

PREDICTIVE VALIDITY STUDY · PREPARED FOR

A Leading BFSI Lender

# Does the assessment predict performance?

A 12-month validation of the PMaps Tele-Sales hiring assessment against real business outcomes — 1,118 hires measured on cumulative disbursement, tracing exactly which competencies separate high performers from low.

**1,118**

HIRES VALIDATED

**12-mo**

OUTCOME WINDOW

**84%**PREDICTIVE WEIGHT · 2  
TRAITS**51%**

HIGH PERFORMERS

TELE-SALES · FRONTLINE BFSI

MAR 2024 – MAR 2025

JUNE 2025

## EXECUTIVE SUMMARY

# The assessment separates high performers from low – and the data proves it.

This study validates the PMaps Tele-Sales hiring assessment for a leading BFSI lender by testing a single question that matters to every hiring decision: **do assessment scores actually predict who performs on the job?** We tracked **1,118 hires** over a full 12-month window (March 2024–March 2025) and measured each against an objective business outcome – **cumulative disbursement**.

**The headline.** When candidates were split into high performers ( $\geq$  ₹45 Lakhs disbursed) and low performers, high performers scored higher on **every** section of the assessment. The two competencies that discriminate most – Personality Profiler and Attention to Detail – together carry **84% of the model's predictive weight**, confirming the instrument is weighting the right things.

**What this means for hiring.** The assessment is a valid pre-hire filter: it reliably identifies the candidates who go on to disburse more, book more appointments, and stay longer (79% retention). Used as a gate, it concentrates quality – sourcing channels that score well yield up to **75% high performers** versus 40% in the weakest.

**Where to tune.** Two sections add little signal – eVA (1% weight) and Sentence Formation (37% avg score). Reclaiming their test time, rebalancing sourcing, and tightening proctoring where integrity flags cluster will sharpen an already-validated instrument.

## AT A GLANCE

Hires validated	1,118
Outcome window	12 mo
High performers	51%
Outcome metric	₹45L+
Still active	79%
Predictive weight · 2 traits	84%
Avg overall score	58%
Best source yield	75%
Genuine (proctoring)	31%



High performers out-scored low performers on every section – and 84% of the score's predictive weight sits in the two traits with the widest performer gap. The assessment is measuring what matters.

— INSIDE THIS DOCUMENT

# Contents.

From study design to a decision framework — how the assessment was validated, which competencies predict performance, where hiring quality varies by source, and what HR leaders should do next.

---

<b>01</b>	<b>Study Design &amp; Data Universe</b> 1,118 hires, a 12-month window, one objective outcome	<b>04</b>
<hr/>		
<b>02</b>	<b>The Assessment Construct</b> 63 questions, 4 competencies, and how they're weighted	<b>05</b>
<hr/>		
<b>03</b>	<b>The Validation Outcome</b> Splitting the cohort on real disbursement	<b>06</b>
<hr/>		
<b>04</b>	<b>Predictive Signal — What Separates Performers</b> Where 84% of the predictive weight lives	<b>07</b>
<hr/>		
<b>05</b>	<b>Section Performance &amp; Refinement</b> Two sections carry the signal, two need tuning	<b>08</b>
<hr/>		
<b>06</b>	<b>Sourcing Quality — Centers &amp; Vendors</b> 40% to 75% high-performer yield by channel	<b>09</b>
<hr/>		
<b>07</b>	<b>The Center Scorecard</b> Yield, productivity, and integrity, side by side	<b>10</b>
<hr/>		
<b>08</b>	<b>Proctoring &amp; Assessment Integrity</b> Where score trust holds — and where it breaks	<b>11</b>
<hr/>		
<b>09</b>	<b>Recommendations</b> Adopt, refine, operationalize — across three horizons	<b>12</b>
<hr/>		
<b>10</b>	<b>A Decision Framework for Hiring Leaders</b> Current baselines and recommended targets	<b>13</b>
<hr/>		
<b>11</b>	<b>Implementing This in Your Organization</b> A five-stage operating loop you can run in-house	<b>14</b>
<hr/>		
<b>12</b>	<b>Why PMaps</b> How the assessments work — and why they fit your hiring goals	<b>15</b>

---

## 01 · STUDY DESIGN &amp; DATA UNIVERSE

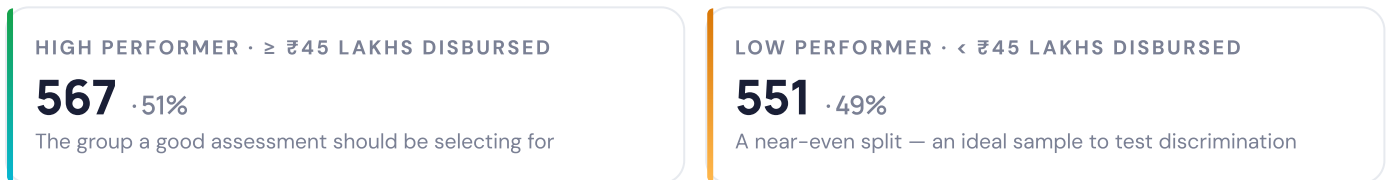
## How the assessment was put to the test.

