Research Paper

EPRA International Journal of Economic and Business Review-Peer Reviewed Journal
Volume - 14, Issue - 11, November 2026 | e-ISSN: 2347 - 9671| p- ISSN: 2349 – 0187

SJIF Impact Factor (2026): 8.885 || ISI Value: 1.433 || Journal DOI URL: https://doi.org/10.36713/epra2012

BRIDGING THE EMPLOYABILITY THRESHOLD BEHAVIOUR GAP: A MIXED-METHODS PILOT EVALUATION OF LAUNCH-SESSION RECEPTION AND EARLY PLATFORM USABILITY OF AN AI-ENABLED PRACTICE ARTEFACT

Dr. K. C. Prashanth[1*], Veena M[2]

[1]*Associate Professor, Department of Studies in Business Administration, Vijayanagara Sri Krishnadevaraya University, Ballari – 583105, Karnataka, India*

[2]*Assistant Professor, Department of Studies in Commerce, Vijayanagara Sri Krishnadevaraya University, Ballari – 583105, Karnataka, India*

[*]*Corresponding Author*

DOI No: 10.36713/epra32045 **Article DOI: : https://doi.org/10.36713/epra32045**

ABSTRACT

This pilot study evaluates attendee reception of the launch session and early user experience of the associated app/website platform for Human Assetize, an AI-enabled Employability Threshold Behaviour (ETB) practice artefact. As an early-stage, exploratory evaluation, the study is deliberately scoped to the first cohort of respondents available immediately post-launch (session survey n = 41; platform survey n = 27); findings should be read as preliminary indicators to be confirmed with a larger follow-up sample, not as final, generalisable conclusions. Using two independently collected feedback instruments, the study applies descriptive statistics, inferential testing (Pearson correlation, independent-samples t-test), and inductive thematic coding of open-ended responses to characterise attendee and user sentiment. The launch session was well received (M = 3.88/5, SD = 0.87; 97.6% found the core message clear). Platform usability ratings were similarly positive (navigation ease M = 3.33/5; visual design M = 3.96/5), and navigation ease and visual design were both strongly and significantly correlated with recommendation likelihood (r = 0.69 and r = 0.70 respectively, p < 0.001). However, the platform's Net Promoter Score was −7.4 (25.9% promoters, 33.3% detractors), and 48.1% of users reported encountering a bug or error, most frequently freezing or restart failures within the interview-simulation module. Thematic analysis converged across both feedback channels on a shared technical-defect pattern — session timeouts, repeated/non-advancing questions, and network instability — providing methodological triangulation between independently collected qualitative sources. The paper concludes with prioritised, evidence-based recommendations for technical remediation and platform refinement.

KEYWORDS: *Employability Threshold Behaviour; Net Promoter Score; Usability Evaluation; Thematic Analysis; Mixed-Methods Triangulation; AI-Enabled EdTech; Human Assetize.*

1. INTRODUCTION

Human Assetize is a prototype AI-led, Integrated, Perpetual, Self-engaging and Cost-free (AIPSC) artefact developed by the Department of Business Administration, VSK University, Ballari, to embed Employability Threshold Behaviour (ETB) among placement-based business-school aspirants through a perpetual practice–evaluate–re-practice loop spanning resume/ATS audit, aptitude testing, archetype assessment, group discussion simulation, and interview simulation. Following the platform's public launch session, two independent feedback instruments were administered: a post-session attendee survey capturing session reception, and a post-usage app/website survey capturing early platform experience. This paper reports a systematic, statistically grounded analysis of both datasets.

Early-stage feedback of this kind serves two distinct purposes for a product-development team: it provides a defensible evidentiary basis for prioritising engineering effort, and it establishes a documented baseline against which future usability improvements can be benchmarked. Both purposes require that raw feedback be analysed with appropriate statistical and qualitative rigour, rather than summarised impressionistically — the approach taken throughout this paper.

