MindBridge

mindbridge.ai

Technical Whitepaper

Assertion-Level Risk Scoring:

A Technical Validation Guide for Audit Firms

MindBridge Analytics Inc.

Version: July 2026

Table of Contents

Executive Summary

Risk assessment at the assertion level is a requirement of every major auditing standard. ISA 315 (Revised 2019), PCAOB AS 2110, and the AICPA's AU-C Section 315 each require auditors to identify and assess risks of material misstatement at the assertion level — understanding not just that an account may be misstated, but which specific assertion is at risk and why. Executing that requirement well has always depended on the quality of the auditor's understanding of the entity, its environment, and the transactions that flow through its general ledger.

MindBridge Assertion-Level Risk Scoring (May 2026 release) is designed to augment that process. By applying a library of control point algorithms to every transaction in the general ledger and surfacing the results at the assertion level, the platform gives auditors a data-driven starting point for their risk assessment — one that is grounded in the full population of GL activity rather than a subset, and that disaggregates risk by assertion rather than presenting a single composite score. This is AI-assisted risk assessment: the platform surfaces patterns and concentrations in the data that inform auditor judgment; the auditor applies that judgment to reach the risk assessment conclusions that the standards require.

This document is intentionally audit-standard neutral. The scoring mechanics described in Sections 3 and 4 do not map to any single standard's risk assessment framework — they are designed to produce assertion-level evidence that is relevant and usable regardless of whether a firm opines against those various standards. **How a firm integrates that evidence into its risk assessment methodology, how it documents reliance on platform scores, and how it trains engagement teams to interpret the output are all firm-level methodology decisions.**

What this document is unambiguous about is the role of human judgment. Assertion-level scores are an input to the auditor's risk assessment process, not a substitute for it. The platform does not assess inherent risk, rather it surfaces transaction-level signals that are relevant to inherent risk assessment. It does not reach conclusions about where material misstatement is likely, it quantifies where anomalous patterns are concentrated. The auditor reads those signals, applies their understanding of the entity and its environment, exercises professional skepticism, and reaches the risk assessment conclusions that the engagement requires. Nothing in this document should be read as suggesting otherwise.

The assertion-level risk scores that power the MindBridge platform are the output of a six-stage pipeline.

The authoritative reference for which control points are active and which assertions they feed is the mapping matrix, reproduced in full in Appendix A.

Purpose and Scope of This Document

The MindBridge client-facing whitepaper describes the platform's capabilities at a level appropriate for audit firm leaders evaluating whether to adopt the tool. This companion technical whitepaper is different in purpose and audience. It is written for the central methodology team whose sign-off is required before engagement teams can rely on assertion-level scores in audit documentation.

The methodology team's role is to answer a specific question: can an auditor point to a MindBridge assertion score in their working papers as evidence supporting a risk assessment decision, and will that reliance survive a regulatory, or firm quality review? Answering that question requires understanding not just what the platform produces, but how and where the current version supports the firm's process and workflows for risk assessments.

What this document covers

- The six-stage technical pipeline from raw GL data to assertion-level percentage scores
- The control point-to-assertion mapping matrix, including which control points are currently inactive
- The three-band point system and demerit point accumulation mechanics
- A structured validation and sign-off framework for methodology teams

What this document does not cover:

- Any client-facing user interface or heatmap visualizations
- Any LLM-driven or agentic workflow layers
- The risk-materiality four-quadrant matrix and procedure recommendations
- Integration with any working paper solutions.

How the Demerit Scoring Pipeline Works

Every assertion-level risk score the platform displays is the output of six sequential stages using a Demerit Scoring system. *Note: the current product does not use the term 'demerit point' but instead refers to it plainly as 'point'.*

Stage	What happens
1 — Assertion mapping	Each MAC Level 3 account category is linked to its relevant financial statement assertions via the Mapping matrix
2 — Control point linkage	Each active control point is linked to the assertions it is relevant to, per the Mapping matrix
2a — Three-band bucketing	Each control point's raw 0–100% score is converted to 0, 1, or 2 demerit points via threshold bands
3 — Demerit accumulation	Demerit points are summed per entry across all CPs in the ensemble
4 — Ensemble aggregation	Entry-level demerit totals are aggregated across the account population to produce an assertion-level percentage
5 — Population and configuration	Account population size and Library Settings configuration shape how scores should be interpreted and applied

Stage 1

The assertion mapping table

Before any transaction is scored, each MAC Level 1 account category is linked to the financial statement assertions relevant to it. This pre-built mapping governs which control points activate for each account type. The full Mapping matrix is reproduced in Appendix A; the discussion below draws directly from it.