A predictive-validity study only means something if the outcome is real and the sample is large. Both conditions hold here. We anchored the analysis to **cumulative disbursement** — an objective measure of a Tele-Sales agent's productivity and financial contribution — and tracked every assessed hire over a full year of on-the-job performance.



### The outcome metric

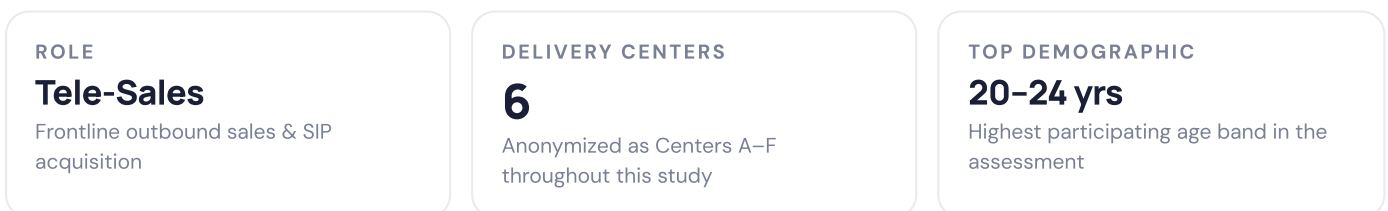
Rather than rely on a subjective manager rating, every candidate's real-world output was used as the ground truth. **Cumulative Disbursement Amount** was selected as the key performance parameter because it objectively reflects what the role exists to do. Candidates were then split into two groups against a **₹45 Lakh** threshold.



**Why a 51 / 49 split is ideal.** A balanced outcome distribution is the strongest possible test bed for predictive validity. If the assessment can cleanly separate two groups of near-equal size, the signal is real — not an artifact of one group dominating the sample.

### Who took the assessment

The cohort skews young, as frontline Tele-Sales hiring typically does — the **20–24 age group** is the highest participating demographic. Candidates were sourced across six delivery centers and a panel of staffing vendors, which lets us isolate not just **whether** the assessment predicts performance, but **where** hiring quality concentrates (Sections 06–07).



## — 02 · THE ASSESSMENT CONSTRUCT

# What the assessment measures – and how it's weighted.

The instrument is a **63-question, 40-minute** battery across four competency sections, each chosen for its relevance to Tele-Sales success. Crucially, the sections are not weighted equally: the **factor loading** column below shows how much each contributes to the predictive composite score – the single most important thing to understand before reading the results.

SECTION · COMPETENCY	QS	TIME	WEIGHT	WHAT IT ASSESSES
<b>Personality Profiler</b> Ambiversion · Achievement Orientation · Influencing Skills · Conscientiousness · Positivity	46	17:00	70%	Balance of assertiveness and adaptability, drive for results, and positive customer interactions.
<b>Sentence Formation</b> English grammar & structuring	6	05:00	15%	Ability to construct clear, grammatically correct communication.
<b>Attention to Detail</b> Focus & accuracy under time pressure	8	06:00	14%	Capacity to focus and handle multiple customer details accurately and efficiently.
<b>PMaps eVA</b> Vocabulary · Grammar · Comprehension · Pronunciation · Fluency	3	12:00	1%	Spoken communication — explaining products and addressing queries persuasively.
<b>Total battery</b>	63	40:00	100%	Composite score benchmarked against a legacy cut of 45%.

## Reading the weighting

Two things stand out immediately, and both are validated by the performance data that follows:

- **Personality Profiler does most of the predictive work — 70% of the composite.** For a relationship-driven sales role, behavioral fit, not raw aptitude, is the dominant signal. This is by design, and the outcome data confirms the design is right.
- **eVA consumes 30% of test time for 1% of the weight.** Spoken-language assessment takes 12 of the 40 minutes but contributes almost nothing to the composite — a clear efficiency opportunity flagged in Section 05.

**Factor loading, in plain terms.** It's the share of the final score each section drives. A section with a high loading moves the hiring decision; a section with a 1% loading barely registers — regardless of how long it takes to administer.

## 03 · THE VALIDATION OUTCOME

# Did the high performers actually score higher?

This is the test of predictive validity. We took the cohort already split by real disbursement and asked whether their assessment scores told the two groups apart **before** they were hired. The answer is unambiguous.