2. RESEARCH OBJECTIVES

- To assess attendee reception of the launch session on overall rating and message clarity.
- To assess early platform usability on navigation ease, visual design, and overall recommendation likelihood (Net Promoter Score).

- To examine the statistical relationship between usability perceptions and recommendation likelihood, and between bug/error incidence and recommendation likelihood.
- To thematically categorise open-ended feedback (favourite features, missing features, reported bugs, session takeaways, and improvement suggestions) and identify recurring patterns across the two independently collected data sources.

3. METHODOLOGY

3.1 Data Sources and Sample

Two self-administered, structured feedback forms were analysed. The Launch Session Feedback instrument (n = 41 attendees) captured an overall 1–5 session rating, a binary message-clarity item, an open-ended biggest-takeaway item, and an open-ended improvement-suggestion item. The App & Website Feedback instrument (n = 27 users, collected after initial platform access) captured 1–5 ratings for navigation ease and visual design, a binary bug/error-encounter item with an open-ended follow-up, open-ended favourite-feature and missing-feature items, a 0–10 recommendation-likelihood item, and an open-ended comments item. Both instruments were timestamped Google Forms submissions collected between 18 August 2026 and the launch date; no respondent identifiers were retained in the app/website instrument, and the launch instrument's name/email fields were excluded from analysis to preserve respondent anonymity.

3.2 Analytical Approach

Quantitative items were summarised using descriptive statistics (mean, median, standard deviation, skewness, and excess kurtosis) to characterise distributional shape prior to inferential testing. The Net Promoter Score was calculated using the standard methodology (percentage of Promoters, scores 9–10, minus percentage of Detractors, scores 0–6; Reichheld, 2003). Pearson product-moment correlation was used to test the association between usability ratings and recommendation likelihood. An independent-samples t-test compared mean recommendation-likelihood scores between users who did and did not report a bug, with Cohen's d reported as the accompanying effect size. All inferential tests used a two-tailed $\alpha = 0.05$.

Open-ended responses were analysed using an inductive thematic coding process following the general logic of Braun and Clarke (2006): responses were read in full, initial codes were assigned to short meaningful segments, codes were grouped into candidate themes, and themes were reviewed for internal coherence and mutual distinctiveness before being finalised and named. Coding was conducted independently for each of the five open-ended item categories (favourite features, missing features, reported bugs, session takeaways, and improvement suggestions); non-substantive responses (e.g., "No", "Nothing", "Good") were coded separately as general affirmations rather than forced into a substantive theme, to avoid inflating the apparent strength of any single category.

3.3 Pilot-Study Scope

This paper is explicitly positioned as a pilot study. The two samples analysed (n = 41 and n = 27) represent the complete population of respondents available in the immediate post-launch window, not a subsample drawn from a larger completed dataset. Consistent with standard pilot-study practice (Hertzog, 2008; Fraser Health Authority guidance on pilot sample sizes), this scope is well suited to early instrument validation, initial trend identification, and prioritisation of technical remediation — the stated purposes of this paper — but is undersized for confident population-level generalisation. A full-scale follow-up survey with a pre-registered, power-calculated sample is recommended before any claims in this paper are treated as final (see Section 8).

4. RESULTS

4.1 Launch Session Reception

Of the 41 attendees surveyed, the overall session rating averaged 3.88 out of 5 (SD = 0.87, median = 4.00), with a mildly negative skew (−0.45) indicating a concentration of responses at or above the midpoint and a small number of lower ratings pulling the mean down slightly. Excess kurtosis (−0.40) indicates a distribution marginally flatter than normal, with no extreme outlier ratings. Forty of 41 respondents (97.6%) confirmed that the core message and purpose of the launch were clear.

