

THE LEARNING PROFESSIONAL

THE LEARNING FORWARD JOURNAL

RESEARCH

Leveraging AI to Evaluate Professional Learning

ONLINE EXCLUSIVE

Thomas R. Guskey

February 2026



Citation: Guskey, T. R. (2026). Leveraging AI to evaluate professional learning: Online Exclusive. *The Learning Professional*, 47(1).
<https://learningforward.org/journal/learning-with-ai/leveraging-ai-to-evaluate-professional-learning-online-exclusive/>

Generative AI has profoundly influenced every phase of the professional learning process, including planning and design, implementation, and follow-up. Perhaps its greatest impact is in the evaluation of professional learning programs and activities. AI can simplify many aspects of professional learning evaluation, from developing participant satisfaction surveys to determining changes in classroom practice and assessing effects on student learning.

However, professional learning leaders' thoughtful reflection and critical analysis are still essential. Even as AI streamlines technical tasks, it heightens leaders' responsibility to ensure evaluation efforts are accurate, meaningful, and aligned with the intended purposes of professional learning.

Determining participants' learning

Consider a central component of professional learning evaluation: assessing participating educators' learning. This is considered Level 2 in the professional learning evaluation framework I developed (see Guskey, 2002, 2024, 2026). Effective leaders seek evidence at this level to determine whether meaningful learning occurred. Specifically, did the experience result in measurable changes in participants' knowledge, skills, attitudes, beliefs, or behaviors?

Evidence of participants' learning serves two essential purposes. First, it offers a primary indicator of effectiveness. Even self-directed or self-initiated professional learning should be guided by clearly articulated learning goals (Guskey, 2014). Second, understanding participants' learning provides the foundation for improving policies and practices.

AI can be helpful in gathering evidence of multiple types of educator learning, including cognitive and affective, and behavioral learning. As with any kind of assessment, the learning goals determine the appropriateness of any method and the validity of the resulting evidence. Online surveys aligned with specific learning objectives are the most commonly used tools. However, interviews, observations, personal learning logs, and reflective journals can also yield rich and informative evidence.

Evaluation surveys of educators' cognitive learning

The most efficient and economical method for gathering evidence of participants' cognitive learning is an online assessment or evaluation survey administered at the conclusion of the professional learning experience or shortly thereafter. To be meaningful, questions addressing cognitive learning must be closely aligned with the articulated learning goals and focused on the knowledge and skills participants were expected to acquire.

AI applications such as ChatGPT, Gemini, Claude, ChatSonic, Perplexity, and others can make survey development considerably easier, generating individual survey items or complete instruments based on clearly defined specifications. Among these tools, ChatGPT is currently the most widely used by educators, largely because of its versatility in tasks related to lesson planning, teaching, and assessment (Toppo, 2023; Youngstown State University, 2026).

For example, consider a professional learning activity designed to help participants understand the core elements of project-based learning and prepare for classroom implementation. To assess participants' mastery of the activity's intended cognitive learning goals, I entered the following prompt into ChatGPT:

Produce a survey with 10 questions to measure knowledge of the primary criteria for implementing project-based learning.

The following survey was immediately generated:

Sure! Here's a 10-question survey designed to assess knowledge of the primary criteria for implementing project-based learning (PBL). These questions focus on key elements such as student voice and choice, real-world relevance, sustained inquiry, reflection, critique, and public product – core components of high-quality PBL as outlined by frameworks like those from PBLWorks (Buck Institute for Education).

Survey: Knowledge of project-based learning (PBL) implementation

Instructions:

For each question, choose the best answer that reflects your understanding of effective PBL practices.

1. What is the primary goal of project-based learning?
2. To prepare students for standardized testing
3. To engage students in real-world, meaningful learning experiences
4. To ensure students follow teacher-led lectures
5. To focus on rote memorization of content
6. Which of the following best describes a “driving question” in PBL?
7. A question with a single correct answer
8. A question designed for weekly quizzes
9. An open-ended question that guides the entire project
10. A teacher-generated question that students must memorize

These items address the essential elements of project-based learning, and ChatGPT even identified a specific source from which those elements were drawn.