RESULT · VALIDATED

## High performers out-scored low performers on every section of the assessment.

Across all four competencies, the group that went on to disburse  $\geq$  ₹45 Lakhs had recorded higher assessment scores at the pre-hire stage. The instrument was producing the right signal consistently — not on a lucky subset, but across the board. That consistency is what makes the assessment reliable enough to base hiring decisions on.

# 100%

OF SECTIONS FAVORED HIGH PERFORMERS

## The signal concentrates where it should

Not every section discriminates equally. The two with the **widest gap** between high and low performers are Personality Profiler and Attention to Detail — which also happen to carry the two largest factor loadings. The instrument's design and its real-world behavior agree.

SIGNAL · STRONGEST PREDICTOR

### Personality Profiler

The widest performer gap of any section, and 70% of the composite weight. Behavioral traits — drive, influence, conscientiousness — are what separate top Tele-Sales talent.

SIGNAL · STRONG PREDICTOR

### Attention to Detail

The second-widest gap and 14% of the weight. High performers focus and stay accurate under time pressure — a direct correlate of clean, complete customer handling.

SIGNAL · LOWER DISCRIMINATION

### eVA & Sentence Formation

Both record the lowest average scores and contribute the least discrimination. They are refinement candidates, not decision drivers (Section 05).



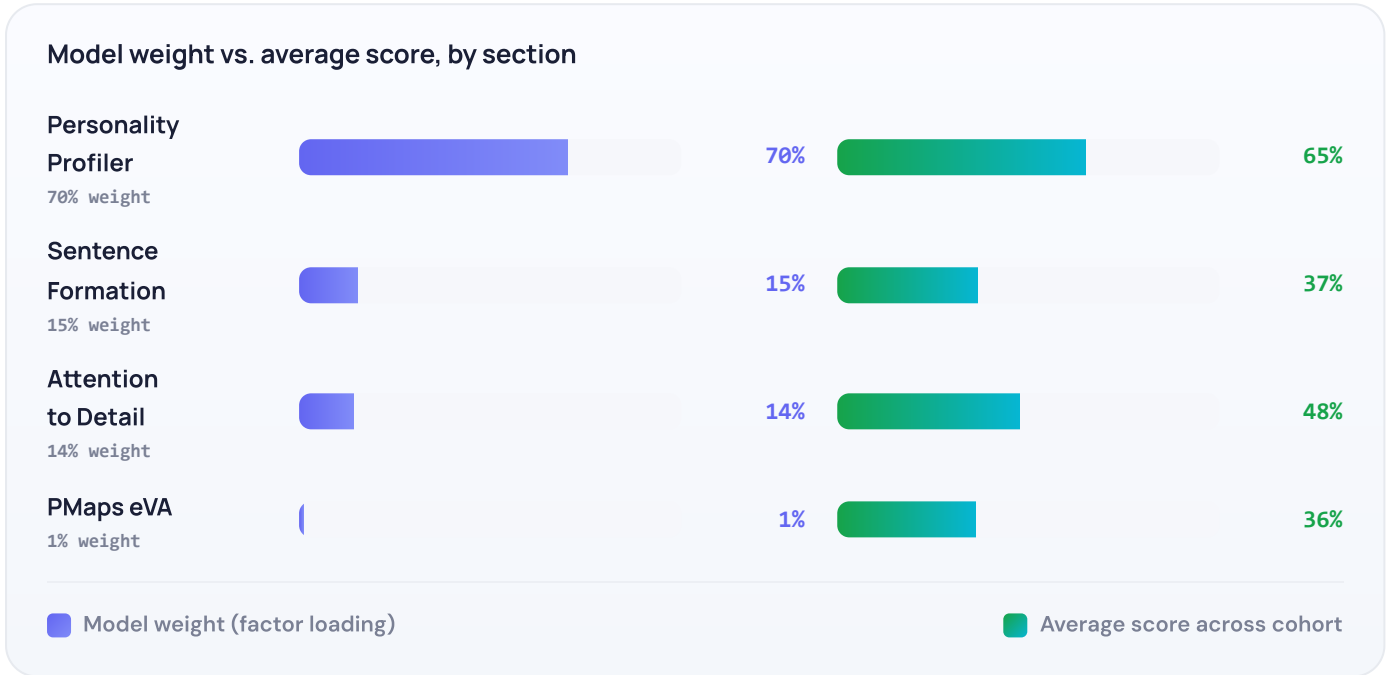
84% of the predictive weight sits in the two competencies with the widest performer gap. The assessment isn't just correlated with performance — it's correlated through the right traits.

**Why this matters for hiring.** A score that separates future high performers from low performers — and does so through job-relevant behavior — is exactly what a pre-hire gate should be. Used at the top of the funnel, it raises the floor on every hire that follows.