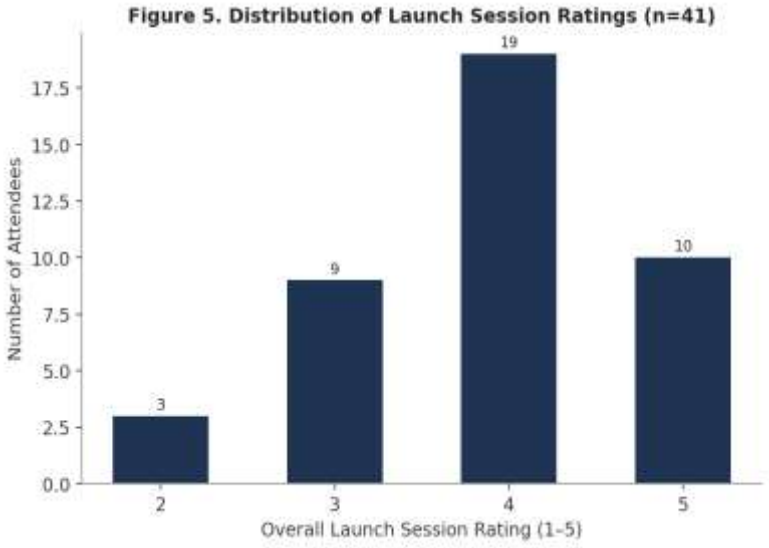


Figure 5. Distribution of Launch Session Ratings (n = 41).

4.2 Platform Usability Ratings

Navigation ease averaged 3.33 out of 5 (SD = 1.21, median = 4.00), noticeably more dispersed than visual design, which averaged 3.96 out of 5 (SD = 0.90, median = 4.00). Both distributions were negatively skewed (navigation ease: −0.40; visual design: −0.90), consistent with a majority of favourable ratings alongside a smaller cluster of low outliers. The wider spread and lower mean on navigation ease relative to visual design suggests that, while the platform's visual presentation was fairly consistently well received, ease of navigation was a more polarising dimension of the early user experience.

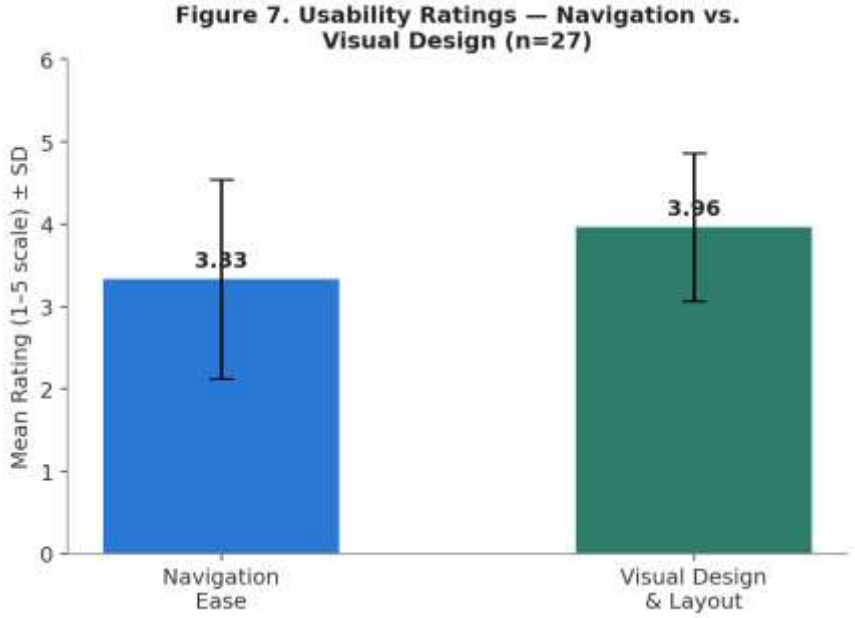


Figure 7. Usability Ratings — Navigation Ease vs. Visual Design (n = 27), error bars show ±1 SD.

