Ignis AI PowerSkillsAssessment™

Advances in Science and AI Technology Enable Measurement of Human Power Skills

Yigal Rosen[1], Ph.D.
Co-Founder & Chief Product Officer

Ilia Rushkin, Ph.D.
Vice President of AI & Data Science

White Paper

Boston, MA
July 2026

[1] Correspondence concerning this white paper should be addressed to Yigal Rosen. Email: yigal@ignisai.ai

1. Executive Summary

Over the past decade, breakthroughs in cognitive science, psychometrics, and artificial intelligence have made it possible to measure and develop human power skills—the essential capabilities that define how people think, create, and collaborate in the age of intelligent technologies. This white paper provides an overview of the scientific foundation for the Ignis AI PowerSkillsAssessment™, a scalable, validated system for assessing creative thinking, communication, collaboration, leadership, analytical thinking, productivity, and AI fluency.

2. The "Blind Spot" in Talent Acquisition and Talent Development

Most organizations lack reliable ways to assess or quantify power skills—the human competencies that underpin creativity, collaboration, communication, leadership, and adaptability. Employers increasingly recognize that success in an AI-augmented economy depends on human strengths—leadership, empathy, adaptability, and analytical thinking—yet few have systems to identify and grow them at scale. This is an evident "blind spot" in workforce analytics.

The Future of Jobs Report 2025 surveys over 1,000 global employers (representing more than 14 million workers) across 22 industry clusters and 55 economies to explore how technological, economic, demographic, and environmental trends will transform jobs and skills through 2030 (World Economic Forum, 2025). The report signals that the future of work demands both technical fluency and robust human capabilities—skills that aid adaptability, collaboration, leadership, judgment, and resilience in a rapidly shifting environment. The report finds that 39% of workers' current skill sets will be disrupted or rendered obsolete by 2030. Employers expect a shift toward more cognitive, adaptive, and social skills—not just technical ones. Human-centered skills such as creative thinking, analytical thinking, resilience, flexibility, leadership, and social influence are emphasized as essential complements to technical proficiency. The Future of Jobs Report 2025 report's emphasis on creative thinking, analytical thinking, and leadership aligns directly with the skill domains measured by Ignis AI. Its recognition that technical skills alone are insufficient supports Ignis AI's mission: to provide visibility, measurement, and development of the human side of performance — the "power skills" that enable people to work with AI, not just alongside it. With nearly 40% of skills expected to shift, the demand for diagnostic tools and adaptive development programs (rather than static training) becomes critical. Ignis AI's approach of measuring and iteratively developing power skills is well-positioned to help organizations navigate that transition.

The 2026 ETS Human Progress Report emphasizes that durable soft skills matter more than ever, leading with an adaptability mindset (ETS, 2026). The report highlights that 77% of the 32,000 workers included in the survey believe job security requires continuous evolution, with three-fourths actively developing new skills to future-proof their careers. As automation, AI, and new forms of work redefine the workforce, ETS identifies a critical need for better tools to measure and develop human-centric capabilities—the very skills that enable individuals to thrive alongside intelligent systems. As AI becomes more central to work, workers want clearer ways to

validate and benchmark their skills. The report stresses that without visibility into these skills, companies struggle to align talent development with evolving job demands. The 2026 ETS Human Progress Report calls for science-based, psychometrically valid tools that move beyond traditional trait assessment and knowledge tests to capture applied human capabilities such as creative problem solving, teamwork, and communication effectiveness. These skills are as measurable and improvable as cognitive ones if grounded in rigorous design.

The findings align closely with the mission of Ignis AI, which provides a validated, data-driven framework to measure and develop creative thinking, communication, collaboration, leadership, productivity, AI productivity, and analytical thinking. The report's call for rigorous, scalable human-skills measurement tools directly supports the importance of platforms like Ignis AI in preparing individuals and organizations for success in the AI-driven future.

3. MIT EPOCH Framework

In a world increasingly shaped by artificial intelligence, a profound shift is underway—not only in what machines can do, but in what humans must do differently. A recent MIT report, The EPOCH of AI: Human–Machine Complementarities at Work, by Isabella Loaiza and Roberto Rigobón (2025), explores this transformation through a new lens. Rather than framing AI as a competitor to human capability, the report introduces the EPOCH framework, a model that measures and celebrates the uniquely human capacities that make collaboration with AI both effective and ethical. EPOCH stands for Empathy, Presence, Opinion, Creativity, and Hope—five pillars of human potential that define the frontiers of value creation in an age of intelligent machines. Drawing on data from U.S. labor markets and occupational databases, the researchers developed a way to quantify these qualities, showing that jobs requiring higher levels of empathy, ethical judgment, creativity, and leadership have grown consistently over the past decade. The work we are creating and the roles that are thriving are increasingly "human-intensive." The report demonstrates that occupations rooted in these capabilities not only resist automation but also expand as technology advances, signaling a clear economic shift: the more AI we integrate, the more we rely on human power skills.

The findings are reinforced by a companion piece from MIT Sloan Ideas Made to Matter (MIT Sloan School of Management, 2025), which argues that the most essential capabilities for the AI era are those that fill AI's blind spots. While machines excel at computation, scale, and precision, they fall short on empathy, moral reasoning, imagination, and human connection. The very qualities that AI cannot replicate—creativity, emotional intelligence, ethical judgment, and vision—are what allow humans to provide context, meaning, and direction in systems increasingly run by algorithms.

Together, these insights tell a clear story about the future of work. The frontier of human-AI collaboration lies not in coding better algorithms, but in cultivating better humans. Yet, as both the MIT report and article point out, organizations today lack the tools to identify, measure, and develop these essential human strengths. Traditional assessments capture knowledge and technical proficiency but fail to illuminate a person's creativity, communication style, leadership

potential, or collaborative intelligence. This visibility gap leaves individuals and organizations ill-prepared to harness the full potential of human-AI partnerships.

This is where the work of Ignis AI becomes particularly important. The EPOCH of AI report calls for psychometrically rigorous and scalable approaches to assess the very human power skills—creative thinking, collaboration, leadership, analytical reasoning, and AI productivity—that Ignis AI was built to measure and develop. This approach is not simply about protecting human work from automation; it is about amplifying human capabilities through science, data, and intentional development.

Ultimately, the MIT research reframes the narrative around AI. Rather than fearing what machines can take away, it invites us to imagine what they can make possible—when paired with human creativity, empathy, and leadership. The future of progress, it suggests, will belong to those who understand that intelligence is no longer a singular concept, but a partnership: one between human insight and artificial precision. To thrive in that future, we must learn to see, measure, and grow the capabilities that make us most human.

4. The Ignis AI Human Power Skills Framework: Measuring What Makes Us Human

In an era defined by the convergence of human creativity and artificial intelligence, success increasingly depends on a new kind of intelligence — one that blends cognitive insight with emotional depth, strategic communication, and ethical leadership. The Ignis AI Human Power Skills Framework (Rosen, & Rushkin, 2026a) captures this fusion by defining and measuring the seven essential capabilities that enable individuals and organizations to thrive in a world where human ingenuity and machine precision coexist.

At its core, the framework is built around **creative thinking**, the ability to imagine what does not yet exist and to transform ideas into actionable solutions (Rosen, & Rushkin, 2026b, Rosen, & Rushkin, 2026c). Creativity is not a solitary act but a disciplined process of exploration and synthesis — one that draws upon curiosity, flexibility, and courage to see beyond the obvious. Individuals at higher performance levels move from generating novel ideas to inspiring innovation in others, shaping environments where imagination becomes a shared language for progress.

Yet creativity alone cannot move ideas into the world without **communication** — the bridge between thought and action. Effective communication in this framework is defined by clarity, empathy, and adaptability. It is the ability to express ideas with purpose, listen deeply, and influence outcomes through shared understanding. Those who excel in this skill move beyond simply transmitting information; they create alignment, trust, and a sense of common mission within teams and across organizations.

From communication flows **collaboration**, the practice of turning collective intelligence into coordinated effort. In the Ignis framework, collaboration is not just teamwork — it is the art of combining diverse perspectives to create something stronger than any one contribution. As individuals grow in this domain, they evolve from cooperative participants to architects of

high-performing, inclusive teams that embody psychological safety, mutual respect, and shared accountability.

Leadership emerges naturally from this progression — not as a title or position, but as a capacity to motivate, guide, and empower others toward shared goals. Effective leaders in the Ignis model balance empathy with vision, direction with openness. They understand that leadership in the age of AI is as much about moral clarity and relational intelligence as it is about strategy and results. At the highest level, they cultivate purpose-driven cultures and nurture others' potential to lead.

Complementing these social and emotional dimensions are the **analytical** and **productivity** skills that transform creativity and collaboration into measurable results. Analytical thinking reflects one's ability to gather, interpret, and synthesize information, applying logic and evidence to solve complex problems. It is about making sense of uncertainty — identifying patterns, questioning assumptions, and arriving at insights that guide decision-making. Productivity extends this analytical capacity into execution, representing an individual's ability to organize, prioritize, and sustain focus to deliver meaningful outcomes. In Ignis AI's framework, productivity is not simply efficiency; it is intelligent action — the ability to convert knowledge, collaboration, and creativity into tangible progress.

Finally, **AI fluency** represents the newest dimension of human capability: the ability to enhance one's effectiveness through the responsible use of intelligent technologies. It combines digital fluency with ethical awareness, enabling individuals to use AI tools as amplifiers of creativity and precision rather than as substitutes for human judgment. Those who excel in AI productivity integrate these technologies seamlessly into their work, guiding others toward innovation while ensuring that human values remain at the center of every intelligent system.

Together, these seven domains form an integrated model of human potential — one that recognizes that the most valuable skills in the 21st-century workforce are not purely technical or purely human, but an intentional blend of both. The Performance Level Indicators (PLDs) embedded in the framework describe the developmental journey across four stages: from emerging awareness to transformative mastery. Each progression reflects growth not only in skill proficiency but in mindset — the evolution from performing tasks to shaping outcomes, from executing processes to inspiring purpose.