04 · PREDICTIVE SIGNAL

# Where the predictive weight actually lives.

Plotting each section’s model weight against its average score makes the structure of the assessment visible. The decision is driven by Personality Profiler; everything else fine-tunes it. This is the chart to keep in mind when setting cut-scores.



## The predictor scorecard

Reading weight, score, and discriminant power together tells HR leaders precisely how to treat each section in a hiring decision.

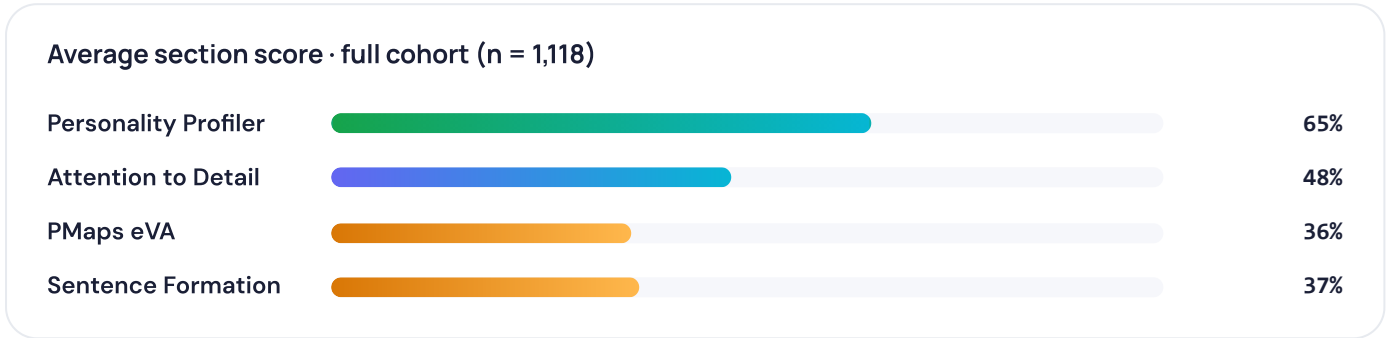
SECTION	WEIGHT	AVG SCORE	DISCRIMINATION	VERDICT
<b>Personality Profiler</b> Behavioral fit	70%	65%	Highest	Core predictor — keep heavily weighted
<b>Attention to Detail</b> Accuracy under pressure	14%	48%	High	Core predictor — protect in scoring
<b>Sentence Formation</b> Written structuring	15%	37%	Low	Low signal, low scores — refine or reweight
<b>PMaps eVA</b> Spoken communication	1%	36%	Minimal	Negligible weight — streamline to save time

**Decision rule.** Set the passing bar primarily on Personality Profiler and Attention to Detail (a combined 84% of the score). Treat Sentence Formation and eVA as developmental signals for onboarding, not as gates — they barely move the prediction.

05 · SECTION PERFORMANCE & REFINEMENT

## Two sections carry the signal. Two need tuning.

Average scores across the full cohort reveal a clear gradient. Personality Profiler sits comfortably highest; the language-oriented sections sit lowest — both an insight about the candidate pool and an opportunity to sharpen the instrument.



### Two refinement opportunities

**OPPORTUNITY · A** EFFICIENCY

**Streamline eVA**

At 12 minutes for 3 questions and a 1% loading, the spoken-language section consumes the most time per unit of predictive value of any section. Shortening it — or moving it to a post-offer developmental check — reclaims roughly a third of total test time without weakening the prediction.

**OPPORTUNITY · B** SKILL GAP

**Address Sentence Formation**

A 37% cohort average points to a genuine language-structuring gap in the talent pool. Two paths: simplify the section to better calibrate difficulty, or keep it and pair low scorers with targeted communication skill-building during onboarding.

**Net effect of tuning.** Streamlining eVA and recalibrating Sentence Formation could shrink the battery from 40 minutes toward roughly 28 — faster for candidates, cheaper to administer at volume — while keeping 84%+ of the predictive power intact in Personality Profiler and Attention to Detail.