Despite its thoroughness, however, close examination reveals a critical limitation. The correct answers to nearly all the survey items are easy to identify and could likely be guessed by participants and nonparticipants alike. For example, in the first item, “What is the primary goal of project-based learning?” three of the four response options appear clearly implausible, making option B the obvious choice. A similar pattern appears in the second item, where option C stands out as unmistakably correct. This tendency for overly transparent answers characterizes nearly all the generated items. To investigate whether higher-quality, more discriminating items could be produced, I entered the same prompt into Gemini, Claude, ChatSonic, and Perplexity. Among these tools, Claude generated the following, more discerning response:

I'll create a survey to measure knowledge of the primary criteria for implementing project-based learning. This will help assess understanding of PBL fundamentals.

This survey covers key criteria for effective project-based learning including driving questions, student voice and choice, authenticity, reflection, critique and revision, public products, standards alignment, sustained inquiry, assessment approaches, and the teacher's role as facilitator.

The questions are designed to distinguish between surface-level understanding and deeper comprehension of PBL principles. You could use this assessment for teacher professional development, to identify areas for additional training, or to establish baseline knowledge before implementing PBL in your educational context.

Would you like me to explain the correct answers or provide more context for any specific criteria included in the survey?

Project-based learning knowledge assessment

Instructions:

This survey aims to assess your understanding of the primary criteria for implementing effective project-based learning. For each question, select the answer that best reflects your knowledge.

1. What is the most essential element of an authentic project-based learning experience?
2. Student choice and voice
3. A driving question or challenge
4. Technology integration
5. Assessment by outside experts
6. In project-based learning, student voice and choice primarily refer to:
7. Allowing students complete freedom with no guidance
8. Students selecting from teacher-created project options
9. Students having meaningful input into what they learn and how they demonstrate their learning
10. Students voting on classroom policies

The items in this survey assess the essential criteria of project-based learning, and correct answers are not as obvious. Furthermore, Claude offered “to explain the correct answers” and “provide more context for any specific criteria included in the survey.” When asked to explain the correct answer for Item 1, for example, it provided the following response:

Correct answer: B) A driving question or challenge.

Explanation: *A driving question or authentic challenge is the foundation of effective project-based learning (PBL). It frames the entire project, gives it purpose, and connects learning to real-world contexts. While student voice/choice (A) is important, it's the driving question that establishes the framework. Technology (C) is a tool, not an essential element, and external assessment (D) is beneficial but not foundational.*

Professional learning leaders could use this information to review items, make any necessary revisions, and guide participants in correcting identified learning gaps or misunderstandings. In addition, they could augment the survey with open-ended response items to assess possible unintended learning goals and gather suggestions for improvement, such as, “What did you learn that will have the biggest impact on your classroom practices?” and “What improvements would you suggest for future professional learning activities?”

This comparison of AI tools shows that ChatGPT and Claude generated notably different outputs. It underscores the value of using multiple AI platforms and thoughtfully comparing their results to determine which tool best meets the specific task at hand. It is also important to remember the databases AI systems rely on are continually evolving and expanding, which can cause a single tool to generate different responses over time. Consequently, entering the same prompt into ChatGPT or Claude today would likely yield a different result.

By simplifying development tasks, AI tools enable professional learning leaders to focus on critical issues related to ensuring the accuracy of what is measured and aligning survey items with program learning goals. Remember, however, that AI-generated evaluation surveys still require careful review, adaptation, and refinement to ensure quality and relevance. Once the final evaluation survey is complete, it can be uploaded into an online survey platform such as Google Forms, Jotform, or SurveyMonkey to distribute the survey efficiently and tally results.

Evaluation surveys of affective learning

AI can also be used to generate surveys to assess a wide range of affective outcomes, including teachers' beliefs about effective instruction, attitudes toward prior professional learning experiences, expectations for students' learning, concerns about change, enjoyment of teaching, or overall affect toward teaching. In many settings, a particularly important affective outcome is teachers' sense of efficacy in teaching and learning contexts (see Guskey, 2021; Tschannen-Moran et al., 1998). Teacher efficacy, which refers to "teachers' belief or conviction that they can positively influence students' learning, even those who may be considered difficult or unmotivated" (Guskey & Passaro, 1994, p. 628), has been found to be strongly associated with successful program implementation (Berman & McLaughlin, 1978). Recent reviews also show a consistently positive association between teacher efficacy and student achievement (Kim & Seo, 2018; Täschner et al., 2025), although the causal nature of this relationship has been questioned (Jerrim et al., 2023).