The Ignis AI Human Power Skills Framework thus offers more than a taxonomy of capabilities; it serves as a blueprint for thriving in the age of intelligent systems. By giving organizations and individuals visibility into their human power skills — and a way to develop them with scientific precision — it redefines what it means to be prepared, productive, and creative in the AI era. It places the human mind and heart at the center of innovation, reminding us that the future will be built not just by machines that think, but by humans who lead, imagine, and connect.

5. The Leadership Science Foundation Behind Ignis AI Power Skills Framework

Leaders today are asked to deliver results in environments defined by uncertainty, interdependence, and competing stakeholder demands. In practice, "leadership" is rarely a single trait or role-based authority; it is a repeatable process of diagnosing what matters, mobilizing people across boundaries, and executing principled action in systems that are simultaneously strategic, political, and cultural. Ignis AI's leadership measurement and development approach is built around this modern view. The Ignis AI 6-Skill Leadership Framework updates earlier, individual-centric leadership constructs by shifting toward a systems- and innovation-oriented perspective that reflects real organizational work—from early-career management through senior executive leadership. It is intentionally designed to (a) align with leadership science, (b) strengthen assessment and development, and (c) resonate with what corporate leadership teams actually need.

5.1 Ignis AI Definition of Leadership

Leadership is the ability to mobilize people and resources through self-awareness, clarity, and integrity to drive meaningful, coordinated action. It integrates self-awareness, purposeful communication, relational influence, systems thinking, innovation, and values-driven decision-making to build trust, align stakeholders, and enable collective progress.

This definition is intentionally operational: it describes what leaders do—how they interpret complex conditions, shape alignment, and move an organization from insight to implementation—rather than describing leadership as a personal identity. In this model, leadership effectiveness is visible in the quality of decisions under uncertainty, the reliability of cross-functional execution, and the ability to build shared movement without relying on positional authority alone. Because these behaviors are observable in language and judgment, Ignis AI can measure them through realistic, open-ended scenarios that elicit how a person actually thinks and acts when the context is ambiguous, politically charged, and high stakes.

Grounded in leadership science, the 6-Skill Leadership Assessment Framework is designed to provide leaders and organizations the definitions and operational guiding principles for the development, data analysis, reporting, and interpretation of Ignis AI PowerSkillsAssessment™ by focusing on systems- and innovation-oriented leadership talent development. The framework expands its value across all leadership levels—from early career to senior executives. In summary, this assessment framework is designed to:

- Align with advancements in leadership science and practices of principled and innovative leaders.
- Provide useful insights for leadership talent development and coaching.
- Offer skills that directly address senior management needs while providing value across management roles, levels, and industries.

5.2 The Six Leadership Skills Ignis AI Measures (and Why They Matter)

Ignis AI operationalizes leadership as six interlocking capabilities. The framework preserves the essence of earlier leadership constructs (reflection, communication, integrity) but adds critical dimensions often missing from traditional assessments—especially those tied to innovation, coalition building, and organizational effectiveness.

Self-Aware Leadership

Definition: The ability to understand one's strengths, blind spots, and impact; reflect on decisions; seek and act on feedback; and adapt one's behavior under changing conditions with humility and learning agility.

Scientific grounding and implication: Self-awareness and learning agility are foundational "enablers" of leadership because they govern whether leaders can update mental models when evidence changes. In complex work, leaders who can rapidly learn—without defensiveness—are better positioned to correct processes early, reduce escalation cycles, and build credibility through accountability. In Ignis AI's measurement logic, self-aware leadership is not scored as introspection in isolation; it is reflected in whether the leader notices impact, invites feedback (including difficult feedback), and converts insight into observable behavioral adjustment over time.

Strategic Communication

Definition: The ability to convey ideas clearly and compellingly, tailor messages to audiences, listen actively, and communicate vision and strategy in ways that mobilize and align others.

Scientific grounding and implication: Strategic communication is the "alignment engine" of leadership: it reduces ambiguity, creates shared meaning, and establishes expectations that enable coordinated action. The framework emphasizes not only sending messages but also active listening as a leadership tool—critical for surfacing weak signals, understanding resistance, and discovering what people actually need to execute. In measurement, Ignis AI looks for structured clarity (purpose, audience adaptation, rationale), the ability to lower defensiveness, and leadership language that turns disagreement into progress rather than blame.

Relational Leadership

Definition: The ability to build trust, collaborate across differences, navigate conflict constructively, and influence others without authority through authentic relationships.

Scientific grounding and implication: Real change requires social influence and coalition-building, especially when incentives differ across groups and formal authority is limited. MIT Sloan's "political lens" emphasizes that progress often depends on understanding interests, networks, and power—not only formal roles (Kellogg & Truelove, 2019). Ignis AI therefore treats relational leadership as a measurable skill: trust-building through reliability and fairness, the ability to

engage diverse perspectives, and constructive conflict navigation that preserves candor and psychological safety.

Systems Thinking & Problem Framing

Definition: The ability to diagnose root causes, understand interdependencies, connect short-term actions to long-term impacts, and frame problems in ways that lead to productive solutions.

Scientific grounding and implication: Many organizational failures are "solved" repeatedly because leaders respond to symptoms instead of system drivers. Leadership education emphasizes sensemaking—interpreting ambiguous situations to reveal leverage points (Ancona, 2012). Similarly, Dynamic Work Design highlights solving the right problem before moving to action (Repenning & Kieffer, 2025). Ignis AI measures whether a leader can map handoffs and interdependencies, identify incentive structures and decision rights, and frame a solvable problem statement that avoids premature solutioning while still enabling action.

Innovation & Adaptive Execution

Definition: The ability to generate creative ideas, test and refine solutions, adapt to changing conditions, and execute amid uncertainty to create meaningful progress.

Scientific grounding and implication: Innovation is treated here not as invention in a vacuum, but as disciplined learning under uncertainty. Repenning and Kieffer's Dynamic Work Design principles—especially "structure for discovery" and rapid learning loops—map directly onto Ignis AI's emphasis on hypothesis-driven experimentation, iteration, and adaptation in real operating constraints (Repenning & Kieffer, 2025). The creativity and innovation literature further supports that new value emerges from the interaction of expertise, intrinsic motivation, and a work environment that supports experimentation and progress (Amabile & Pratt, 2016). Ignis AI measures whether leaders can design low-risk, high-learning pilots (metrics + feedback loops), regulate churn, and convert experimentation into scalable operating practices.

Values-Driven Judgment

Definition: The ability to make principled decisions, act with fairness and transparency, uphold commitments, and build trust through alignment between values and behavior.

Scientific grounding and implication: Values-driven judgment is central to sustaining trust in high-stakes environments where short-term pressures compete with ethical obligations and long-term reputational risk. In the cultural lens, leaders shape meaning systems and norms through what they tolerate, reward, and communicate as "acceptable." Ignis AI measures whether leaders can articulate tradeoffs clearly, choose integrity under pressure, communicate with transparency, and create credible commitments that others believe and follow.

Ignis AI's 6-skill architecture explicitly aligns with major leadership science frameworks. First, it reflects Ancona's view of leadership as a distributed capability set—sensemaking, relating, visioning, and inventing—rather than a role (Ancona, 2012). Sensemaking aligns with Systems Thinking & Problem Framing, relating aligns with Relational Leadership, and visioning/inventing reinforces Strategic Communication and Innovation & Adaptive Execution. Second, it incorporates Repenning and Kieffer's Dynamic Work Design, which reframes leadership as designing the conditions for better work: solving the right problem, structuring discovery, and regulating flow (Repenning & Kieffer, 2025). These principles strengthen the "insights-to-implementation" bridge: the leader is evaluated not only on diagnosing and persuading but also on building learning cycles and operating rhythms that reduce friction, create clarity, and enable execution. Third, the framework internalizes a three-lens approach (strategic, political, cultural) by embedding those lenses within measurable skills: systems thinking captures structural and incentive-driven diagnosis (strategic), relational leadership captures influence and coalition-building (political), and values-driven judgment captures norms, meaning, and integrity under pressure (cultural) (Kellogg & Truelove, 2019; Schein & Schein, 2023). This matters for assessment because leadership breakdowns rarely originate in only one lens; reliable leaders can shift lenses and still mobilize coordinated action. Finally, the framework integrates innovation science via Amabile and Pratt's Dynamic Componential Model of Creativity and Innovation, which highlights that creativity and innovation are driven by expertise, motivation, creative-thinking skills, and the work environment (Amabile & Pratt, 2016). Ignis AI therefore treats innovation as a leadership behavior: fostering psychological safety, enabling progress through iterative experimentation, and creating conditions where ideas can be developed into outcomes.

6. Advances in Science and AI Technology Enable Measurement of Human Power Skills

Over the past decade, breakthroughs in cognitive science, psychometrics, and artificial intelligence have made it possible to measure and develop human power skills—the essential capabilities that define how people think, create, and collaborate in the age of intelligent technologies. This section highlights a series of research programs led by Yigal Rosen and Ilia Rushkin in collaboration with the OECD, Harvard University, Microsoft, The LEGO Foundation, and the U.S. National Science Foundation. Together, these initiatives established the scientific foundation for the Ignis AI PowerSkillsAssessment™, a scalable, validated system for assessing creativity, communication, collaboration, leadership, analytical thinking, productivity, and AI fluency (Rosen, et al, 2025). Each subsection below presents a milestone in this journey—demonstrating how rigorous research and applied innovation have converged to make human potential visible, measurable, and improvable.