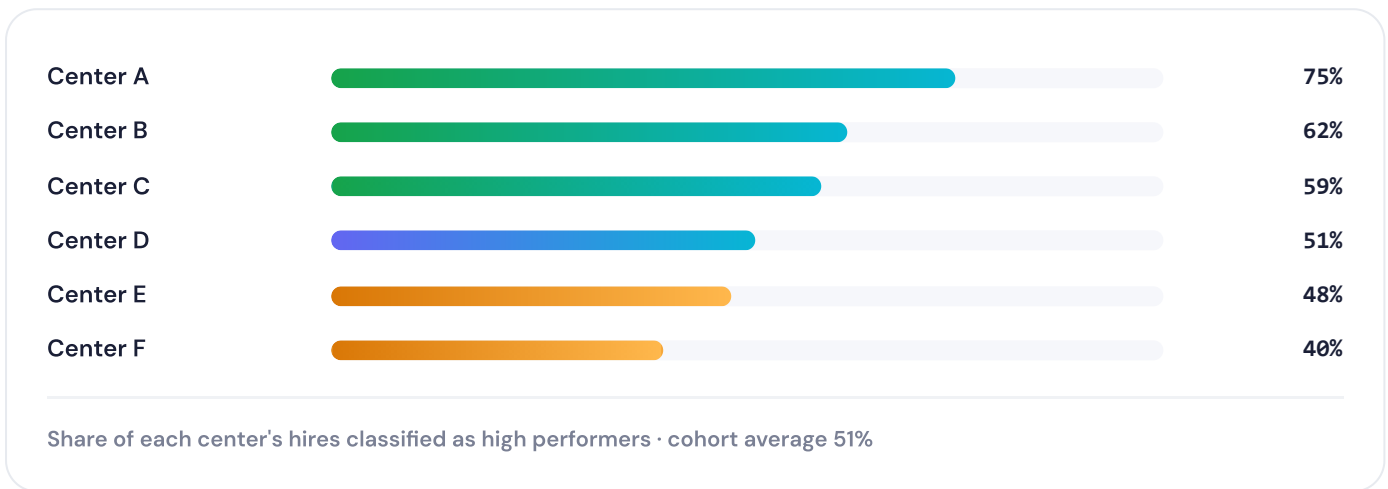
“A shorter, sharper assessment isn’t a weaker one. Cutting time from the lowest-signal sections concentrates the test on what actually predicts performance.”

06 · SOURCING QUALITY

# The same assessment, very different yields.

Because the assessment is validated, its high-performer classification becomes a clean lens on sourcing quality. The spread is dramatic: the best center yields almost **twice** the high-performer rate of the largest one. Where you hire from matters as much as how you screen.

## High-performer yield by center



## Yield tracks real productivity

The classification isn't academic — centers with more high performers also disburse more and book more appointments per agent. Center A leads on every dimension; Center F, the largest source by volume, trails on all of them.

<p><b>CENTER A · TOP YIELD</b></p> <p><b>₹164 L</b></p> <p>Avg disbursement · 75% high performers · 194 avg appointments</p>	<p><b>COHORT AVERAGE</b></p> <p><b>₹97 L</b></p> <p>Avg disbursement · 51% high performers · 114 avg appointments</p>	<p><b>CENTER F · LARGEST VOLUME</b></p> <p><b>₹64 L</b></p> <p>Avg disbursement · 40% high performers · 90 avg appointments</p>
--	---	---

## The vendor view

By staffing vendor, **Vendor 1** supplies the largest share of hires (37.5%) but 60% of them land in the low-performer band — as does **Vendor 2** (18.4% of volume, 60% low). In contrast, **Vendor 3** and **Vendor 4** deliver stronger outcomes, with 58% and 56% high performers respectively. Volume and quality are pulling in opposite directions.

**The sourcing paradox.** The two vendors supplying the most candidates also supply the most low performers. Rebalancing intake toward proven high-yield channels — even modestly — would lift overall hire quality without changing the assessment at all.

## 07 · THE CENTER SCORECARD

# Yield, productivity, and integrity – side by side.

One table to compare every delivery center on the dimensions that drive a sourcing decision: volume, high-performer yield, real disbursement, and assessment integrity. Sorted by high-performer yield, best at the top.

CENTER	N	SHARE	HIGH PERF	AVG DISB	GENUINE
Center A	102	9.1%	75%	₹164L	31%
Center B	84	7.5%	62%	₹137L	49%
Center C	102	9.1%	59%	₹138L	22%
Center D	255	22.8%	51%	₹82L	25%
Center E	223	20.0%	48%	₹103L	30%
Center F	352	31.5%	40%	₹64L	34%
All centers	1,118	100%	51%	₹97L	31%

## What the scorecard says

- Quality and volume are inversely correlated.** The two largest centers — Center F (31.5%) and Center E (20.0%) — post the two lowest high-performer yields. The smallest, Center A and Center B, post the highest.
- Center A is the benchmark.** Highest yield (75%) and highest disbursement (₹164L). Study its sourcing and screening practices and replicate them.
- Center C carries an integrity caveat.** Strong yield (59%) but the lowest "genuine" proctoring rate (22%) — its scores need closer verification before they're trusted at face value (Section 08).

