

RESEARCH REPORT

Accurately Measuring Skills: Isolated vs Multi- Skill Scoring to Predict Performance

Inside this Report

01	Introduction	03
----	---------------------	----

02	Skills Cannot Be Evaluated in Isolation	07
----	--	----

03	Proving a More Realistic Model of Workplace Performance	11
	Methodology	12
	Results	13
	Table 1: Correlation between scoring approaches	15
	Research Notes	16

04	About HiringBranch Assessments	17
	From Job Description to Assessment Scenario Generation	18
	Explainability Matters	20

05	Learn More About HiringBranch	21
----	--------------------------------------	----

In real workplace settings, communication and soft skills are rarely used in isolation.

A typical customer or colleague interaction unfolds as a conversation whereby multiple skills are used at the same time.

A retail employee, for example, does not simply “build rapport” or “demonstrate product knowledge” as separate behaviours. A customer may enter a store and appear to be browsing. The employee approaches, identifies whether help is needed, explores what the customer is looking for, translates that need into suitable product options, adjusts the recommendation through dialogue, and, where appropriate, identifies an opportunity to cross-sell or close the sale.



This single interaction involves multiple skills working together:

- | | |
|-----------------------|------------------------|
| • rapport building | • product matching |
| • needs discovery | • benefit articulation |
| • active listening | • objection handling |
| • attention to detail | • persuasion |
| • critical thinking | • closing |

THE CORE POINT

**None of these skills
fully explains the
interaction on its own.**

What matters is how they are combined,
sequenced, and adapted within
the conversation.

THE SAME PATTERN, ANOTHER INDUSTRY

In a customer service setting, if a customer complains, empathy is important, but empathy alone is not enough. Similarly, in a healthcare setting, if a patient complains about their ailment, empathy is important, but empathy alone is not enough. The healthcare worker must understand the issue, ask the right follow-up questions, use clinical judgement to determine the best response, reassure the patient, manage expectations, and sometimes deliver difficult messages.

Evaluating “empathy” in isolation misses the actual performance requirement of resolving the issue through an effective conversational strategy.

The same principle applies to any role in any industry. It is never one simple skill that does all the work.

Rather, it is the combination of the right skills that makes an employee effective at their job.

Empathy without problem solving could feel polite but ineffective.

Problem solving without reassurance may feel abrupt or dismissive. Product knowledge without needs discovery may lead to irrelevant recommendations. From this perspective, evaluating only one skill at a time makes no sense.

Traditional assessments often treat competencies as independent traits. A candidate may be asked to do an assessment for problem solving, another one for active listening, a third for language, and so on. The problem is, separating these competencies fails to reflect how communication works in practice.

Independent skill evaluation also forces recruiters to reconcile multiple scores among potentially hundreds or thousands of applicants. Stacking single-skill assessments creates real operational overhead at volume, and that overhead becomes a legal risk specifically when the combination method isn't pre-validated and consistently documented, simply because multiple scores exist. Using one validated instrument removes the combination-methodology risk entirely. For candidates, they end up with a stack of assignments to complete for just one application. It's a lose-lose situation.

TWO WAYS TO MEASURE THE SAME CONVERSATION

**Isolated scoring
breaks the interaction
apart. Combined
scoring keeps it whole.**

TRADITIONAL ASSESSMENT – FOUR DISCONNECTED MEASUREMENTS

Acknowledgement

Assessment 1

Positive language

Assessment 2

Empathy

Assessment 3

Active listening

Assessment 4

Four separate scores that the recruiter must reconcile – and a stack of assignments for one application.

HIRINGBRANCH – ONE CONVERSATION, SKILLS MAPPED ALONG IT



Empathy



Critical thinking



Negotiation



Collaboration



Language proficiency



Spoken fluency



Active listening

One conversational flow, evaluated end to end, producing a single integrated score.

Using conversation flows to measure performance

Effective workplace communication depends on skills sequencing, adaptation, and integration.

That means candidates need to know not only what to say, but when to say it, and how to adjust based on the other co-worker or customer's response. When applicable, they also need to understand how to move the interaction toward a successful outcome. This is why skills-based hiring teams need to evaluate performance across full conversational flows.