6.1 OECD's First Global Assessment of Collaboration Skills through Conversational Agents Technology: The PISA 2015 Collaborative Problem Solving (CPS) Assessment, led by Rosen and Foltz (2014), pioneered the use of AI-driven conversational agents to measure complex social-cognitive skills at global scale. Implemented by the OECD, this innovation replaced traditional human-to-human collaboration tasks with human-to-agent (H-A) interactions, allowing consistent, equitable evaluation across more than 70 countries (OECD, 2017). Rosen's

studies demonstrated that simulated agents could authentically replicate human collaboration—sharing understanding, managing conflict, and coordinating group action—while removing interpersonal variability (Rosen, 2014, 2015a, 2015b). Test-takers collaborating with agents performed comparably to those with human partners, but produced richer, more reliable behavioral data. This human-to-agent framework marked a turning point in the science of human skills measurement, proving that "hard-to-measure" interpersonal competencies could be assessed rigorously and fairly at scale (Rosen, Ferrara, & Mosharraf, 2015). Further research demonstrated a wide range of stealth assessment features and innovative assessment designs to acquire authentic evidence of collaboration skills (Stoeffler, Way, & Rosen, 2022; Stoeffler, et al., 2020). The scientific foundation established here contributed to the development of novel assessment technologies of human power skills.

6.2 Measuring Power Skills in Civic and Experiential Learning at Harvard: At Harvard College's Center for Public Interest Careers (CPIC), Rosen, Turkay, and Eidelman (2016) designed one of the first frameworks to assess leadership, collaboration, communication, and adaptability in workforce internship and talent development contexts. The system integrated behavioral analytics, situational judgment tasks, and mentor evaluations, replacing subjective self-reports with observable performance data in support of impactful internship programs. Across multiple implementation cycles, it demonstrated that structured reflection and feedback linked to power-skill indicators enhanced Harvard student learning and long-term career readiness. This project provided an early model for psychometrically validated assessment in authentic settings, paving the way for scalable approaches for the assessment of human power skills.

6.3 Harvard and Microsoft Adaptive and AI-Driven Assessment: Further research led by Rosen and Rushkin at Harvard University's Office of the Vice Provost for Advances in Learning created new standards in adaptive assessment and AI-based proficiency estimation—core scientific underpinnings of Ignis AI's technology. Through the Adaptive Learning Open Source Initiative (ALOSI), developed with Microsoft Corporation and HarvardX, the team built an open adaptive engine using Bayesian knowledge tracing and real-time proficiency modeling. The system dynamically adjusted learning pathways, leading to significantly higher gains than static course structures (Rosen et al., 2017, 2018). This work introduced continuous, transparent skill-tracking—an innovation contributed to continuous proficiency estimation models, which quantifies complex human power skills such as creativity and collaboration.

6.4 Scaling AI-Powered Communication Assessment in Microsoft IT Support Program: In partnership with Microsoft Corporation, Rosen (2017) developed and deployed a conversational agent-powered assessment for communication and interpersonal effectiveness, reaching more than 100,000 global trainees at Microsoft. The system simulated customer interactions, analyzing clarity, empathy, and adaptability through natural language processing. It achieved a 20% higher predictive validity for job performance compared with traditional certification grades. This initiative proved that AI-based human skill assessment can forecast real-world performance—bridging the gap between education and workforce, providing the technical and psychometric foundation for enterprise applications of human power skills assessment.

6.5 PISA 2022 and The LEGO Foundation: Measuring and Tracking Creative Thinking: The PISA 2022 Creative Thinking Assessment, co-led by Rosen and Lansing-Stoeffler, was the first international study to measure creativity as a multidimensional cognitive process rather than a self-reported trait. Conducted across 64 countries, it established creativity as a valid, reliable, and comparable construct (OECD, 2024; Rosen, Stoeffler, & Simmering, 2020). Building on this achievement, Rosen, Rushkin, and colleagues (2023)—supported by The LEGO Foundation and BrainPOP—developed a dynamic proficiency estimation model for tracking creative thinking growth. This allowed real-time monitoring of learners' progress from idea generation to complex innovation. The findings marked a paradigm shift from static measurement to talent developmental analytics, directly informing the framework for continuous human power skills development.

6.6 Human–AI Collaboration in Medical Education: The five-part "What Can Artificial Intelligence Teach Us?" series, co-led by Rosen and Carlile (2023) at the Brigham Education Institute, examined how agentic AI can transform medical training by assessing communication, empathy, and teamwork in simulated clinical interactions. Rooted in Rosen's earlier research at Harvard and Microsoft and Carlile's innovative leadership in medical education, this work sparked new collaborations with Boston Children's Hospital and Brigham and Women's Hospital. These initiatives illustrate how AI-driven assessment can enhance both learning and evaluation of human power skills in high-stakes professional domains—expanding into healthcare providers and medical education.

6.7 Advancing the Science of Human Skills in Collaboration with the University of Cambridge: In March 2023, Rosen delivered an invited keynote at the University of Cambridge Digital Education Futures Initiative Annual Event exploring the evolution of creativity measurement and the role of AI in human capability development (Rosen, 2023). The keynote inspired new research collaborations with the University of Cambridge focused on advancing the science of AI-powered human skills assessment and modeling. These collaborations continue to inform Ignis AI's research agenda, extending its scientific framework to measure and nurture human creativity, collaboration, and reasoning within AI-enhanced environments.

6.8 Measuring Reasoning and Creative Thinking Skills at Scale: Rosen and Rushkin led pioneering work on developing innovative assessment and AI scoring for reasoning skills at BrainPOP (Rosen et al., 2023; Rushkin, & Rosen, 2024). Building on prior advances in adaptive assessment and dynamic proficiency modeling and with support from the U.S. National Science Foundation in 2024, their research introduced a new generation of agent-based and data-driven tasks designed to evaluate how learners generate and refine ideas through iterative exploration and AI feedback (Rosen, 2024; Rosen, Mosharraf, & Rushkin, 2025). The system combined cognitive modeling with AI analytics to infer creative thinking strategies in real time, capturing not just what learners knew, but how they applied their knowledge and skills to novel challenges.

Collectively, these advancements represent a new frontier in human capability science—demonstrating that creativity, communication, collaboration, and leadership are not

abstract traits but measurable, developable competencies. From international education and civic engagement to corporate training and medical education, these projects have forged a unified research-to-innovation pathway that culminates in the Ignis AI PowerSkillsAssessment™: a scientifically validated, patent pending AI-powered platform (Rosen, Kadar, & Rushkin, 2026) that makes the human side of intelligence visible, actionable, and transformative. Through these innovations, science and technology now converge to empower individuals and organizations to unlock their full potential in the age of intelligent systems.

7. Design of Ignis AI PowerSkillsAssessment™

Ignis assessments are built on an expansive and robust Ignis AI Power Skills Framework. The Skills Framework consists of skills and subskills that are defined at a fine-grained level and designed to align with real-world workplace tasks. Each skill in the Framework has a range of performance level descriptors that outline how a skill might develop over the course of a career through gained responsibility, experience, and job complexity. Ignis AI assessments focus on identifying the current level of skill for an individual at this point in their career journey. The assessments use a rich and complex range of task types to provide individuals with an opportunity to demonstrate their skills. Together, these seven skills form a comprehensive Framework for workforce readiness: analytical and creative cognition, interpersonal and leadership capacity, autonomous professional development, and fluency with AI-augmented productivity.

To collect insights about an individuals' proficiencies with these skills as they might relate to a job position, Ignis AI Assessment Framework defines these skills in a way that is both fine-grained and contextualized to the workforce and real-world tasks. The Ignis skills framework consists of 7 PowerSkills and their subskills, each Power Skill in the Ignis AI Power Skills Framework has a range of Performance Level Descriptors (PLDs) that provide progressive, cumulative, fine-grained descriptions of how a skill might develop over the course of a career through gained responsibility, experience, and job complexity. Insights gained from the assessment, therefore, support not only an understanding of an individual's current proficiency with a skill, but also provide insights into how the skill can be further developed.

7.1 Assessment Design for Ignis AI PowerSkillsAssessment™

Ignis assessments focus on identifying the current level of skill for an individual in their career journey. The assessments use a complex range of task types to provide individuals with an opportunity to demonstrate both their interpersonal and intrapersonal skills. Due to the value of these skills for success in the workforce, assessment of these skills should be held to a high standard of reliability and validity without introducing irrelevant variance or bias that technology might cause (International Test Commission and Association of Test Publishers, 2025)

Ignis assessments use an innovative approach to situational judgement tasks, multiple choice, and constructed response questions by using conversation-like chatbot functionality, leveled responses, and rubrics validated by subject matter experts. Rather than a response being evaluated as correct, partially correct, or incorrect, Ignis assessments allow individuals to demonstrate their level of skill along a progressive continuum. Ignis assessments achieve this

through the use of ordered multiple-choice responses and leveled rubrics. Ordered multiple-choice (OMC) task types explicitly arrange the answer options along a continuum of quality or sophistication. Each option corresponds to a particular level of quality or sophistication, making it suitable for assessing developmental progressions in skills (e.g., creative thinking or collaboration). Similarly, constructed (open) response task types explicitly evaluate answer responses along a continuum of quality or sophistication. The continuum of quality or sophistication is outlined by the Ignis AI Power Skills Framework which represents ordered, cumulative, progressions of understanding and ability that are based on thresholds between categories (e.g., "at least this level" of quality), where categories are seen as successive intervals of achievement, which through the Framework are aligned with different levels of responsibility, or complexity of a job.

Ignis AI assessments go beyond situational judgement tasks, multiple choice, and constructed response questions by embedding multi-dimensional features to acquire authentic evidence of skill. These features allow the collection of information about skills while mitigating social desirability bias - the tendency for individuals to answer questions in assessments in a way that makes them look good or appear favorable, rather than truthfully reflecting their actual behaviors or thinking processes. This allows for a deeper and more authentic collection of evidence about an individuals' Power Skills.

Real-world contexts

Conversational scenarios are designed to reflect real-world contexts. The scenarios replicate the dynamic ways knowledge and information are shared in modern workplaces. Individuals may be asked to interpret and respond to workplace scenarios presented through multimedia elements such as video clips featuring individuals having discussions in virtual meetings, virtual one-on-one meetings with a manager, or interactive meetings transcripts. Other scenarios may include reviewing workplace graphics, summaries from project reports, or internal office emails that contain nuanced interpersonal or informational cues. By embedding these diverse information sources within the assessment, Ignis items engage individuals in cognitively and socially authentic tasks that reflect the complexity of communication, decision-making, and collaboration in real work settings. This design not only increases ecological validity but also enables the assessment to capture a more genuine expression of the skill being evaluated.