To evaluate participants' sense of teacher efficacy as an intended professional learning goal, I entered the following prompt into ChatGPT, Gemini, Claude, ChatSonic, and Perplexity:

Develop a survey with 15 questions to measure teacher efficacy.

Claude produced the following well-targeted response that I judged to be the best of the five:

I'd be happy to create a teacher efficacy survey for you. Teacher efficacy refers to teachers' beliefs in their ability to positively affect student learning outcomes. Here's a comprehensive 15-question survey:

This teacher efficacy survey is designed to measure teachers' beliefs about their capabilities across three key dimensions:

1. **Instructional strategies efficacy** (Questions 1–5)
 - Focuses on teachers' confidence in implementing effective teaching methods.
 - Addresses differentiation and assessment capabilities.

2. **Classroom management efficacy** (Questions 6–10)
 - Measures teachers’ beliefs about their ability to maintain order and productive learning environments.
 - Includes behavior management and establishing routines.
3. **Student engagement efficacy** (Questions 11–15)
 - Evaluates teachers’ perceived ability to motivate and connect with students.
 - Includes items on working with challenging students and families.

Teacher efficacy survey

Instructions:

Please indicate your level of agreement with each statement below using the following scale: 1 = *Strongly disagree*, 2 = *Disagree*, 3 = *Somewhat disagree*, 4 = *Somewhat agree*, 5 = *Agree*, 6 = *Strongly agree*.

Instructional strategies efficacy

1. I can implement alternative teaching strategies effectively when students are not understanding the material.
2. I am confident in my ability to gauge student comprehension of what I have taught.
- ...

Classroom management efficacy

6. I can effectively manage disruptive behavior in the classroom.
7. I am confident in my ability to establish routines that keep classroom activities running smoothly.
- ...

Student engagement efficacy

11. I can motivate students who show low interest in schoolwork.
12. I can help students value learning and believe they can succeed academically.
- ...

Additional comments

Please share any additional thoughts about your teaching efficacy:

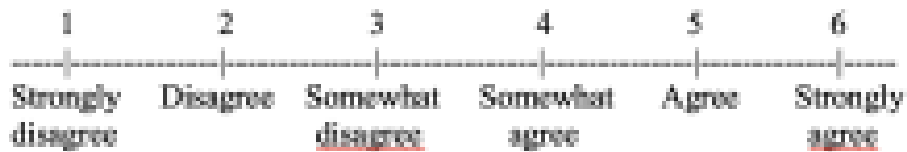
Thank you for completing this survey. Your responses will help inform professional learning and support systems.

This survey is well-organized and differentiates three key dimensions of teacher efficacy: instructional strategies, classroom management, and student engagement. Nevertheless, closer review suggests several revisions that would make the instrument easier to complete and more likely to yield accurate measures of teacher efficacy.

First, the directions would be clearer if presented in a simple visual format. For example:

Instructions:

Please indicate your level of agreement with each statement below using the scale below:



Second, all items in the current version are positively worded, most beginning with the phrase “I can.” When every statement is phrased positively, respondents tend to exhibit acquiescence bias, a tendency to agree with items regardless of their true beliefs (Costello & Roodenburg, 2015), which can distort results and limit the accuracy of interpretations.

Early scholars Rensis Likert (1932) and Jum C. Nunnally (1967) recommended including a balanced mix of positively and negatively worded items to reduce this bias. Balanced scales generally produce more valid scores than those composed entirely of positively phrased statements (Greenberger et al., 2003).

Accordingly, the survey would be strengthened by rewording one-third to one-half of the items to reflect a less efficacious perspective and be reverse scored. For instance, Item 2 might be revised to read, “I sometimes struggle to find ways to gauge student comprehension of what I have taught.” Similarly, Item 6 could become, “I have difficulty effectively managing disruptive behavior in the classroom.”

