From Measurement to Action: A Learning Engineering Approach to AI-Powered Assessment for Human Power Skills Development

Yigal Rosen, PhD and Ilia Rushkin, PhD

Abstract

As AI reshapes work, human "power skills"—such as creative thinking, communication, collaboration, leadership, and AI fluency—are becoming essential to employability and performance. Yet these skills remain difficult to measure and develop at scale. Grounded in Learning Engineering, this work presents an AI-enabled assessment framework that integrates evidence-centered design, agentic AI, Bayesian proficiency estimation, and fairness-aware modeling. Building on prior large-scale assessments of creative thinking and collaboration, we describe the construct model, task design, scoring architecture, and early pilot results, demonstrating how learning engineering can translate learning science into scalable tools for talent development in the AI era.

Theoretical and Empirical Foundation

AI-Powered, Scenario-Based Assessment: Large-scale research shows that conversational agents and open-ended scenarios can validly measure complex social–cognitive skills like collaboration and creative thinking at scale (Rosen, 2015; OECD, 2017; Rosen et al, 2020).

Human skills as Performance, Not Self-Report: International assessments demonstrated that creative thinking can be assessed through generation, evaluation, and improvement of ideas in realistic contexts using rigorous scoring and psychometric models (OECD, 2024, Rosen et al., 2023).

Bayesian, Adaptive Measurement: Adaptive learning research established Bayesian proficiency estimation as a way to model evolving skill mastery with uncertainty, enabling scalable, evidence-based feedback and development (Rosen et al., 2017, 2018; von Davier et al., 2022).

Learning Engineering in Action: The Ignis AI PowerSkillsAssessment™ is developed through an iterative learning engineering cycle that connects research insights to real-world implementation—using pilot data to continuously refine task design, scoring, proficiency models, and deployment decisions to balance rigor, fairness, and usability at scale (Baker et al., 2022; Craig et al., 2025; Thai et al., 2023).

Instruments

Ignis AI PowerSkillsPrint™, a structured visualization designed to make human power skills visible, interpretable, and actionable for professionals and leaders based on their performance on Ignis AI PowerSkillsAssessment™. The central "talent flower" represents an individual's multi-dimensional performance profile across critical human power skills assessed. Each petal corresponds to a distinct skill domain, with its relative size and numerical proficiency estimate reflecting performance inferred from scenario-based, agentic AI interactions rather than self-report.

Method

Study I (Low-Stakes Pilot, n = 353): A diverse U.S. adult sample completed scenario-based tasks across seven power skills. Results showed strong reliability (α = .82–.91), a clear multidimensional factor structure, high item quality (96% acceptance), and meaningful performance differences in creative thinking and leadership behaviors.

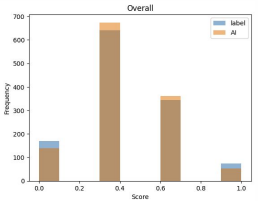
Study II (Applied Pilot, n = 36): Conducted with professionals and leaders in technology organizations, this study validated the assessment in a higher-stakes context. Reliability remained strong (avg. α = .85), factor structure was confirmed, and Bayesian proficiency models produced stable, interpretable estimates, with optimal precision for identifying emerging and high-potential leaders.

Results

Evaluation of AI model scoring assessments on a human-labeled dataset: We measure the mean square error of the score (on the 0-to-1 scale), and the Pearson correlation between the AI score and the label score.

Power Skill	Mean Absolute Error (MAE)	Pearson Correlation with Label	n
AI Fluency	0.082 (0.131)	0.737 (0.610)	285
Analytical Thinking	0.068 (0.179)	0.760 (0.325)	286
Collaboration	0.077 (0.124)	0.730 (0.620)	321
Communication	0.067 (0.225)	0.850 (0.427)	210
Leadership	0.103 (0.167)	0.718 (0.503)	120
Overall	**0.077** (0.160)	**0.774** (0.497)	**1229**

Values in parentheses show the results when the underlying AI model is used without fine-tuning, for comparison.



We can also compare the distributions of assessment responses by score, using the human-label scores ("label") and the scores from the AI model ("AI"). These histograms show that the distribution is not heavily skewed, and that difference between the label distribution and the AI score distribution is small.

Conclusions

Results indicate that AI-powered, performance-based assessment of human power skills is feasible, psychometrically sound, and practically useful for talent development. Ongoing work focuses on longitudinal change, predictive validity linking skills to real-world outcomes. From a learning engineering perspective, continued work is required to refine the feedback layer—how to transform posterior distributions and profile patterns into actionable, personalized recommendations for learners, coaches, and organizations.

2026 The Learning Engineering Research Network (LERN) Convening, Learning Engineering Institute, Arizona State University

yigal@ignisai.ai