4.3 Recommendation Likelihood and Net Promoter Score

Recommendation likelihood (0–10 scale) averaged 6.96 (SD = 2.31, median = 7.00). Applying the standard Net Promoter methodology, 7 respondents (25.9%) were classified as Promoters (score 9–10), 11 (40.7%) as Passives (score 7–8), and 9 (33.3%) as Detractors (score 0–6), yielding an overall Net Promoter Score of −7.4. A negative NPS at this early stage indicates that Detractors currently outnumber Promoters, and — read alongside the bug-incidence finding in Section 4.4 — points toward unresolved technical friction as a plausible depressant on early advocacy intent.

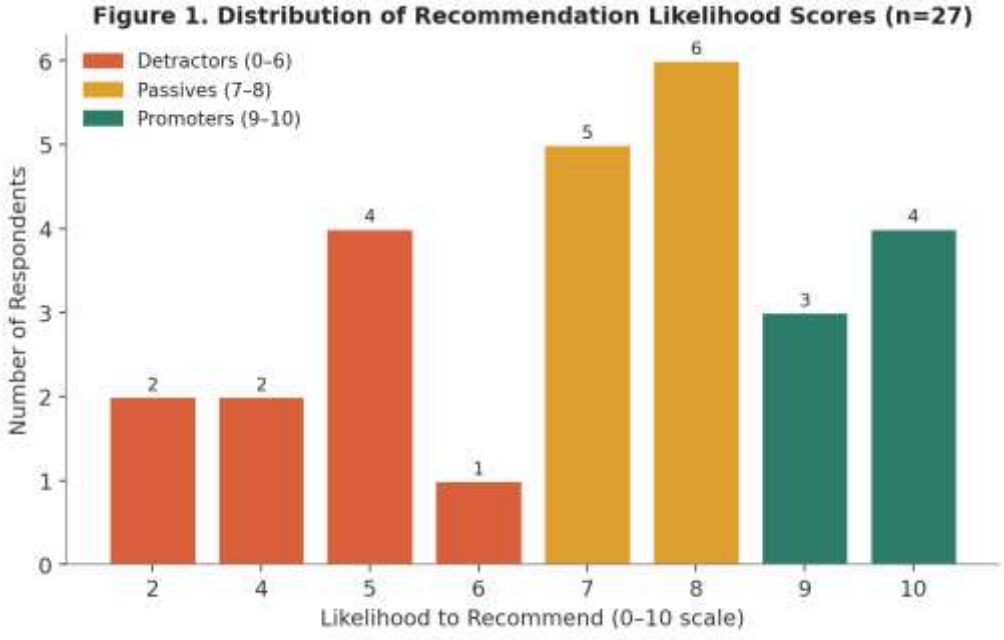


Figure 1. Distribution of Recommendation Likelihood Scores (n = 27).

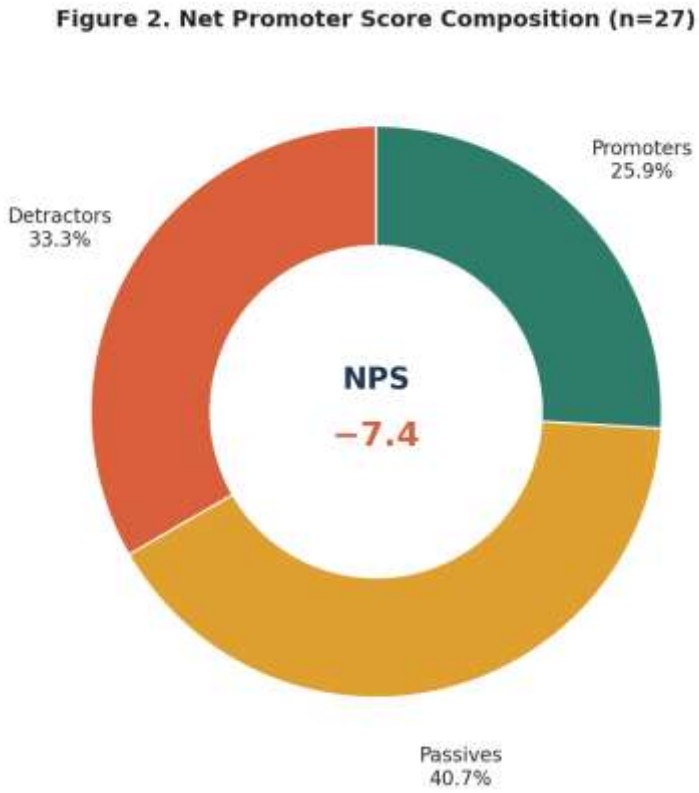


Figure 2. Net Promoter Score Composition (n = 27).