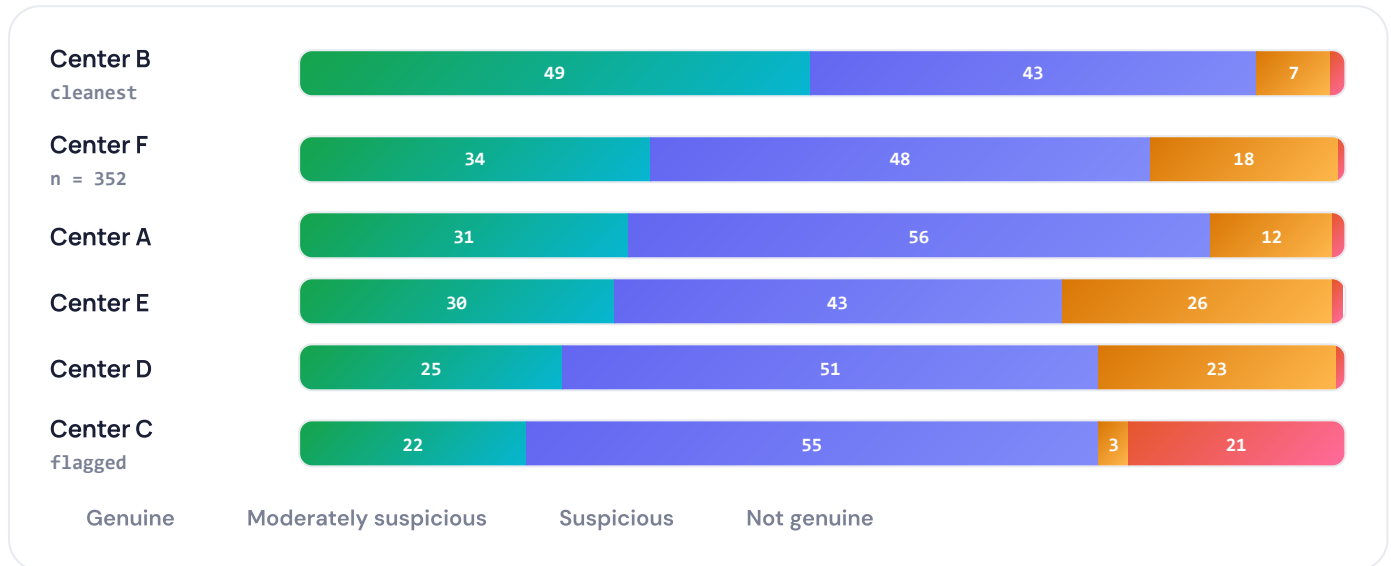
**How to read "Genuine."** The genuine rate is the share of a center's assessments completed with full adherence to proctoring standards. A high yield paired with a low genuine rate is a flag — strong scores that may not be fully trustworthy.

08 · PROCTORING & INTEGRITY

# A valid score is only useful if it's a genuine one.

Predictive validity assumes the scores were honestly earned. Proctoring data lets us check that assumption center by center. Most assessments cluster as genuine or moderately suspicious, but two integrity patterns demand attention.

## Integrity distribution by center



## Two patterns to act on

**FLAG · 01** CRITICAL

### Center C – not-genuine spike

Center C records a **20.6% "not genuine"** rate — by far the highest, against ~1% everywhere else. Its strong high-performer yield (59%) cannot be taken at face value until this is investigated. Urgent corrective action and stricter on-site controls are warranted.

**FLAG · 02** MONITOR

### Centers E & D – suspicious cases

Center E (26%) and Center D (23%) carry the highest **"suspicious"** shares, pointing to process lapses or gaps in proctoring effectiveness. Tighter monitoring and candidate-awareness briefings would lift score trust at both.

**The bright side.** Centers B and F hold the highest genuine counts, showing the proctoring model works when standards are enforced. The fixes here are operational discipline, not a redesign — and they make every downstream validity claim stronger.

09 · RECOMMENDATIONS

# Adopt, refine, operationalize.

The assessment is validated — so the headline recommendation is to **use it with confidence** as a hiring gate. The rest sharpen the instrument and translate its signal into better sourcing and integrity. Grouped by horizon.

## 1 Adopt — act on the validated signal

NOW

- Gate hiring on the proven predictors**  
 Set cut-scores primarily on Personality Profiler + Attention to Detail — a combined 84% of predictive weight.
- Keep ₹45L disbursement as the north star**  
 Re-validate scores against this outcome each cycle so the gate stays honest as the market shifts.
- Raise the bar above the legacy 45%**  
 The validated construct averages 58%. Calibrate a passing threshold to the high-performer score distribution.
- Treat language sections as developmental**  
 Use Sentence Formation & eVA to guide onboarding, not to reject — they barely move the prediction.

## 2 Refine — sharpen the instrument

THIS QUARTER

- Streamline eVA**  
 12 minutes for 1% of weight is poor leverage. Shorten or move to a post-offer check to reclaim ~1/3 of test time.
- Recalibrate Sentence Formation**  
 A 37% average signals a difficulty or pool gap. Re-pitch the section or reweight it.
- Target a ~28-minute battery**  
 A shorter test improves completion and candidate experience without losing predictive power.
- Confirm section weights annually**  
 Re-estimate factor loadings on each new cohort to keep the model current.