Ordered multiple-choice options for efficiency and reflection.

Some of Ignis items are Ordered multiple-choice (OMC) items that provide a structured yet developmental way for individuals to reflect on and demonstrate their current level of skill. Each response option is carefully calibrated along a continuum of quality or sophistication, aligned to the developmental progressions of the Ignis AI Power Skills Framework. Rather than focusing on a single "correct" answer, these options reflect how people with different levels of understanding or proficiency might realistically respond to a situation. This approach allows for efficient scoring while still supporting reflection Because each option represents a meaningful level of

performance, OMC tasks yield interpretable, developmentally sensitive data that can inform growth and learning, not merely judgment.

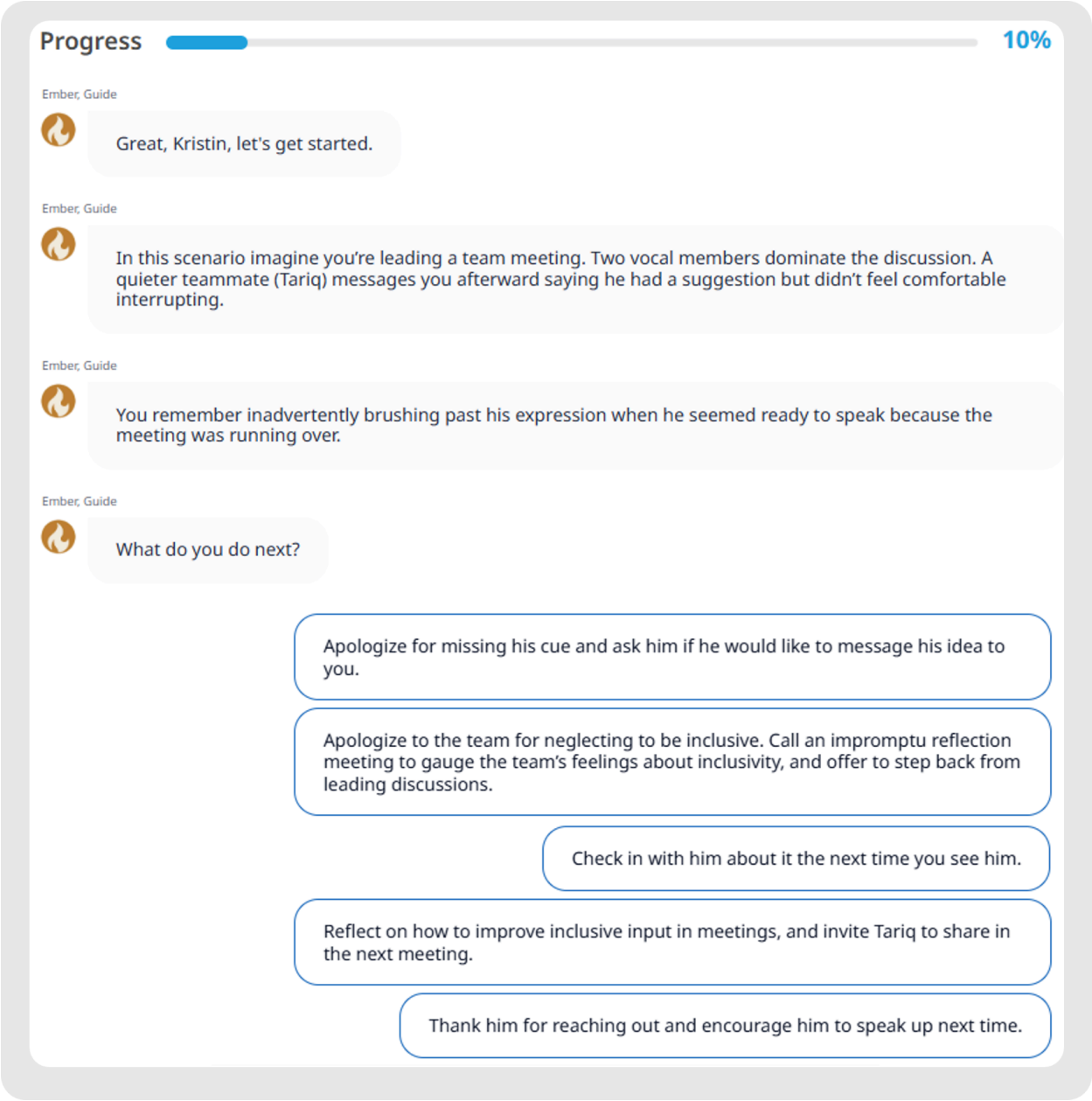

Figure 1. Sample leadership OMC item

Constructed response options for authentic demonstration of skill.

Constructed response (open-ended) items invite individuals to articulate their own thinking, strategies, and reasoning in context. This task type captures both what a person knows and how

they apply that knowledge—providing a more holistic understanding of the underlying skill. Individuals may be asked to draft an email response, outline an action plan, or explain how they would address a challenge raised in a meeting. Such responses are evaluated using leveled rubrics developed and validated by subject matter experts, ensuring that each level reflects a meaningful progression in skill sophistication. This open format supports the observation of subtle indicators of ability—such as empathy, strategic thinking, or adaptability—that are often missed in traditional multiple-choice formats.

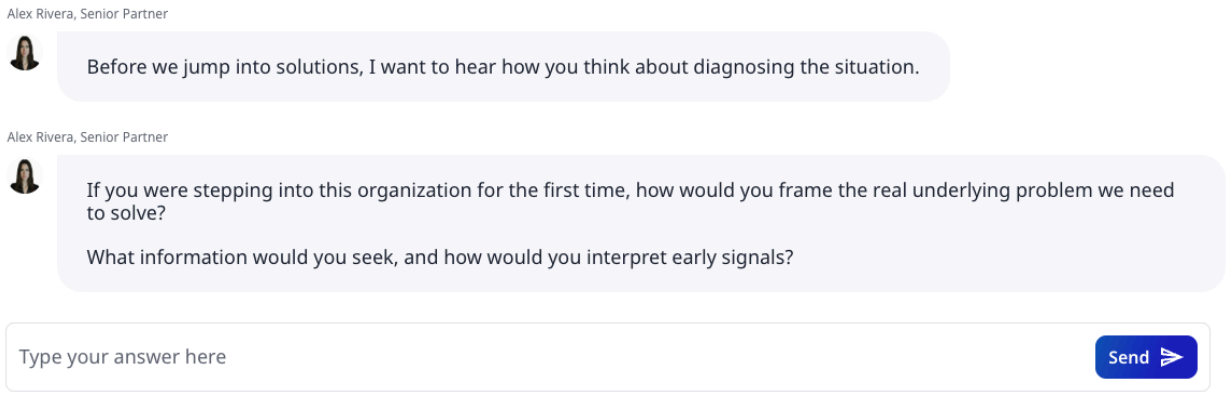

Figure 2. Sample Leadership CR item

Together, these innovations allow Ignis assessments to measure skills in ways that are authentic, efficient, and developmentally informative—capturing how individuals think, communicate, and act in the complex social environments of modern work.

Beyond task design, accurate and efficient scoring and reporting of assessments is critical for providing timely and useful insights. Ignis assessments leverage AI to scale the measurement of social and cognitive skills in a way that was not previously possible. Having assessment and subject matter experts in the loop ensure that the insights from Ignis assessments are expertly designed and expertly reviewed to ensure accuracy. Experts play a critical role in training AI models that score assessments, particularly those involving complex human skills such as communication, creative thinking, and collaboration. Their input ensures that the model's judgments are accurate, reliable, and aligned with human standards of performance. Ignis experts are responsible for training and validating the AI models as well as the ongoing calibration and continuous model improvement to ensure consistency, quality, and alignment.

7.2 How Ignis AI Measures Leadership Skills

A core challenge in leadership assessment is that self-report instruments often measure identity claims ("I am collaborative") more than real leadership behavior. Ignis AI addresses this by grounding measurement in (a) explicit Performance Level Descriptors (PLDs) and (b) structured, conversational scenarios that force tradeoffs, prioritization, and decision reasoning—conditions under which leadership shows itself.

At the integrated level, Ignis AI defines leadership proficiency across levels 0–3. At PLD 3, the individual exemplifies leadership excellence through deep self-awareness, strategic clarity, relational influence, and values-driven execution. They proactively seek diverse feedback, model intellectual humility, and continuously refine their leadership practice. Their communication is inspiring, precise, and adaptive—connecting vision to strategy while elevating voices around them through exceptional listening and inquiry. They excel at sensemaking: reframing complex, ambiguous challenges, identifying systemic drivers, and generating innovative pathways forward. They build strong, trust-based relationships across stakeholder groups, skillfully aligning interests, navigating political dynamics, and mobilizing others toward shared goals. Through authenticity, transparency, and principled consistency, they create psychological safety, accelerate collective learning, and foster a culture of purpose, accountability, and innovation.

At PLD 1, otherwise the individual shows emerging leadership capabilities but applies them unevenly. They demonstrate some self-awareness and occasionally seek feedback, though reflection is limited and behavior change is inconsistent. Communication shows effort but may lack clarity, strategic framing, or adaptation to different audiences. They can identify straightforward problems but struggle to see systemic root causes or propose innovative approaches. Leadership contributions somewhat move work forward but do not reliably create alignment or sustained momentum.

This structure supports both diagnostic measurement and actionable coaching, because it identifies which capabilities drive outcomes and which patterns limit effectiveness. Because the framework is designed around observable behaviors, it naturally supports development pathways. This is also why the framework scales across levels. Early-career professionals practice these skills in project contexts (e.g., influencing without authority, communicating across functions). Mid-level leaders apply them to cross-team operating systems (handoffs, incentives, escalation paths). Senior leaders apply them to enterprise transformation (portfolio choices, cultural norms, political coalitions, and ethical tradeoffs). The same six capacities are present—the context expands.

