ludora LAB



WHITE PAPER NO. 3 · NOVEMBER 2026

Measuring Soft Skills: From Attendance to Evidence

How to show that people handle difficult situations better after training, not just that they attended.

5-minute read

ludoralab.com

In short

1. **Most soft-skills training is measured by attendance and satisfaction.** Neither tells you whether anyone now handles a difficult client or a tense meeting better.

2. **Decisions in realistic situations are measurable.** Situational judgment tests, which ask people what they would do in work scenarios, predict job performance, and simulations add consequences to the same idea.

3. **A simple plan uses five kinds of evidence**: first-attempt decisions, a before-and-after test, skill levels, a manager check-in after 30 days, and one business measure.

.34

the average validity of situational judgment tests for predicting job performance, across 102 studies with 10,640 people

IN THIS PAPER

1. The attendance trap
2. Why soft skills seem hard to measure
3. Decisions are observable
4. Five kinds of evidence
5. A measurement plan in five steps
6. Cautions

1 The attendance trap

Ask an HR team how a leadership or communication program went, and the answer often has two numbers: how many people completed it, and how much they liked it. Both are easy to collect. Neither answers the question a CEO actually asks: are our people better at this now?

Donald Kirkpatrick's four levels, first published in the 1950s and still the most common evaluation model, separate four questions: did people like it (reaction), did they learn (learning), do they act differently (behavior), and did it make a difference to the business (results). Completion and satisfaction sit at the first level. The value sits at the third and fourth.

2 Why soft skills seem hard to measure

Technical skills have clear tests: the code runs or it doesn't, the reconciliation balances or it doesn't. Soft skills are judgment in context. A good answer to an angry client depends on the client, the history and the stakes. So many organizations conclude that soft skills can only be judged informally: "we know it when we see it."

The problem is that we rarely see it. Difficult conversations happen behind closed doors, and the manager who avoided one leaves no trace in any report. What we need is a way to see judgment in action, under controlled conditions, before and after training.

3 Decisions are observable

Psychologists have measured workplace judgment this way for decades. A situational judgment test describes a realistic work situation and asks what the person would do. In a meta-analysis by Michael McDaniel and colleagues, based on 102 validity coefficients and 10,640 people, these tests predicted job performance with an average validity of .34, a useful level for any workplace assessment. Tests built from an analysis of the actual job did better than generic ones.

> **.34**
>
> **average validity of situational judgment tests for job performance (McDaniel and colleagues, 2001)**

A later meta-analysis by Michael Christian, Bryan Edwards and James Bradley found that the tests most often measure leadership and interpersonal skills, and that those measuring teamwork and leadership had relatively high validities. Video-based tests tended to predict performance better than paper ones.

A branching simulation is a situational judgment test with consequences. Instead of choosing an answer and moving on, the learner sees what happens next, and has to decide again. That makes it both a practice tool and a source of evidence.

4 Five kinds of evidence

Evidence	What it shows	Kirkpatrick level
First-attempt decisions	What people do before they know the "right" answer: their real instinct	Learning
Before-and-after test	Whether judgment improved on new situations, not the ones practiced	Learning
Skill levels	Beginning, Developing or Proficient for each skill, from first attempts only	Learning
30-day manager check-in	Whether the manager sees the behavior at work, using a short checklist	Behavior
One business measure	A number chosen at the start, such as complaint resolution time or project delays	Results

First attempts matter most. If learners can retry until they get a perfect score, the score measures persistence, not judgment. Ludora Lab records the first decision in every exercise for evidence, and lets learners replay as much as they like for practice.

5 A measurement plan in five steps

- **Name the behavior.** "Better communication" is too broad. "Gives specific, timely feedback to team members" can be observed.
- **Choose the evidence before you start.** Decide which of the five kinds you will use, and who will collect each one.
- **Take a baseline.** A short before-test and the first-attempt decisions show where each person starts.
- **Check at 30 days.** Ask managers three or four concrete questions, based on the module toolkit, rather than a general rating.
- **Report honestly.** Show the spread, not just the average, and say what the data can and cannot show.

6 Cautions

- **Small groups, careful claims.** A pilot with 20 people shows direction, not proof. Say so.
- **Practice is not performance appraisal.** Simulation scores should guide development, never ratings or pay. People only practice honestly where mistakes are safe.
- **Keep reflections private.** Learners' written reflections help them learn. They should never be shown to the company.
- **Context still matters.** A good score in a simulation is evidence of judgment, not a guarantee. The 30-day check-in exists for this reason.

ludora LAB

See the evidence for your own team

In a four-week Ludora Lab pilot, your team dashboard shows first-attempt decisions, before-and-after scores and skill levels for everyone, and we help you choose one business measure. Start at ludoralab.com/companies.

ludoralab.com

hello@ludoralab.com

Sources

1. Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology*, 63(1), 83–117.
2. Kirkpatrick, J. D., & Kirkpatrick, W. K. (2016). *Kirkpatrick's Four Levels of Training Evaluation*. ATD Press.
3. McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology*, 86(4), 730–740.
4. McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. L. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology*, 60(1), 63–91.

About Ludora Lab. Ludora Lab makes short, bilingual workplace simulations for early and mid-career professionals in the Arab world. People practice real decisions, see the consequences and build skills that last, in Arabic or English.