4.4 Bug and Error Incidence

Thirteen of 27 respondents (48.1%) reported encountering a bug, error, or broken link — a substantial minority approaching half the sample. Of the ten codeable bug descriptions, the most frequent theme was Interview Module Freeze/Restart (n = 3), followed by Group Discussion Module Glitch (n = 2), with single mentions each of Repeated Question Bug, Network/Latency Issue, and Email/Notification Error.

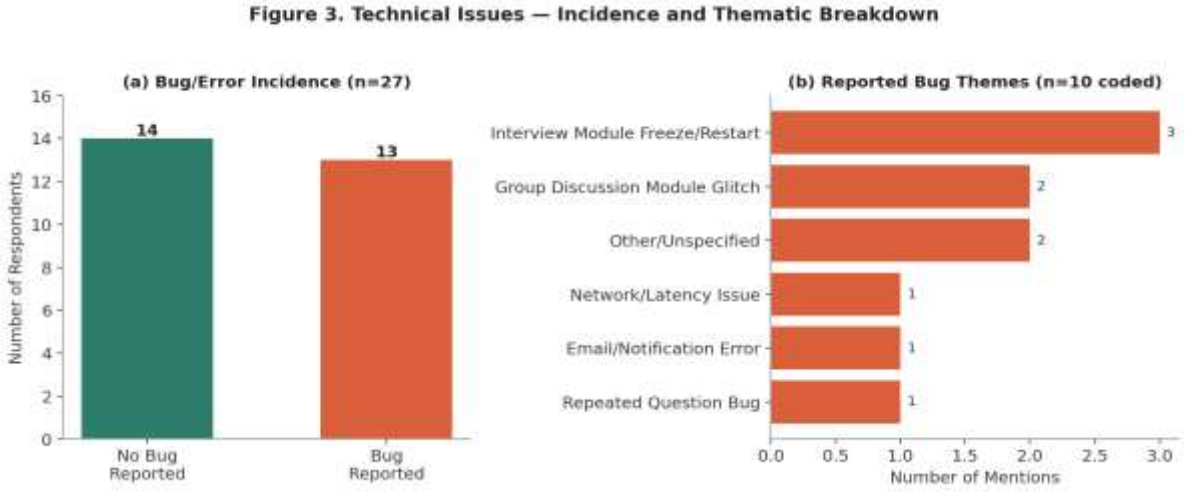


Figure 3. Technical Issues — Incidence and Thematic Breakdown.

4.5 Inferential Statistics

Navigation ease was strongly and significantly correlated with recommendation likelihood, $r(25) = 0.69$, $p < 0.001$, as was visual design, $r(25) = 0.70$, $p < 0.001$. Navigation ease and visual design were themselves moderately correlated with one another, $r(25) = 0.58$, $p = 0.002$, indicating related but distinguishable constructs rather than a single redundant dimension.