7.3 How Ignis AI Measures Communication Skills

Communication is the ability to acquire, structure, and convey information in ways that create clarity, trust, and coordinated action. In the Ignis AI framework, communication is not simply transmission; it is an intentional leadership process of listening deeply, making meaning, adapting to audience and context, and mobilizing shared understanding.

In the Leadership Edition of Ignis AI PowerSkillsAssessment™ , communication is evaluated in executive-level contexts where stakes are high and ambiguity is real: strategic drift, cross-functional conflict, customer risk, cultural tension, and rapid change. The focus is on clarity, empathy, and adaptability as observable drivers of trust and alignment.

At the integrated level, our Leadership Edition defines communication proficiency across levels 0–3. At PLD 3, communication reflects exceptional clarity, adaptability, and ethical influence. The

individual demonstrates deep listening that surfaces hidden tensions, political realities, and weak signals early. They synthesize complexity into compelling narratives that connect vision, rationale, principles, and action, and they do so with calm presence under scrutiny. In conflict, they transform threat into collective problem-solving, elevate others' voices, and build durable alignment. Their communication accelerates trust, decision quality, and execution across diverse stakeholder groups.

At PLD 1, otherwise communication shows effort and occasional clarity but lacks consistency or strategic depth. The individual listens but misses weak signals, political dynamics, or the emotional meaning behind objections. Messages are generally understandable yet may be under-structured, insufficiently tailored, or overly verbose. In tense moments, communication can drift into defensiveness, premature problem-solving, or avoidance of hard truths. Alignment is sometimes achieved, but shared ownership and sustained momentum are not reliably created.

7.4 Scoring Models

Ignis AI scoring models ensure that responses reflecting increasingly sophisticated reasoning or skill are assigned correspondingly higher scores. This approach enables Ignis assessments to capture fine-grained distinctions in performance, supporting reliable and interpretable measures of growth and proficiency across complex, human-centered skills. Just as the assessment tasks are structured to capture nuanced demonstrations of skill, the scoring models are designed to reflect the developmental nature of performance across progressive levels of understanding and ability. Rather than treating responses as simply correct or incorrect, Ignis employs psychometric models that preserve and interpret the ordered nature of task responses to yield precise and interpretable measures of skill proficiency.

Ordered multiple-choice (OMC) task types explicitly arrange the answer options along a continuum of quality or sophistication. Each option corresponds to a particular level of understanding, making OMC items ideal for assessing developmental progressions in complex skills such as creative thinking, communication, or collaboration. Similarly, Constructed Response (CR) assessment items evaluate written or produced responses along the same continuum of quality or sophistication, with each score point representing an ascending level of demonstrated ability. We use a GenAI-powered method to score responses to CR items. We distinguish several sub-types of items based on their scoring method. Generally speaking, an item may be tagged with several skills, but scoring is performed independently for each skill, so that the scoring method applies to a single skill and, consequently, is a property not of the item, but of the item-skill pairing.

The rubric-based scoring method uses scoring rubrics produced by a human expert. The rubrics describe in plain language the criteria by which a certain numeric score should be assigned. The rubrics are the fundamental element of the scoring method, but not the only one. We also make use of the skill description, look up similar scored and human-checked responses in historical data and insert them as instructive examples (a RAG pipeline), and make use of a fine-tuned GenAI model. In the interest of lowering variance, we make it possible to use several scoring agents in

parallel and aggregate their decisions (if there are any disagreements, it may give rise to fractional score values not present in the scoring rubrics, and it may flag this scoring event in particular for the next human-in-the-loop review). Each score is accompanied by a rationale, which simplifies the human-in-the-loop review process, used to grow the pool of labeled data.

The topic-based scoring method applies to a specific type of questions where the test-taker is asked to write one or several concepts, which are then classified according to a prepared list of topics. The topics and their prevalences are compiled from training data either by a human expert or by application of a topic model, and the remainder of the scoring process is a deterministic mathematical calculation that uses the topic classification outputs (class probabilities or simply class labels, depending on the implementation).

Based on a preliminary labeled dataset, and without the RAG loop that will assist the model scoring by providing relevant scored examples, the model has achieved the mean absolute error of 0.095 on predicted scores (defined in the range [0, 1]). The Pearson correlation between the predicted scores and labels was >0.75, and the comparison of histograms of AI scores and labels is below, showing a good agreement. As the histogram shows, the generated scores take only a discrete set of values, same as in the labeled dataset. This reflects the model using the same scoring framework as humans (in a more general case, the model may generate intermediate values as well). This allows us to determine the accuracy and, as a measure of agreement corrected for chance agreement, the quadratic weighted kappa (QWK): both achieve the values >0.77.

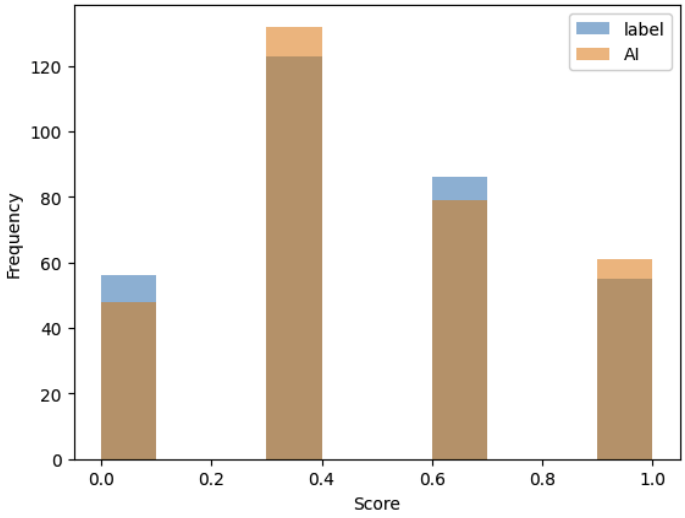

Figure 3. The comparison of histograms of AI scores and labels across all power skills

7.5 Assessment creation with AI

We can generate assessment content with generative AI, based on the flexible specifications (such as the industry, the functional area of an employee, the seniority level). The generation process is informed and controlled by our skill framework, including not only the data specific to

the particular skill but also a body of knowledge about skills and skill measurement that we maintain. Upon the initial generation of several alternative versions of an assessment activity, each item in them is validated for its properties, such as the tight alignment to the required skills, the level of linguistic complexity, etc. The items are generally multi-tagged, meaning that an item is relevant for several skills at once, one of which is the primary one, and the rest of which have smaller weights. Use of instructive examples of previously generated items that passed human review, further constrains the generation process and ensures that the psychometric properties of the new items will remain consistently close to those of the example items.

7.6 2026 Study III: Psychometric properties of Ignis AI PowerSkillsAssessment™

Psychometric properties of the assessment are examined to validate both the AI scoring system and the assessment content, creation of which is becoming more and more AI-assisted. We pay particular attention to fairness and biases. Choosing a respondent grouping variable (such as ethnicity), and selecting an item set developed for the same Power Skill, we set up our Differential Item Functioning (DIF) analysis by modeling the score of a respondent on an item as *score ~ rest_score + group + rest_score:group*. We use a fractional-binomial generalized linear model, separate for each item in the item set. The rest score is defined as the sum of the user's scores on all items except the one being modeled. The fit coefficient for the term group represents *uniform DIF* (a potential bias of the item for or against an entire group, regardless of their overall performance in the item set) , and the coefficient for the interaction term represents the non-uniform DIF (a potential bias of the item that depends on the respondent performance, but also on their group). An item is marked if any of these coefficients is statistically significant, based on a p-value threshold, which we set to 0.05.[2] The model is further used to predict the score using the group-averages of rest scores, so that the differences between predictions represent the gaps among groups. An item is marked if the maximum group gap exceeds a threshold. An item that received one of the marks is flagged for human review but not removal: either statistical evidence is weak but implied consequences are large, or vice versa. An item that received both marks is flagged for removal. The item-level results are then aggregated to the level of the item set.

In a recent study (Study III) conducted in March-April 2026 with a balanced diverse group of respondents recruited through the Prolific research platform (n=613) who took the full battery of assessment activities that we generated using our hybrid AI+Human SME (Subject Matter Expert) system (54 items in total). The item sets in the analysis were defined by the item's primary association with one of our 7 Power Skills. The DIF analysis was based on **sex/gender (Male vs. Female), ethnicity (Asian/Black/White), age group (18-40 yo vs. 41-69 yo), employment status (binary) or student status (binary)**. Overall, **no significant difference was found among any groups.**

Furthermore, a 2PL-IRT analysis is applied to estimate the item difficulty and discrimination power. In the same study, no obvious outliers were found, and the item difficulty was found to

[2] We would like to remind here that, perhaps counter-intuitively, a higher p-value means a more stringent DIF test. It means that the test will flag even a DIF with weaker evidence.

cover a wide range (approx. 0.2-0.6 on the 0-1 scale) and the discrimination was approx. 0.7-0.8 on the 0-1 scale (see Figure 4) .