## 3 Operationalize — sourcing & integrity

ONGOING

- Rebalance toward high-yield sources**  
 Shift intake toward Centers A, B & C; put Centers E & F on quality-improvement plans.
- Investigate Center C integrity**  
 The 20.6% not-genuine rate needs urgent root-cause review and stricter on-site proctoring.
- Tighten monitoring at Centers E & D**  
 Reduce suspicious-case rates with awareness briefings and live proctoring.
- Replicate the Center A playbook**  
 Document what the top center does on sourcing and screening, and roll it out.

10 · DECISION FRAMEWORK

# From a validated assessment to better hiring outcomes.

The study confirms the assessment works. The opportunity now is to compound that signal — through cut-scores, sourcing discipline, and integrity — into measurable gains. Current baselines and recommended targets below.

## Current → recommended target

LEVER	CURRENT	TARGET	TRAJECTORY	Δ
<b>High-performer yield</b> Share of hires disbursing ≥ ₹45L	51%	65%+		+27%
<b>Predictor concentration</b> Weight in the two proven traits	84%	Hold		keep
<b>Assessment time</b> Battery length per candidate	40 min	~28 min		-30%
<b>Genuine proctoring rate</b> Cohort-wide assessment integrity	31%	60%+		+94%
<b>Not-genuine rate</b> Driven almost entirely by Center C	3%	< 1%		-67%
<b>Low-yield sourcing share</b> Intake from sub-45% centers (Center F)	31%	< 20%		-35%

## The bottom line for hiring decisions

**01**

**Trust the gate**

The assessment predicts performance through job-relevant behavior. It is sound enough to base Tele-Sales hiring decisions on today.

**02**

**Hire from the right places**

Quality varies 40%–75% by source. Rebalancing intake lifts outcomes before a single test question changes.

**03**

**Protect the signal**

Tighten proctoring where flags cluster and re-validate annually so the gate stays trustworthy and current.

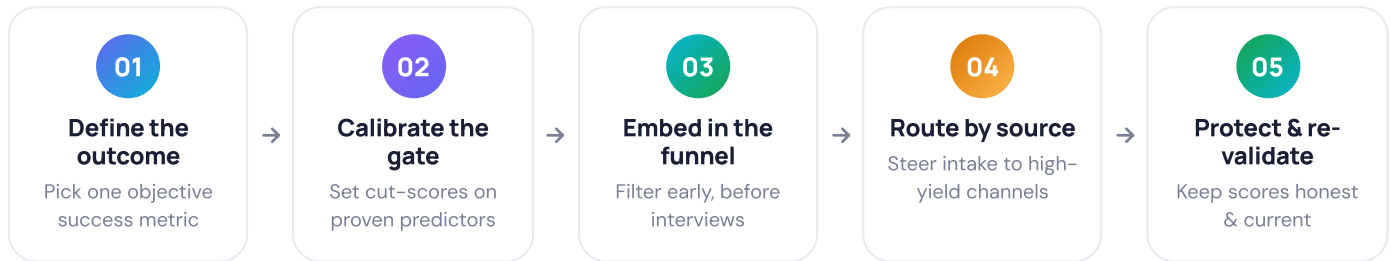
**Targets are recommended, not contractual.** They assume the adopt-refine-operationalize plan ships together and are intended as a tracking frame for quarterly review with the hiring leadership team.

11 · IMPLEMENTING THIS IN YOUR ORGANIZATION

# You've seen the result. Here's the system behind it.

This study didn't validate a single test — it ran a repeatable loop you can stand up inside your own hiring function, for any frontline role. Five stages, each feeding the next, refreshed every hiring cycle. It's a quarter of focused work, not a year-long programme.

The implementation loop — from outcome definition to a self-correcting hiring gate ↻ REPEAT EVERY CYCLE