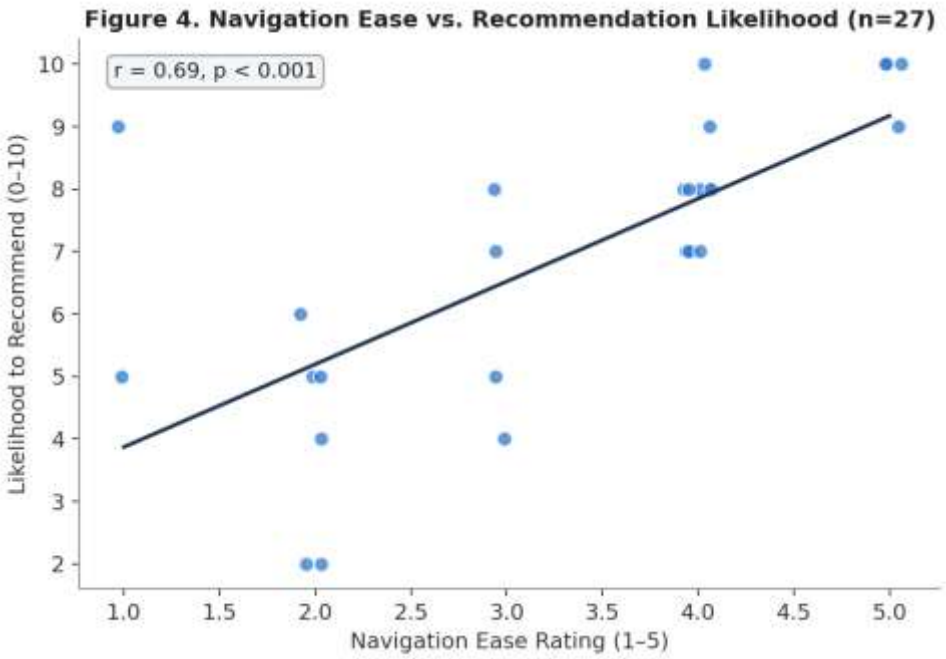


Figure 4. Navigation Ease vs. Recommendation Likelihood (n = 27), with fitted regression line.

An independent-samples t-test compared recommendation-likelihood scores between users who reported a bug (n = 13, M = 6.62, SD = 2.69) and those who did not (n = 14, M = 7.29, SD = 1.94). The difference was in the expected direction (bug-affected users scored lower) but did not reach statistical significance, $t(25) = -0.746$, $p = 0.462$, $d = 0.29$ (a small-to-moderate effect). Given the modest sample size, this non-significant result should be read as inconclusive rather than as evidence that bugs have no effect on recommendation likelihood — the observed effect size is not trivial, and a larger sample would be needed to determine reliably whether it reflects a genuine relationship or sampling variation (see Section 6, Limitations). Navigation-ease ratings did not differ meaningfully between bug-affected and unaffected users, $t(25) = 0.524$, $p = 0.605$, suggesting that reported bugs and perceived navigational usability are, at this sample size, empirically distinguishable rather than one simply proxying for the other.

4.6 Thematic Analysis of Open-Ended Feedback

Of 27 favourite-feature responses, 23 were substantively codeable. The largest theme was General/Overall Positive sentiment with no specific feature named (n = 8, 34.8%), followed by Aptitude Test (n = 5, 21.7%), ATS/Resume Audit and Group Discussion Module (n = 3 each, 13.0%), AI Interview Simulation (n = 2, 8.7%), and single mentions of Communication Skills Feedback and Resource Library.