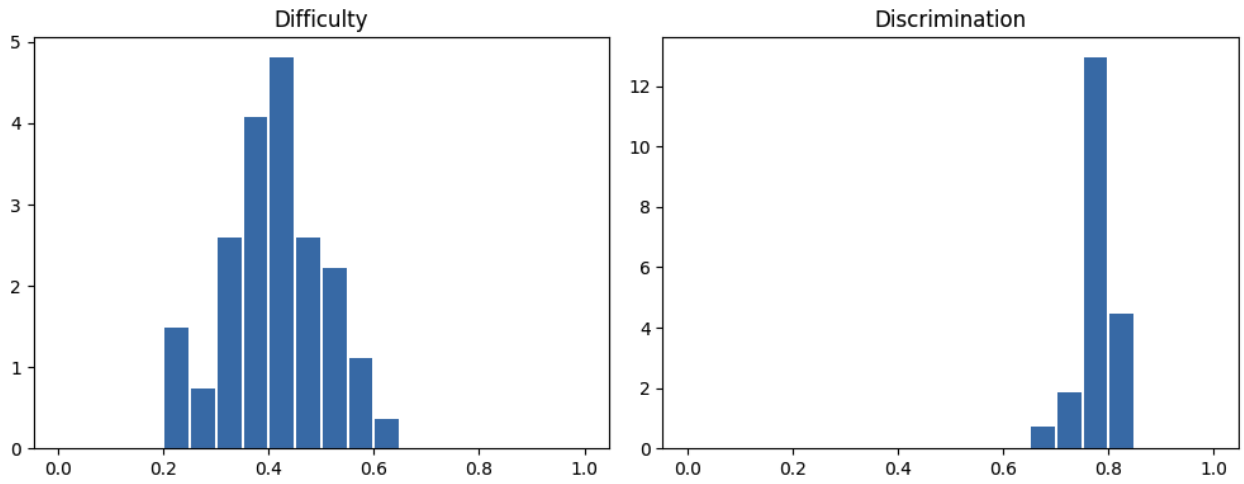

Figure 4. Difficulty and discrimination parameters of 54 AI-generated, SME reviewed items in the Study III.

A separate model was fitted for each Power Skill. For interpretability, both parameters are transformed after the model to [0, 1] range. For "difficulty", the transformation is the standard normal CDF (same as the population distribution of ability assumed by the IRT model), for "discrimination" it is x/(1+x).[3]

In a subsequent study (Study IV), also conducted in April 2026, the same methodology was used with the content developed to measure Power Skills in the Leadership Edition (40 items across 3 Power Skills, n=431). **No significant difference was found among any groups.** The IRT parameter distribution was similar to the one found in Study III. The difficulty range was approx. 0.2-0.8, and the discrimination range approx. 0.7-0.85.

The internal consistency of assessment items within each Power Skill was found to be high: the lowest Cronbach's alpha was 0.92 in Study III and 0.95 in Study IV.

Another study (Study V) was conducted in April 2026 with the goal of evaluating the agreement between human scorers and our AI scoring model. The responses obtained from Study III were used. The study was conducted in a multi-round setup. 20 scorers were initially recruited through

[3] For IRT discrimination, a common interpretation scheme, going back at least to Baker, 2001 (p. 34), uses the labels "very low," "low," "moderate," "high," very high" with cut-points 0.35, 0.65, 1.35, 1.7 un the untransformed space, meaning that the transformed values in the range [0.63, 1) are interpreted as "very high" (0.63 ≈ 1.7/(1 + 1.7)). For transformed difficulty and participant ability one can use the same labels with cut-points 0.15, 0.4, 0.6, 0.85. These are virtually equivalent to cut-points -2, -1, 1, 2 in the untransformed space, which are motivated as the convenient integer *z*-values for the standard normal distribution of ability in the model.

the Prolific research platform, screened for management experience, reasoning skills and AI annotation experience. After the first calibration round, 10 were selected based on scoring agreement metrics and results inspection. After the second calibration round, 6 were selected, who then participated in 3 more rounds. In total, 400 responses, randomly selected from the data, were scored, each by the entire group of scorers. The consensus score was formed as the mean score. The results were:[4]

- Inter-rater reliability among human scorers: mean quadratic weighted kappa (QWK) of 0.708 and Krippendorff's alpha of 0.714.
- Consistency between AI scoring and individual human evaluations: mean QWK of 0.723.
- Agreement between AI scoring and the consensus human score: QWK of 0.788.

Substantial inter-rater reliability indicates that the assessment responses are not strongly open to interpretation. Substantial agreement of AI with the human raters indicates that the model evaluates responses in largely the same way as human scorers. We note that the QWK between the mode and a human scorer is on average higher than among different human scorers.

7.7 Estimation of Skill Proficiency: Skill Proficiency Model

The approach to estimating skill proficiency is Bayesian: scores obtained by a candidate or an employee on any assessment item (as well as from any other potential source, such as from the analysis of their resume) are the measurements of the skill proficiency as a latent variable defined on the interval [0,1]. The posterior distribution of this latent variable, for each skill and for each individual, is stored as the current estimate of their proficiency, and reportable statistics can be extracted from it as needed. For instance, the best single-value estimate of proficiency is obtained as the mean of the posterior distribution, while the standard deviation of this distribution provides a simple definition of the credible interval of this estimate. We allow for partial scores by generalizing the traditional Bayesian framework of binary outcomes, without losing its main characteristics. Multi-tagged items are treated by separating skills: an assessment item tagged with multiple skills generates multiple scores (one per skill), so that the proficiency calculation perceives such an item as a collection of measurements.

The system of skills is assumed to be hierarchical, and an incoming measurement of a skill triggers a recalculation of not only that skill but also of its ancestor skills. For each skill we maintain the posterior distribution due to direct measurements, a posterior distribution induced by measurements of descendent skills, and an operation of combining them: the combined posterior distribution is the one normally used downstream from the skill proficiency model.

[4] For calculating metrics that require the scores to have a certain ordered set of values, rather than any value within a range, the consensus scores of human scorers, as well as the AI model scores, were rounded to the nearest such value.

Figure 5. Example for visual representation of power skills proficiencies.

7.8 Estimation of JobSuccessFit™

Given a posterior distribution of a skill proficiency, we calculate the value of the skill-specific fit to a job. We define the link function $L(x)$ which is the value of the fit for the proficiency value x. This function encodes the job requirements on the skills. Like the proficiency itself, the fit is a latent variable, whose primary reportable characteristic is its mean. Thus, for a skill s, the reported fit value is $f_s = E_{post}[L(x_s)]$, where we take the expectation value w.r.t. the posterior distribution of that skill's proficiency x_s. Other characteristics, such as credible intervals, can be extracted as well, if needed. The method does not impose any substantial conditions on the form of the link function, other than it should be defined on [0,1], and its range should be in [0,1]. A simple form of the link function is a flat cap: $L(x) = min(x/X, 1)$, where X is the threshold level of proficiency required by the job. Each individual's fit to the job is characterized by a set of skill-specific fit values $\{f_s \in [0,1]\,,\ s = 1, 2,...\}$. These can be aggregated ("scalarized") into a single value, for instance by a weighted average, the skill weights being parameters of the job. This value can be used to rank candidates by their fit to the job.

The aggregated fit can be supplemented by other characteristics that show other aspects of multi-dimensional fits. Selection of candidates (or talent development insights for employees) in the multidimensional space of fits f_s can be regarded as a multi-objective optimization problem, where we can short-list the Pareto-optimal candidates. A candidate X is said to dominate a candidate Y if $f_s^{(X)} > f_s^{(Y)}$ for all skills s, and a candidate is called Pareto-optimal if s/he is not dominated by any candidate for the same job. Thus, short-listing the Pareto-optimal candidates has the following practical meaning to the recruiter: it is guaranteed that for any candidate not on the short-list there is at least one short-listed candidate who is a superior fit on *all* the skills we are considering. Moreover, having removed the Pareto-optimal candidates from the pool, one can repeat the operation on the remainder, thus producing a secondary short-list, and so on until the entire pool of candidates is split into a sequence of cohorts (known as Pareto fronts). Within each cohort, the candidates can still be ranked by the aggregate fit value.

Another possibility is to perform the principal component analysis (PCA) on the multi-dimensional fits of candidates in the pool. The principal components can be viewed as the candidate archetypes, described by their loadings. The PCA scores characterize a candidate in the space of such archetypes. A practical advantage of this approach lies in its use as a dimensionality-reduction technique. Limiting ourselves to the first two principal components, we can visualize the pool of candidates in a 2-D diagram with two described archetypes.

8. 2025 Studies for Ignis AI PowerSkillsAssessment™

The development of Ignis AI PowerSkillsAssessment™ follows a rigorous, evidence-based research and validation process grounded in cognitive science, psychometrics, and AI modeling. The system uses multi-method data collection, advanced psychometric modeling (including item response theory and latent skill estimation), and validation across diverse populations to ensure fairness, reliability, and construct validity. Together, these studies demonstrate how Ignis AI is building a scalable, scientifically robust platform for measuring and developing human power skills such as creativity, communication, collaboration, leadership, analytical thinking, productivity, and AI fluency both in low - and high-stakes settings (Study I and Study II, respectively).

8.1 Study I: Pilot Validation Using Prolific Sample - Low Stakes Setting

Method

The first pilot study aimed to establish initial validity and reliability evidence for the Ignis AI PowerSkillsAssessment™ using a diverse, US adult population recruited through the Prolific research platform in July through September 2025 (total n=353, minimum of 50 per skill). The study focused on measuring seven Power Skills—creative thinking, collaboration, communication, leadership, analytical reasoning, productivity, and AI fluency in a low-stakes setting (i.e., where participants treat the assessment as low-stakes with no implications to hiring decisions or talent review). Each participant completed the assessment for a single skill, which included scenario-based tasks designed to elicit behaviors aligned with the skill domain. The sample

represented a range of demographic and professional backgrounds, ensuring variability necessary for psychometric calibration and cross-group fairness testing

First, subject matter experts examined the construct validity of the assessment and data collected ensuring that an assessment accurately measures the specific skill or construct it was designed to capture, aligning the assessment's content, tasks, and scoring with the underlying theoretical framework of that construct. Second, psychometric analyses tested the underlying structure of the PowerSkills model, and reliability coefficients (Cronbach's α) were calculated for each domain. Next, key psychometric item properties such as item total correlations were investigated in order to inform item selection and inference optimization. Last, scores were passed through the Ignis AI skill proficiency model with human-in-the-loop evaluation.

Key Findings

Results demonstrated strong internal consistency across domains (α ranging from .82 to .91). Standard item selection criteria for workforce assessments revealed 96% acceptance rate in this preliminary study in a low-stakes setting (based on key psychometric properties such as item difficulty and item consistency with the intended subskill). Additionally, human evaluation of responses from constructed-response items revealed meaningful qualitative differences between high and low scorers, particularly in creativity and leadership indicators such as originality, adaptability, and integrative decision-making.

