboard.

Separate assistance from mastery. If AI support is allowed, define what independent judgment still needs to remain visible.

Use multiple time horizons. Collect immediate evidence, short-term transfer signals, and later indicators of sustained performance.

A 90-day agenda for redesigning the scorecard

The goal is not to measure everything. It is to replace the default scorecard with a small set of indicators that matter to sponsors, faculty, managers, and learners themselves. In most programs, that means choosing one or two authentic performance tasks, one post-program transfer check, and one observable work outcome that leaders can review with confidence.

STEP 1

Name the real outcome

Define what better performance actually means in the role, not just what the course covers.

STEP 2

Replace one weak metric

Swap a legacy proxy, such as completion alone, for one stronger task or transfer measure.

STEP 3

Design authentic evidence

Create a rubric, decision memo, simulation, observed practice, or work-sample review tied to real standards.

STEP 4

Add a transfer checkpoint

Check thirty to sixty days later for actual use in context, not just immediate post-course confidence.

STEP 5

Review and refine

Use the first cycle to improve the scorecard, reduce noise, and make the evidence easier to interpret.

What strong redesign looks like

After the first 90 days, leaders should be able to show a visibly different measurement model: fewer vanity metrics, clearer definitions of success, stronger performance tasks, and at least one credible follow-up measure tied to transfer or work quality. The scorecard should read less like a course operations report and more like an evidence summary about changed capability.

That shift also improves program design. When teams know what evidence matters at the end, they are more likely to build learning experiences that create that evidence in the first place.

A COMMON MISTAKE

Adding more dashboards without improving the underlying evidence. Better measurement does not come from visualizing the same weak proxies more elegantly. It comes from selecting stronger signals and designing the program around them.

Signals to track early

TASK QUALITY

Participants complete authentic assignments that reveal reasoning, not just polished output.

ROLE TRANSFER

Learners can point to where and how they used the capability in their own work.

EXTERNAL OBSERVATION

Managers, peers, faculty, or reviewers can verify a concrete change in practice.

SUSTAINED USE

The capability remains visible after the course instead of disappearing once the structure ends.

Leadership implication: stop asking learning teams to defend impact primarily with completion and satisfaction. Ask them to show how the program produces transfer, how that transfer is observed, and what evidence suggests performance improved in context.

Adult learning earns its value when practice changes

Completion, satisfaction, and platform analytics will remain part of learning operations. They can help teams monitor access, experience, and delivery quality. But in executive and professional education, those measures should not be mistaken for proof of impact. The real value of a program appears when participants make better decisions, perform more capably in their role, and sustain those gains over time.

AI raises the standard for measurement because it widens the distance between surface fluency and genuine capability. Leaders therefore need scorecards that are harder to game and easier to trust: authentic performance tasks, transfer checks in real context, external observation, and evidence of durability beyond the learning event.

The right question is not whether learners finished. It is whether the program changed what they can do when the work becomes real.

When institutions measure only what is easiest to collect, they understate the value of strong programs and overstate the value of weak ones. Better measurement helps both design and accountability because it forces learning systems to aim at real performance from the start.

ABOUT THE AUTHOR

Emma Sanchez, Ed.M.

Emma Sanchez is an education strategy and learning systems advisor with more than 20 years of experience designing high-impact learning for executive and professional audiences. She works across curriculum strategy, adult learning, instructional design, AI-enabled program architecture, and measurable outcomes.

She holds a Master of Education from Harvard and helps institutions build rigorous, scalable learning systems that connect strong educational design to meaningful organizational results.

FOCUS

Executive learning, curriculum strategy, and AI-enabled education systems.

EXPERIENCE

20+ years designing professional and executive learning.

SPECIALTY

Learning science, curriculum architecture, assessment, and responsible AI workflows.

SUGGESTED USE

This paper is intended as a leadership viewpoint for universities, executive education teams, corporate learning groups, and program leaders who want stronger evidence of impact in AI-era adult learning.

QUESTIONS FOR LEADERS

QUESTION 1

If completion and satisfaction disappeared from the scorecard, what evidence would remain that the program improved capability?

QUESTION 2

Where in the experience must learners demonstrate judgment, not just polished output supported by AI tools?

QUESTION 3

Who besides the learner can verify transfer in context, and when should that evidence appear?