If real skills are expressed simultaneously through communication, why aren't more assessments measuring all role-relevant skills in one short candidate experience, like HiringBranch does?

At HiringBranch, we believe the best pre-hire evaluation is not a checklist of isolated soft skills.

It is a sequence of decisions and behaviours embedded in real conversations, and that's how it should be measured in a hiring process to get the best quality hire. So we set out to prove it.

Here's what our first-party research found. Empirical validation confirms the advantage of assessing skills in combination with one another (see Table 1) over using singular skills as hiring signals.

Methodology

To do this analysis, the HiringBranch research team focused on one frontline role essential for large organizations - customer service. The analysis focused on four skills that are fundamental to great customer service representatives: acknowledgement, positive language, empathy, and active listening.

457

real candidates
(n = 457)

4

expert human evaluators,
manually annotating

Four expert human evaluators were asked to score a batch of nearly 500 conversations (n=457) as the ground truth. Rather than grading isolated skills, the evaluators assigned a holistic score to represent the candidate's overall conversational performance. Each evaluator manually annotated conversations, with a high inter-annotator agreement¹ to prove validity.

Each individual trait (of the four highlighted skills) is automatically evaluated by a dedicated in-house model trained specifically for that soft skill. Because each individual skill score captures only one dimension of the multi-faceted judgment human raters form, the human-evaluator correlations were expected to be modest on their own, and they serve as a diagnostic baseline rather than an indictment of single-skill scoring.

The research team then compared the human scores individually and as an average to HiringBranch's proprietary AI results, which scores all soft skills and communication skills simultaneously. In other words, each method was evaluated for how well it could read the candidate's skill in question.

Results

When individual soft skill scores are correlated with human expert ratings of the selected customer service skills, each captures only a partial signal (**Pearson r = 0.21–0.55**). Even averaging all four skill scores together only modestly improves prediction (**Pearson r = 0.68**).²

THE HEADLINE RESULT

0.85 Pearson r

$r^2 = 0.71$

variance explained

0.68

simple average of 4 skills

0.21–0.55

individual skill scores

By contrast, the HiringBranch proprietary AI achieves a substantially stronger agreement (**Pearson r = 0.85, r^2 = 0.71**) than the other single scoring methods, as seen in Table 1.

TABLE 1

Correlation Between Automated Scoring Approaches and Human Expert Ratings:

Individual Skill Scores, Averaged Skill Score, and the HiringBranch Proprietary AI

COMPARISON ITEMS		PEARSON R
Human Score vs. ML Acknowledgement Score		0.5416
Human Score vs. ML Positive Language Score		0.5082
Human Score vs. ML Empathy Score		0.2127
Human Score vs. ML Active Listening Score		0.5551
Human Score vs. Simple Average of ML 4 skills above scores		0.6760
Ridge regression on the above 4 skill scores ($R^2 = 0.499$)		0.706
HiringBranch Proprietary AI Combined Skills Score³ 0.8464		

This demonstrates that the HiringBranch Proprietary AI produces a more accurate measure of communication and skill quality by analyzing them together, rather than any single skill assessment or their simple combination.

RESEARCH NOTES

- 1 To evaluate the inter-rater reliability (IRR) of the collected data in our corpus, a two-way random-effects Intraclass Correlation Coefficient (ICC) model was calculated across 457 items. Since the final corpus utilizes the average score of all four annotators, the average-measures coefficient is our primary indicator of data quality. The analysis yielded an excellent average-rater absolute agreement of **$ICC(2,4) = 0.895$ (95% CI[0.88, 0.91], $F(456, 1368) = 9.77, p < .001$)**. For full transparency, the single-rater reliability was **$ICC(2,1) = 0.680$ (95% CI[0.64, 0.72])**, representing moderate-to-good baseline agreement before aggregation. These metrics justify the use of the averaged scores as a highly reliable ground truth for subsequent machine learning analysis.
- 2 The proprietary model's correlation was evaluated on a held-out test set, ensuring the reported $r=0.85$ reflects out-of-sample generalization rather than in-sample fit.
- 3 This scoring formula includes more than just these four soft skills and also allocates optimal weights to the four skills.