Implications

The findings from Study I provide compelling initial validation that Ignis AI's PowerSkillsAssessment™ reliably captures measurable differences in human power skills across diverse individuals in a low-stakes setting. The results support the assessment's construct validity and internal coherence while providing the foundation for longitudinal and cross-context validation. Importantly, the data reinforce the feasibility of using AI-driven, open-response tasks to generate meaningful behavioral indicators—moving beyond traditional self-report measures toward authentic skill measurement.

8.2 Study II: Applied Pilot with Employees - High(er) Stakes Setting

Method

The second validation study extended the evidence base through a pilot conducted in September 2025 with employees in two technology companies, focusing on psychometric calibration, performance distribution, and applied validity across real-world learning in a high-stakes setting (i.e., where participants treat the assessment as higher-stakes with potential implications to goals setting as part of talent review). Participants included professionals from product/project management, software development, UX/AI design and senior management (n=34). The pilot involved employees who were instructed to complete the assessment honestly and provide detailed feedback. The goal was to evaluate how the Ignis AI PowerSkillsAssessment™ functions in applied workforce environments and how its results can inform learning goals setting and targeted human power skill development initiatives.

The assessment battery included interactive, scenario-based simulations representing work-relevant challenges that required communication, creativity, collaboration, and decision-making under uncertainty. Automated scoring models were trained using natural language processing (NLP) and Bayesian proficiency estimation to evaluate responses along key performance indicators (PLDs). Unlike Study I, each participant worked on the full assessment battery, covering the full set of power skills.

Key Findings

Inter-domain reliability remained robust (average $\alpha = .85$).

A preliminary correlation matrix among the proficiencies of power skills was estimated from this data. The typical (median) correlation value was 0.21 (low). The highest correlation (0.63) was observed between the skills Leadership and Communication.

Standard item selection criteria for workforce assessments in a high-stakes setting revealed 75% acceptance rate in this applied pilot with employees providing invaluable insights for further assessment development and optimization.

Response analytics revealed that the combination of AI-assisted scoring and task sequencing allowed efficient measurement while maintaining participant motivation and data quality. Pilot scores were passed through the skill proficiency model to provide individuals with power skills proficiency estimates on all skills. A proficiency estimate has the appearance of a bell-curve, which indicates the uncertainty spread around the main estimated value. The proficiency estimates are based only on the items that met the psychometric criteria. Below is an illustrative example for proficiency estimates for one participant.

The plots show the posterior probability distributions of the participant's proficiency level on each skill, measured on the scale from 0 to 1. In Bayesian statistics the posterior distribution represents the state of our belief about the value of the proficiency, based on all the evidence (scores) that the model has consumed so far. The true proficiency is never known exactly (it is a latent variable), and the posterior distribution describes the probabilities assigned to various possible values of proficiency. From it, we can extract the reportable currently-best estimate of proficiency (typically, the mean) and the interval of uncertainty (the spread of the distribution around the mean). As usual in Bayesian statistics, more evidence tends to shrink the uncertainty: it would make the posterior distributions have sharper, more narrow peaks, which would mean that there is very little uncertainty left. To get closer to that state, we need to increase the amount of evidence that goes into the model. This can be done in three complementary ways:

1. Increase the number of assessment items.
2. Develop more multi-tagged assessment items, so that the same response serves as evidence for multiple skills, thus increasing the amount of evidence for each skill.
3. Supplement the assessment with the analysis of descriptive data sources for the candidate or employee (resumes, cover letters).

The dashed vertical lines with numeric values indicate the means of those distributions, which serve as a single, interpretable proficiency scores. However, the calculation of the Ignis AI JobSuccessFit™ is based on the entire posterior distributions and not just on the means, thus capturing the uncertainty as well.

Implications

Findings from Study II reinforce the Ignis AI PowerSkillsAssessment™ as both a scientifically valid and practically useful tool for assessing critical human capabilities in a high-stakes setting. The pilot demonstrated a strong foundation for authentic skill measurement and user trust. The alignment between measured proficiency levels and perceived task authenticity provides encouraging evidence that participants' perceptions are predictive of psychometric signals, a positive validation outcome. By linking psychometric precision with applied relevance, the study highlights how Ignis AI's platform can provide organizations with actionable insights to guide talent development, coaching, and leadership growth. The integration of adaptive item delivery and AI-based response modeling represents a major step forward in scaling valid, reliable human skills assessment.

8.3 Directions for Future Studies

Together, these three pilot studies conducted in 2025-2026 provide promising preliminary evidence for the reliability, validity, fairness, and utility of the Ignis AI PowerSkillsAssessment™ in both low- and high-stakes settings. Across populations and contexts, results consistently demonstrate that human power skills—long considered intangible—can now be measured with scientific rigor and interpreted for actionable growth. Future work will expand this validation across industries, jobs, job levels and career trajectories and refine assessment design through a wider range of scenarios and task types, and explore longitudinal outcomes to link power skills development with performance and talent development at scale.

9. Summary

The development of the Ignis AI PowerSkillsAssessment™ represents a major advancement in the science and practice of measuring human capabilities. Drawing on more than a decade of interdisciplinary research across cognitive science, psychometrics, and artificial intelligence, Ignis AI has established a validated, evidence-based platform for assessing and developing essential human power skills—creativity, communication, collaboration, leadership, analytical thinking, productivity, and AI fluency.

Studies I-III provide promising preliminary evidence for the validity, reliability, and fairness of the Ignis AI PowerSkillsAssessment™ in both low- and high(er)-stakes environments. *Studies I and III*, conducted with a diverse adult population, confirmed the internal consistency, and psychometric soundness of the model, demonstrating that these human skills can be measured objectively through performance-based tasks rather than self-reports. *Study II*, implemented with employees in professional settings, reinforced these findings and highlighted the system's capacity to generate authentic, actionable insights for workforce development through AI-driven

scoring and Bayesian proficiency estimation. *Study III*, provided evidence on a full psychometric validation of the power skills assessment (n=613) demonstrating industry standard psychometric qualities of the assessment and evidence of no adversary effects (i.e., biases) in the context protected groups (gender, race/ethnicity, and age).

Together, the studies underscore Ignis AI's rigorous evidence-based approach, combining human expertise with adaptive AI to capture the complexity of human potential. The research demonstrates that human capabilities, once considered intangible, can be measured, visualized, and cultivated with scientific precision. As organizations navigate the transformation of work in the age of intelligent technologies, Ignis AI offers a scalable and validated framework to bridge the human–AI divide, enabling individuals and organizations to identify, grow, and apply the very skills that make us most human.

Together, the studies underscore Ignis AI's rigorous evidence-based approach, combining human expertise with adaptive AI to capture the complexity of human potential. The research demonstrates that human capabilities—once considered intangible—can be measured, visualized, and cultivated with scientific precision. As organizations navigate the transformation of work in the age of intelligent technologies, Ignis AI offers a scalable and validated framework to bridge the human–AI divide, enabling individuals and organizations to identify, grow, and apply the very skills that make us most human.

10. References

Amabile, T. M., & Pratt, M. G. (2016). The dynamic componential model of creativity and innovation in organizations: Making progress, making meaning. *Research in Organizational Behavior, 36,* 157–183.

Ancona, D. (2012). Sensemaking: Framing and acting in the unknown. In S. Snook, N. Nohria, & R. Khurana (Eds.), *The handbook for teaching leadership* (pp. 3–19). Harvard Business Press.

Association of Test Publishers and International Test Commission. (2025, July). *ITC Guidelines on Technology-based Assessment (version 1.1).* https://www.intestcom.org/upload/media-library/tba-guidelines-ver-11-july-2025-1754044782Z3vV9.pdf

Baker, F. B. (2001). *The basics of item response theory*. ERIC Clearinghouse on Assessment and Evaluation. For full text: https://files.eric.ed.gov/fulltext/ED458219.pdf

Drucker, P. F. (1993). The rise of the knowledge society. The Wilson Quarterly, 17, 52-72.

Educational Testing Service (ETS). (2026). *ETS Human Progress Report*. Princeton, NJ: ETS.

International Test Commission and Association of Test Publishers (2025). *Guidelines for technology-based assessment*. International Test Commission.

Kellogg, K., & Truelove, E. (2019). *A political perspective on organizations*. MIT Sloan School of Management.

Liu R, Liu H, Shi D, Jiang Z. Diagnostic Classification Models for a Mixture of Ordered and Non-ordered Response Options in Rating Scales. Appl Psychol Meas. 2022 Oct;46(7):622-639. doi: 10.1177/01466216221108132. Epub 2022 Jun 24. PMID: 36131839; PMCID: PMC9483220.

Loaiza, I., Rigobón, R. (2025). *The EPOCH of AI: Human–machine complementarities at work* (Working Paper No. 5028371). MIT Sloan School of Management. Cambridge, MA. https://ssrn.com/abstract=5028371

MIT Sloan School of Management. (2025, June 17). *These human capabilities complement AI's shortcomings. MIT Sloan Ideas Made to Matter*. https://mitsloan.mit.edu/ideas-made-to-matter/these-human-capabilities-complement-ais-shortcomings

Organisation for Economic Co-operation and Development (OECD). (2017). *PISA 2015 Results (Volume V): Collaborative Problem Solving.* OECD Publishing.

Organisation for Economic Co-operation and Development (OECD). (2024). *PISA 2022 Results (Volume III): Creative Minds, Creative Schools.* OECD Publishing.

Repenning, N. P., & Kieffer, D. C. (2025). *There's got to be a better way: How to deliver results and get rid of the stuff that gets in the way of real work*. PublicAffairs.

Rosen, Y. (2014). *Comparability of conflict opportunities in human-to-human and human-to-agent online collaborative problem solving. Technology, Knowledge and Learning, 19*(1–2), 147–174.

