Validity & Criterion Defense Memo

Nicole Salem

PSY 720

Crystal Connors (Ph.D)

April 3, 2026

Memorandum

To: Executive Leadership Team, SilverLine Health Technologies

From: Nicole Salem, External I-O Psychology Consultant

Date: April 3, 2026

Subject: Evaluation of Selection System for Healthcare Implementation Consultants

Executive Summary

This report evaluates the scientific and legal defensibility of SilverLine Health Technologies' proposed selection system for Healthcare Implementation Consultants. A defensible system must demonstrate not only predictive capability but also consistency, accuracy, and fairness in decision-making.

The proposed system incorporates established selection tools, a cognitive ability assessment, a personality inventory, and structured interviews, yet several limitations introduce measurable risk. The current approach does not fully account for measurement error in individual-level decisions, interview evaluations show inconsistent reliability, and internal validation and documentation practices remain underdeveloped. These issues increase the likelihood of inaccurate hiring decisions and elevate exposure to legal challenge.

The system shows promise, but refinement is necessary. I recommend approval contingent upon targeted revisions that strengthen measurement precision, standardization, and validation practices.

Measurement Error and Decision Risk

All psychological assessments contain measurement error, meaning that observed scores reflect an estimate rather than a precise representation of an individual's true standing (SIOP,

2018). The standard error of measurement (SEM) provides a practical way to interpret this uncertainty by defining a range within which a candidate's true score likely falls (Highhouse et al., 2016).

In this case, an SEM of 5.1 for the cognitive ability assessment suggests that individual scores should be interpreted with caution. Candidate X (70) and Candidate Z (72) fall within overlapping score ranges, indicating no meaningful difference in underlying ability. Candidate Y (78) appears stronger, though the margin of error still limits how confidently that difference can be interpreted.

The cutoff score of 65 introduces additional complexity. Scores near this threshold may fall within the measurement error band, meaning that acceptance or rejection decisions could hinge on differences that are not statistically meaningful. SEM is specifically intended to inform whether score differences support defensible decisions (Highhouse et al., 2016).

This becomes particularly relevant if a hiring decision is challenged. A rejected candidate could reasonably argue that the decision was based on unstable score differences rather than true performance potential. When score interpretation does not account for measurement error, the decision process becomes more difficult to justify.

Reliability and Standardization

Reliability reflects the consistency of measurement and serves as a foundation for defensible selection decisions. The cognitive ability assessment demonstrates acceptable reliability, with internal consistency of .84 and test–retest reliability of .77, indicating relatively stable measurement.

The structured interview presents a different pattern. An interrater reliability of .54 indicates notable inconsistency across interviewers, suggesting that candidate evaluations may

vary depending on who conducts the interview. In applied settings, this level of variability introduces uncertainty into hiring decisions and weakens confidence in the results. Rater disagreement is a well-established source of measurement error in observational methods (Highhouse et al., 2016).

Observed interview practices further contribute to this inconsistency. Interviewers report discussing impressions prior to scoring, departing from structured prompts, and relying on general impressions such as perceived charisma. These behaviors shift the process away from standardized evaluation and increase the likelihood of bias.

The personality inventory shows moderate reliability, though its use is not supported by internal validation data. Professional standards emphasize that reliability evidence should be established within the context in which the tool is used, particularly when decisions affect individual outcomes (SIOP, 2018).

Taken together, the cognitive assessment provides relatively consistent measurement, while the interview introduces substantial variability. Even moderate inconsistencies matter in selection contexts because decisions are made at the individual level rather than across groups (Highhouse et al., 2016).

Validity and Professional Standards

Validity concerns the extent to which a selection procedure predicts relevant job performance outcomes. The cognitive ability assessment demonstrates a moderate relationship with performance ($r = .39$), indicating practical predictive value. The structured interview shows a weaker relationship ($r = .24$), and the personality inventory relies on publisher-reported estimates without internal validation.

Measurement theory establishes that reliability places an upper limit on validity (Highhouse et al., 2016). Because the structured interview demonstrates relatively low reliability, its predictive contribution is constrained. This is especially important given its substantial weight in the overall decision process.

Professional standards, as outlined in the Principles for the Validation and Use of Personnel Selection Procedures, require organizations to establish multiple forms of validity evidence and maintain clear documentation supporting the use of each selection method (SIOP, 2018). This includes demonstrating that assessments are job-related, supported by empirical evidence, and applied consistently.

SilverLine has taken initial steps by examining the validity of cognitive tests; the evidence supporting the personality inventory and interview remains limited. This gap becomes more consequential if subgroup differences in hiring outcomes emerge. In those cases, organizations must demonstrate that selection procedures are both job-related and necessary for business operations.

At present, the system demonstrates limited predictive value and lacks the depth of validation required for strong defensibility.

Legal Defensibility Analysis

Several elements of the proposed system present defensibility concerns. Interview administration lacks consistency, increasing the likelihood of unequal treatment across candidates. Documentation practices are also limited, with no formal rationale for weighting selection components and inconsistent records of scoring decisions.

The connection between predictors and job performance varies across tools. The cognitive assessment provides moderate support, while the personality inventory and interview

lack sufficient internal validation. The absence of adverse impact analysis introduces additional risk, particularly given the organization's diverse workforce and established expectations for monitoring fairness (SIOP, 2018).

Score interpretation practices also present challenges. Decisions made near cutoff thresholds do not currently account for measurement error, which weakens the justification for accepting or rejecting candidates based on small score differences.

Among the components, the structured interview represents the greatest vulnerability. Its lower reliability, combined with inconsistent administration and limited validation evidence, makes it the most difficult to defend if challenged.

Executive Recommendation

The proposed system reflects a solid conceptual foundation but requires refinement before implementation. Proceeding without adjustment would introduce avoidable risk in both decision quality and legal defensibility.

The most immediate priority is strengthening the structured interview process. This includes formalizing interviewer training, enforcing consistent use of structured questions, and requiring independent scoring before any discussion occurs. Clearer behavioral anchors and ongoing monitoring of rater agreement will further improve consistency.

Internal validation should also be expanded. Establishing predictor–criterion relationships for the personality inventory and interview within SilverLine's workforce will strengthen the evidence base supporting their use. Incorporating SEM into score interpretation, particularly for candidates near cutoff thresholds, will improve both accuracy and fairness in decision-making.

Documentation practices should be formalized to support transparency and defensibility. This includes developing a comprehensive validation report, clearly articulating weighting

decisions, and conducting adverse impact analyses to monitor outcomes across demographic groups.

With these adjustments, the selection system can support more consistent, accurate, and defensible hiring decisions. In its current form, it reflects strong intent but requires additional rigor to meet professional and legal standards.

References:

Highhouse, S., Dalal, R. S., & Salas, E. (2016). *Essentials of Personnel Assessment and Selection* (2nd ed.). Routledge.

Society for Industrial and Organizational Psychology. (2018). *Principles for the validation and use of personnel selection procedures* (5th ed.). Author.

Sackett, P. R., & Lievens, F. (2008). Personnel selection. *Annual Review of Psychology, 59*, 419–450.https://doi.org/10.1146/annurev.psych.59.103006.093716