## How to run each stage

<p><b>01</b></p>	<p><b>Define the outcome</b></p> <p>Choose a single, objective measure of on-the-job success your business already tracks — disbursement, retention, conversion, CSAT. Avoid subjective manager ratings; this metric becomes the ground truth everything is validated against.</p>	<p><b>INPUTS</b> 6–12 months of performance data</p> <p><b>OUTPUT</b> <b>One agreed metric + threshold</b></p>
<p><b>02</b></p>	<p><b>Calibrate the gate</b></p> <p>Score a cohort whose outcomes you already know. Identify which sections actually discriminate high from low performers, and set the passing bar on those — not on equal weighting across every section.</p>	<p><b>INPUTS</b> Assessed historical cohort</p> <p><b>OUTPUT</b> <b>Validated cut-scores</b></p>
<p><b>03</b></p>	<p><b>Embed in the funnel</b></p> <p>Place the assessment early — before interviews — so it filters at volume and protects recruiter time. Automate pass/route logic in your ATS so the gate runs without manual effort on every applicant.</p>	<p><b>INPUTS</b> ATS integration + recruiter SOPs</p> <p><b>OUTPUT</b> <b>Live pre-interview gate</b></p>
<p><b>04</b></p>	<p><b>Route by source</b></p> <p>Tag every hire by channel and track high-performer yield per source. Shift intake toward proven channels and put weak ones on improvement plans — quality often moves before the test changes at all.</p>	<p><b>INPUTS</b> Source-tagged hire outcomes</p> <p><b>OUTPUT</b> <b>Rebalanced sourcing mix</b></p>
<p><b>05</b></p>	<p><b>Protect &amp; re-validate</b></p> <p>Enforce proctoring so scores stay trustworthy, and re-check the score-to-outcome link each cycle. When the role or market shifts, your cut-scores shift with it — and the loop returns to stage 01.</p>	<p><b>INPUTS</b> Proctoring logs + fresh outcomes</p> <p><b>OUTPUT</b> <b>Re-validated model → 01</b></p>

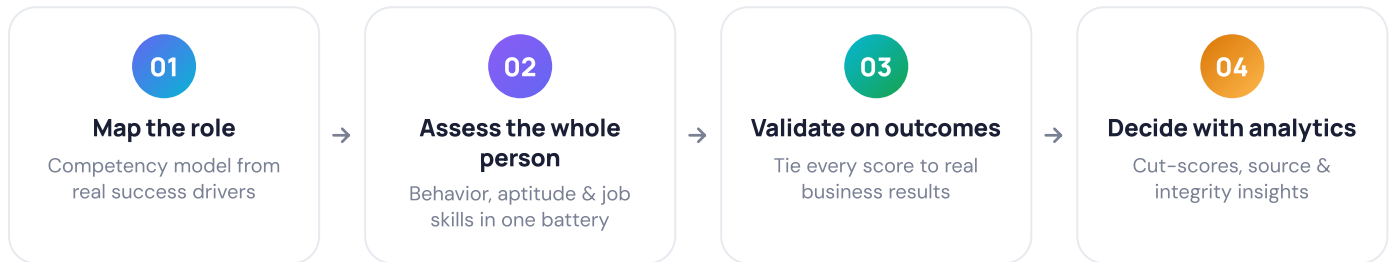
**This is exactly the loop completed in this study.** 1,118 hires, one objective outcome, a validated gate, and a clear read on where to source and where to tighten. Standing the same loop up in your organization turns hiring from a judgment call into a measurable, improving system.

## 12 · WHY PMAPS


# Assessments built to **predict performance** — not just measure traits.

This study isn't a one-off. It's what PMaps does by design: connect what a candidate scores **before** hiring to what they deliver on the job. That outcome-first discipline is why the gate in this report can be trusted — and it's available for every priority role you hire for.


## How PMaps works




## Why hiring leaders choose PMaps




**Predictive, not descriptive**  
Scores correlate with on-the-job performance, so you hire for outcomes instead of gut feel.  
**Proven here: high performers out-scored low on every section.**



**Role-specific science**  
Competency frameworks tuned to each role's true success drivers and weighted by what actually predicts — not equal-weighted guesswork.



**Integrity you can trust**  
AI-assisted proctoring flags suspicious attempts so a passing score is a genuine one — and every downstream decision holds up.



**Analytics that drive action**  
Decision-ready dashboards show where quality concentrates — by competency, source, and center — so you act on signal, not anecdote.

## Make your next hire a **predictable one.**

Validate a hiring gate for your priority roles — the same loop that produced this study, applied to your goals. See it live with our team, or start free today.

[Book a demo →](#)
[Start 7-day free trial →](#)
[Contact us](#)
[ssawant@pmaps.in](mailto:ssawant@pmaps.in)

+91 8591320212

[pmapstest.com](http://pmapstest.com)

# A hiring assessment you can make decisions on.

Across 1,118 hires and a full year of real disbursement data, the PMaps Tele-Sales assessment reliably separated high performers from low — through the right competencies, with 84% of its predictive weight in behavioral fit and attention to detail. The path forward is to adopt it as a gate, sharpen its lowest-signal sections, and source from the channels it shows are strongest.

## STEP 01

### Set the cut-scores

Calibrate the passing bar on Personality Profiler + Attention to Detail and deploy as a hiring gate.

## STEP 02

### Rebalance sourcing

Shift intake toward high-yield centers and place the lowest on quality-improvement plans.

## STEP 03

### Tighten integrity

Investigate Center C, monitor Centers E & D, and re-validate against disbursement annually.