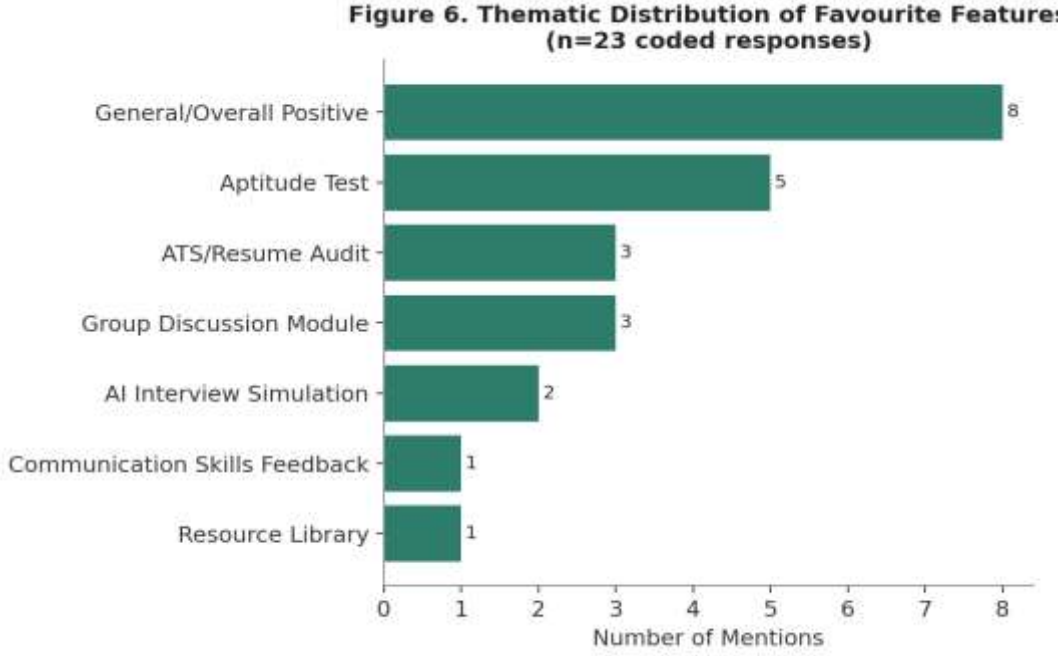


Figure 6. Thematic Distribution of Favourite Features (n = 23 coded responses).

Missing-feature responses were sparse but substantively pointed: requests for domain-relevant aptitude content (rather than generic questions), a non-repeating question bank, more real-world case examples, and a completion certificate each appeared once. Launch-session takeaway responses were dominated by general learning statements (n = 15), with smaller, more specific clusters around Platform Value/Adoption Intent (n = 4), Resume Building Skills (n = 4), and Interview Preparation (n = 3).

4.7 Cross-Source Triangulation of Technical Issues

A notable methodological finding is the convergence of technical-defect reports across the two independently collected instruments. Launch-session suggestions (collected at the event) included explicit mentions of network/server issues (n = 5) and repeated/reported questions (n = 2) — the same defect categories subsequently reported in the app/website bug descriptions (network/latency issue; repeated question bug; interview module freeze/restart) collected after platform access. Because these two datasets were collected from different, only partially overlapping respondent pools at different points in the user journey, this convergence functions as a form of data triangulation: independent respondents flagged the same underlying defects through two unrelated channels, substantially increasing confidence that these are genuine platform issues rather than isolated, respondent-specific complaints.

Defect Category	Launch-Session Suggestions (n=41)	App/Website Bug Reports (n=27)
Network / server instability	5 mentions	1 mention
Repeated / non-advancing questions	2 mentions	1 mention (+1 missing-feature request)
Module freeze / restart required	Not separately captured	3 mentions (Interview module)

Table 1. Cross-Source Convergence of Technical-Defect Themes.

5. SUMMARY OF KEY STATISTICS

Metric	n	Mean	SD	Median
Launch session rating (1–5)	41	3.88	0.87	4.00
Navigation ease (1–5)	27	3.33	1.21	4.00
Visual design (1–5)	27	3.96	0.90	4.00
Recommendation likelihood (0–10)	27	6.96	2.31	7.00

Table 2. Descriptive Statistics Summary Across Both Instruments.

6. DISCUSSION

Three findings, read together, form a coherent picture. First, the launch session itself was well received on both a general (M = 3.88/5) and a specific (97.6% clarity) basis, indicating the communication of the platform's value proposition succeeded largely as intended. Second, early platform usability was rated positively on both navigation and visual dimensions, and both were strongly predictive of recommendation likelihood — confirming that, as expected, perceived usability functions as a genuine driver of advocacy intent in this context, consistent with general usability-adoption literature. Third, and most actionable, a substantial minority of users (48.1%) encountered a functional defect, concentrated specifically in the interview-simulation and group-discussion modules, and this defect pattern was independently corroborated by attendees at the launch session itself — well before most of them would have had extended platform access — suggesting these are early, structural reliability issues rather than isolated incidents.