Some professional learning leaders administer teacher efficacy surveys at the beginning of a program or immediately after its completion and use the results to predict educators’ likely success in implementing new strategies or innovations. However, more meaningful measures of teacher efficacy are typically made after participants have implemented the strategies and received feedback on their impact (Guskey, 2020). Because experiences of success — what Albert Bandura (2001) termed “mastery experiences” — have the strongest influence on efficacy beliefs, measures collected after implementation and reflection on student outcomes provide a more accurate indication of professional learning’s impact on educators.

A powerful tool in the hands of responsible leaders

These examples demonstrate that generative AI is a powerful tool for enhancing the efficiency and quality of professional learning evaluations, but it is not flawless. Its outputs still require the careful, informed judgment of professional learning leaders. AI does not decrease the importance of professional judgment or diminish professional learning leaders’ responsibility; rather, it reshapes it.

Professional learning leaders must remain focused on identifying the evidence that best addresses their evaluation questions, determining the most efficient ways to collect that evidence, and interpreting it accurately and meaningfully when judging the merit and value of programs and

activities. While AI streamlines many aspects of development, leaders must continue to exercise critical judgment regarding the validity of the evidence gathered and its appropriate use in evaluation decisions.

References

- Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, 52, 1-26.
- Berman, P. & McLaughlin, M.W. (1978). *Federal programs supporting educational change. Vol. VIII: Implementing and sustaining innovations*. RAND. learnfwd.org/184cc1
- Costello, S. & Roodenburg, J. (2015). Acquiescence response bias – Yeasaying and higher education. *The Australian Educational and Developmental Psychologist*, 32(2), 105-119.
- Greenberger, E., Chen, C., Dmitrieva, J., & Farruggia, S.P. (2003). Item-wording and the dimensionality of the Rosenberg Self-Esteem Scale: Do they matter? *Personality and Individual Differences*, 35(6), 1241-1254.
- Guskey, T.R. (2002). Does it make a difference? Evaluating professional development. *Educational Leadership*, 59(6), 45-51.
- Guskey, T.R. (2014). Planning professional learning. *Educational Leadership*, 71(8), 10-16.
- Guskey, T.R. (2020). Flip the script on change: Experience shapes teachers' attitudes and beliefs. *The Learning Professional*, 41(2), 18-22.
- Guskey, T.R. (2021). The past and future of teacher efficacy. *Educational Leadership*, 79(3), 20-25.
- Guskey, T.R. (2024). Look beyond the satisfaction survey: A framework to evaluate results of professional learning. *The Learning Professional*, 45(1), 28-33.
- Guskey, T.R. (2026). *Evaluating professional learning* (2nd ed.). Corwin & Learning Forward.
- Guskey, T.R. & Passaro, P.D. (1994). Teacher efficacy: A study of construct dimensions. *American Educational Research Journal*, 31(3), 627-643.
- Jerrim, J., Sims, S., & Oliver, M. (2023). Teacher self-efficacy and pupil achievement: Much ado about nothing? International evidence from TIMSS. *Teachers and Teaching*, 29(2), 220–240.
- Kim, K.R. & Seo, E.H. (2018). The relationship between teacher efficacy and students' academic achievement: A meta-analysis. *Social Behavior and Personality: An International Journal*, 46(4), 529-540.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22(140), 5-55.

Nunnally, J.C. (1967). *Psychometric theory*. McGraw-Hill.

Täschner, J., Dicke, T., Reinhold, S., & Holzberger, D. (2025). “Yes, I can!” A systematic review and meta-analysis of intervention studies promoting teacher self-efficacy. *Review of Educational Research*, 95(1), 3-52.

Toppo, G. (2023). National ChatGPT survey: Teachers accepting AI into classrooms & workflow — Even more than students. *The 74*. learnfwd.org/88544d

Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W.K. (1998). Teacher efficacy: Its meaning and measure, *Review of Educational Research*, 68(2), 202-248.

Youngstown State University (2026) *The AI confidence gap among teachers*.learnfwd.org/124



Thomas R. Guskey, PhD, is Professor Emeritus in the College of Education, University of Kentucky. He is a longtime member of Learning Forward, best known for his work on teacher change and on planning, implementing, and evaluating effective professional learning. Contact him by email at guskey@uky.edu, on X at [@tguskey](https://twitter.com/tguskey), or at www.tguskey.com.