Rosen, Y. (2015a). *Computer-based assessment of collaborative problem solving: Exploring the feasibility of human-to-agent approach. International Journal of Artificial Intelligence in Education, 25*(3), 380–406.

Rosen, Y. (2015b). *Collaborative problem solving in the PISA 2015 assessment: From human-to-human to human-to-agent measurement*. Invited presentation at the Expert Workshop on Innovative Assessment, National Institute of Education, Nanyang Technological University, Singapore.

Rosen, Y. (2017). *Conversational agent assessments of human skills in Microsoft training programs: Efficacy and further research and development directions.* Redmond, WA: Microsoft Corporation.

Rosen, Y. (2023, March 27). *Approaches to creativity assessment – A journey* [Invited keynote]. DEFI Annual Event, University of Cambridge, Cambridge, UK. [Video]. YouTube. https://www.youtube.com/watch?v=wATW9Mr0z4s

Rosen, Y. (2024, June 4). *Unlocking creativity in the age of AI: Innovative approaches to assessment, scoring, and learning* [Opening keynote]. 5th EdTech & Computational Psychometrics Summit (ECPS 2024), Association of Test Publishers, Washington, DC.

Rosen, Y., & Carlile, N. (2023). *What can artificial intelligence teach us?* [Five-part seminar series]. Brigham Education Institute, Brigham and Women's Hospital. [Video]. YouTube. https://youtu.be/bdRrYBqhw_g

Rosen, Y., Ferrara, S., & Mosharraf, M. (Eds.). (2015). *Handbook of Research on Technology Tools for Real-World Skill Development.* IGI Global.

Rosen, Y., & Foltz, P. (2014). Assessing collaborative problem solving through automated technologies. *Research and Practice in Technology Enhanced Learning*, 9(3), 389–410.

Rosen, Y., Kadar, A., & Rushkin, I. (2026). *Artificial Intelligence Systems and Methods for Talent Talent Development and Talent Acquisition* (U.S. Patent Non-Provisional Application). U.S. Patent and Trademark Office.

Rosen, Y., Mokyr Horner, E., Purifoy, C., & Thomsen, B. S. (2023). *Developing & assessing creative skills with AI*. SXSW EDU Conference. Austin, TX.

Rosen, Y., Mosharraf, M., Rushkin, I. (2025). *AI-driven assessment and feedback to foster creative thinking development*. U.S. National Science Foundation.

Rosen, Y., & Rushkin, I. (2026a). *Measuring Creativity in the Age of Generative AI: Distinguishing Human and AI-Generated Creative Performance in Hiring and Talent Systems*. Paper Presented at the 2026 BIG.AI@MIT Conference, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA.

Rosen, Y., & Rushkin, I. (2026b). *From Measurement to Action: A Learning Engineering Approach to AI-Powered Assessment for Human Power Skills Development*. Paper Presented at the 2026 Learning Engineering Research Network Convening, Learning Engineering Institute, Arizona State University, Tempe, AZ.

Rosen, Y., & Rushkin, I. (2026c). *Proposed Framework for Agentic AI Assessment of Creative Thinking for the U.S. Air Force.* Paper Presented at the 2026 Artificial Intelligence Workshop: Accelerating Data and Analytics Capabilities for AI Military Operations Research Society (MORS), Carnegie Mellon University, Pittsburgh, PA.

Rosen, Y., Rushkin, I., Ang, A., Fredericks, C., & Tingley, D. (2017). *Designing adaptive assessments in MOOCs. Proceedings of the Fourth ACM Conference on Learning @ Scale,* Cambridge, MA.

Rosen, Y., Rushkin, I., Ang, A., Munson, L., Lopez, G., & Tingley, D. (2018). *The effects of adaptive learning in a massive open online course on learners' skill development. Proceedings of the Fifth ACM Conference on Learning @ Scale.* London, UK.

Rosen, Y., Stoeffler, K., & Simmering, V. (2020). *Imagine: Design for creative thinking, learning, and assessment in schools. Journal of Intelligence, 8*(2), 18.

Rosen, Y., Turkay, E., & Eidelman, I. (2016). *Evaluation of human power skills development in experiential learning programs: The Center for Public Interest Careers (CPIC) efficacy studies.* Harvard University, Cambridge, MA.

Rosen, Y., Jaeger, G., Newstadt, M., Bakken, S., Rushkin, I., Dawood, M., & Purifoy, C. (2023). *A multidimensional approach for enhancing and measuring creative thinking and cognitive skills. International Journal of Information and Learning Technology, 40*(4), 334–352.

Rushkin, I., & Rosen, Y. (2024). *Automated evaluation of free-form answers and generation of actionable feedback to multidimensional reasoning questions* (U.S. Patent No. 12,020,593). U.S. Patent and Trademark Office.

Schein, E. H., & Schein, P. A. (2023). *Humble Leadership (2nd ed.): The power of relationships, openness, and trust*. Berkeley, CA: Berrett-Koehler Publishers.

Stoeffler, K., Rosen, Y., Bolsinova, M., & von Davier, A. A. (2020). Gamified performance assessment of collaborative problem solving skills. Computers in Human Behavior, 104, 106036.

Stoeffler, K., Way, J., & Rosen, Y. (2022). *When actions speak louder than words: Stealth assessment of social and emotional skills.* In *Assessing Competencies for Social and Emotional Learning* (pp. 135-150). Routledge.

World Economic Forum. (2025). *The Future of Jobs Report 2025*. Geneva, Switzerland: World Economic Forum.

About the Authors

Dr. Yigal Rosen is an internationally recognized leader in the science of human capability assessment, known for pioneering work at the intersection of artificial intelligence, computational psychometrics, and learning sciences. As a Co-Founder and Chief Product Officer at Ignis AI, he leads the development of the PowerSkillsAssessment™—a next-generation platform that uses AI and psychometrics to measure and develop human capabilities such as creativity, communication, collaboration, leadership, analytical thinking, productivity, and AI fluency. He is also pursuing his Executive MBA at the MIT Sloan School of Management (Class of 2027), further strengthening his capacity to lead global innovation at the intersection of science, strategy, and human development. Before founding Ignis AI, Dr. Rosen served as Chief Academic Officer and Senior Vice President of AI & Data Science at BrainPOP, where he led the company's transition into an AI- and data-driven learning platform. His leadership was instrumental in integrating research-based design, adaptive learning, and machine learning models to power creativity and reasoning assessment at scale—advancements that helped position BrainPOP as a global leader in digital learning and contributed to its $875M acquisition by KIRKBI in 2022, the owners of the LEGO Group. Earlier in his career, Dr. Rosen held senior leadership roles at Pearson, ACT, and Harvard University, where he served as Managing Senior Research Scientist at the Office of the Vice Provost for Advances in Learning. At Harvard, he led research and development of innovative AI technologies that made complex human skills measurable and improvable. He pioneered agentic AI for learning and assessment, designing conversational agent systems capable of assessing collaboration, communication, and creative problem solving—technologies later deployed in Microsoft, Harvard, and HarvardX professional training programs. He also taught courses on innovative assessment design at Harvard University's Graduate School of Education, where he helped prepare the next generation of assessment and computational psychometrics experts. Dr. Rosen also led two landmark global assessments for the OECD's Programme for International Student Assessment (PISA)—the PISA 2015 Collaborative Problem Solving Assessment and the PISA 2022 Creative Thinking Assessment—which transformed how education systems worldwide measure and cultivate essential human capabilities for the future of work. He has published extensively in peer-reviewed journals such as Journal of Intelligence, Computers in Human Behavior, Journal of Educational Measurement, and International Journal of Artificial Intelligence in Education. Dr. Rosen is also the inventor on multiple U.S. patents in AI-based scoring, feedback, and adaptive assessment technologies, and his work has been supported by the U.S. National Science

Foundation, Microsoft, and The LEGO Foundation. Dr. Rosen earned his Ph.D. in Education from the University of Haifa (Israel), where he was recognized as the youngest recipient of a doctoral degree in the university's history, and completed postdoctoral fellowships at Tel Aviv University's School of Education and Harvard University's Graduate School of Education.

Dr. Ilia Rushkin is a distinguished scientist and innovator at the intersection of machine learning, psychometrics, and data science. With a rich background in mathematics and data science, Dr. Rushkin has spent his career developing scalable computational models that illuminate how learners think, learn, and solve problems. His work advances not only the capacity of AI to assess human skills, but also the integrity, fairness, and interpretability of those systems. Before joining Ignis AI, Dr. Rushkin served as a Senior Data Scientist at Harvard University's Office of the Vice Provost for Advances in Learning, where he co-led the development of the Adaptive Learning Open Source Initiative (ALOSI) in partnership with Microsoft and HarvardX. This initiative introduced open, transparent adaptive architectures, Bayesian skill-tracking methods, and real-time analytics—laying critical groundwork for today's sophisticated AI-based proficiency models. His peer-reviewed publications and conference presentations span venues such as the ACM Learning @ Scale, Educational Data Mining (EDM), and international AI in education forums, illustrating his commitment to bridging theory and practice.

As an inventor and a former Principal AI/ML Engineer at BrainPOP, Dr. Rushkin holds multiple U.S. patents—such as "Systems and Methods for Learner Growth Tracking and Assessments" and "Automated Evaluation of Free-Form Answers and Generation of Actionable Feedback to Multidimensional Reasoning Questions"—which reflect his deep contributions to AI-enabled assessment of complex problem-solving and reasoning skills.

As a Vice President of AI and Data Science at Ignis AI, Dr. Rushkin leads the design and development of the AI/ML, computational psychometrics, and data science infrastructure that powers the PowerSkillsAssessment™. He ensures that each measurement is underpinned by rigorous analytics, transparent algorithmic processes, and secure modeling practices. His leadership ensures that the platform delivers not only scalable, valid assessments of human power skills—but also actionable insights that support individual growth and organizational development.

Dr. Rushkin's career is anchored in the belief that human intelligence and technology must evolve together, and his work reflects a commitment to advancing both the science and the ethical application of AI in learning and talent development.