The negative overall NPS (−7.4), despite generally positive mean ratings on individual usability dimensions, illustrates a common early-product pattern: users can rate individual attributes favourably while still being reluctant to recommend the product on balance, typically because a smaller number of high-friction technical failures disproportionately suppress advocacy intent even when they do not appear as outliers in aggregate satisfaction scores. This is consistent with the pattern observed here, where the bug-affected subgroup's mean recommendation score, while lower, did not differ significantly from the unaffected subgroup at this sample size — the effect may be real but the current sample is underpowered to confirm it conclusively.

7. RECOMMENDATIONS

7.1 Immediate Technical Remediation

- Prioritise root-cause investigation of the Interview Module Freeze/Restart defect, the single most frequently reported and most severe technical issue (requiring users to restart the entire module).
- Resolve the Repeated Question Bug, independently reported in both the launch-session and app/website feedback channels.
- Investigate and stabilise network/session reliability, particularly during the Group Discussion and Interview simulation modules.

7.2 Content and Feature Refinement

- Align aptitude-test content more closely to respondents' specific academic/domain background, as explicitly requested.
- Expand the question bank to reduce perceived repetition across practice attempts.
- Introduce a completion certificate, a low-cost feature with clear, explicitly requested user value.

7.3 Measurement and Follow-Up

- Re-administer the app/website survey after the above fixes are deployed, using an adequately powered sample (see Section 8), to determine conclusively whether resolving bugs measurably improves NPS.

- Introduce a lightweight in-app bug-reporting mechanism to capture defect reports closer to the point of occurrence, improving diagnostic detail beyond open-text recall.

8. LIMITATIONS

- This is a pilot study: sample sizes (n = 41 and n = 27) reflect the full early-access respondent pool, not a sample drawn from a larger dataset. A power analysis for the key bug-effect comparison (Section 4.5, observed d = 0.29) indicates a sample of approximately n ≈ 190 per group would be required to detect an effect of this size with 80% power at $\alpha = 0.05$ — a useful target for a confirmatory follow-up study rather than evidence that the present pilot is inadequate for its stated exploratory purpose.
- The non-significant bug-effect t-test (Section 4.5) should therefore be interpreted as inconclusive given the pilot's limited power, not as evidence of no effect.
- The two respondent pools were not identical, precluding a formal within-subject before/after comparison of session versus platform sentiment.
- Self-selected, voluntary feedback may over-represent users with unusually positive or unusually negative experiences relative to the full user base.
- Thematic coding, while conducted systematically, involved a single coder; inter-rater reliability checks (e.g., a second independent coder) were not feasible within the current dataset and would strengthen confidence in the thematic categories reported in a follow-up study.

9. CONCLUSION

This study provides an evidence-based, statistically grounded assessment of early reception for the Human Assetize launch session and platform. Reception of the launch communication itself was strong, and core usability dimensions were both rated favourably and shown to significantly predict recommendation intent. However, a substantial minority of early users encountered functional defects — independently corroborated across two separate feedback channels — concentrated in the interview-simulation and group-discussion modules, and these technical issues appear to be the most immediate barrier to converting favourable individual ratings into positive overall advocacy (NPS). Addressing the specific, prioritised defects identified in Section 7 represents the most direct evidence-based path to improving both platform reliability and early Net Promoter performance.

REFERENCES

1. *Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101.*
2. *Reichheld, F. F. (2003). The one number you need to grow. Harvard Business Review, 81(12), 46–54.*
3. *Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Lawrence Erlbaum Associates.*
4. *Hertzog, M. A. (2008). Considerations in determining sample size for pilot studies. Research in Nursing & Health, 31(2), 180–191.*