A Proprietary Taxonomy Built Around Real Conversations That Drive Performance

The central insight behind HiringBranch's approach to assessments is simple: hiring signal success is not a checklist of isolated soft skills. It is a sequence of decisions and behaviours embedded in conversations.

By using a proprietary, research-based skills taxonomy that translates jobs into conversation flows, HiringBranch creates assessments that are realistic, role-specific, and focused on the skills that actually matter for performance.

HiringBranch assessments evaluate how candidates combine language proficiency, pragmatic competence, soft skills, and critical thinking in the context of role-play scenarios relevant to the position. Recruiters are enabled by interpretable scoring, because each assessment is grounded in clearly defined conversational behaviours rather than vague impressions of communicative ability.

Unlike traditional language proficiency tests, psychometric assessments, or behavioural tests, HiringBranch evaluates how candidates use language, judgment, and soft skills together to perform job-relevant communication tasks. HiringBranch assesses workplace communication as it actually occurs, through realistic, real-time conversations.

HOW AN ASSESSMENT IS BUILT

From Job Description to Assessment Scenario Generation

Rather than beginning with a generic list of soft skills, the HiringBranch system begins with the job itself. Our experts start off by analyzing a job description and identifying the conversational situations that are typical for the role. This allows HiringBranch to move from a static job description to a dynamic model of workplace communication.

01 **Job description**

02 **Identify conversational situations typical for the role**

03 **Translate situations into realistic conversation flows**

04 **Map relevant skills to the stages of the interaction**

05 **Realistic assessment scenario**

06 **Interpretable performance scoring**

Situations are then translated into realistic conversation flows, with the relevant skills mapped to the different stages of the interaction. Each flow contains multiple embedded skills correlated to performance. The assessment therefore measures whether the candidate can manage the conversation effectively from beginning to end, rather than whether they can demonstrate a single soft skill in isolation.

A candidate applying for a sales role is not evaluated through abstract questions about sales ability. They are placed in a simulated sales conversation. A candidate applying for a customer service role is not simply asked whether they are empathetic. They must handle a realistic customer issue and demonstrate empathy, critical thinking, reassurance, and resolution management within the same interaction. This makes the assessment more authentic and more closely connected to workplace performance.



Beyond Generic LLM-as-a-Judge Scoring

Many modern AI assessment approaches rely heavily on generic “LLM-as-a-judge” scoring. While such methods can produce broad evaluations, they often lack transparency. They may indicate that a response was strong or weak, but they do not reliably explain which job-relevant behaviours were demonstrated, which were missing, or how the candidate performed across the interaction.

HiringBranch addresses this interpretability gap through its proprietary taxonomy and automatic scoring architecture. Our system maps candidate behaviour to observable communicative actions within specific conversation flows. This allows the assessment to produce structured, explainable, and job-relevant feedback without relying solely on opaque holistic judgments.

The result is a more transparent evaluation of workplace skills.

HiringBranch can identify not only whether a candidate performed successfully, but also which aspects of the interaction contributed to that performance.



Learn More About HiringBranch

HiringBranch is an award-winning AI-powered talent assessment platform headquartered in Canada with global operations. Trusted by enterprise organizations including Bell Canada and Infosys, HiringBranch helps frontline and high-volume hiring teams identify top talent faster - before the first interview. By assessing candidates' critical thinking, communication, and soft skills at 98% accuracy, HiringBranch gives recruiters a clear picture of who can actually do the job, reducing time-to-fill, bad hires, and attrition. Founded in 2022, HiringBranch has delivered over one million assessments worldwide and stands behind every hire with a results guarantee.

98%

scoring accuracy

1M+

assessments delivered

80+

ATS integrations



GET IN TOUCH WITH OUR TEAM:

hello@hiringbranch.com

Research conducted by Yubin Xing, PhD Candidate in Computational Linguistics and NLP expert under the supervision of Assaf Bar Moshe, PhD, and Chief Research Officer at HiringBranch.