The Mapping matrix covers five account categories — Asset, Liability, Equity, Revenue, and Expense — each with five assertion assigments, giving 25 account-assertion pairs in total. Balance Sheet accounts use Existence, Completeness, Rights and Obligation, Valuation/Accuracy, and Presentation; Income Statement accounts use Occurrence, Completeness, Accuracy, Cut-off, and Presentation/Classification.

The count of active control points per assertion, taken directly from the COUNT row, is a critical input to understanding score reliability:

Account category	Assertion	Active CPs	Max demerit points per entry
Asset	Existence	6	12
Asset	Completeness	8	16
Asset	Rights and Obligation	4	8
Asset	Valuation / Accuracy	9	18
Asset	Presentation	5	10
Liability	Existence	8	16
Liability	Completeness	7	14
Liability	Rights and Obligation	4	8
Liability	Valuation	5	10
Liability	Presentation	5	10
Equity	Existence	8	16
Equity	Completeness	8	16
Equity	Rights and Obligation	1	2
Equity	Valuation	8	16
Equity	Presentation	5	10
Revenue	Occurrence	8	16
Revenue	Completeness	7	14
Revenue	Accuracy	8	16
Revenue	Cut-off	5	10
Revenue	Presentation / Classification	4	8
Expense	Occurrence	8	16

Account category	Assertion	Active CPs	Max demerit points per entry
Expense	Completeness	7	14
Expense	Accuracy	5	10
Expense	Cut-off	9	18
Expense	Presentation / Classification	4	8

Stage 2

Control point-to-assertion linkage

The Mapping matrix contains 39 named control points. Of these, 12 are currently inactive. While they are available for mapping by a user, but the default configuration is that they do not have any active assertion assignments based on what we have observed during the research and testing phase on the datasets analysed.

The remaining 27 active control points vary significantly in how broadly they are deployed across the 25 assertion. Outlier Anomaly is active in 14 of 25 assertion assignments; Manual Entry in 12; Suspicious Keyword in 12. At the other extreme, Weekend Post is active in a single assignment (Expense Cut-off).

Several control points may overlap such as: Outlier Anomaly and Rare Flow frequently co-occur on the same entries; Unusual Digit Sequence and Unusual Digit Combination address similar signals via different statistical approaches. Where two control points measure overlapping signals within the same ensemble, their joint inclusion means a single underlying event generates demerit points from both. Users can use Library Settings to remove the any redundant signal during their validation stage.

Stage 2a

The three-band point system: how raw scores become demerit points

Before any accumulation or aggregation occurs, each control point's raw 0–100% score (based on how MindBridge currently determines anomaly score for each type of control point) is assigned a discrete demerit point value via three default bands:

Raw control point score	Demerit points assigned	Design rationale

0 – 24.99%	0 points	Treated as noise or little contribution to risk — the algorithm may have detected a marginal signal but not enough to constitute meaningful risk evidence. Non-firing CPs do not penalise the score.
25 – 74.99%	1 point	Moderate signal — the algorithm detected a pattern consistent with the assertion risk but not at high confidence. Partial evidence is captured rather than discarded.
75 – 100%	2 points	Strong signal — the algorithm detected a pattern at high confidence. Maximum contribution to the entry's demerit total.

The design rationale for this three-band system is explicit under a traditional weighted-average approach, a control point that never fires for a particular client would dilute the overall score. Consider Weekend Post on a client where no transactions are ever posted on weekends — it is included in the Expense Cut-off ensemble because it is relevant in some cases, but it should not lower the score simply because it doesn't fire. The demerit point system solves this: a non-firing CP contributes 0 points and is effectively invisible to the scoring, rather than acting as a drag on the ensemble average.

A second consequence of this design is that raw demerit point totals are cross-engagement comparable in a way that final percentages are not. A total of 17 demerit points on an entry means exactly the same thing — the same number and intensity of control point signals — regardless of which client or engagement produced it. The final percentage score, by contrast, is relative: it is normalised to the ensemble maximum, so a given percentage means different things for different ensemble sizes. This distinction is important for auditors reading the Demerit Point breakdown in the platform's detail view — the raw points carry more directly comparable information than the headline percentage.

Library Settings control the band boundaries

The default thresholds [0–25%], [25–75%], [75–100%] are configurable via Library Settings. An administrator can adjust where the band boundaries sit — for example, raising the lower boundary to 35% to apply a stricter noise threshold, or narrowing the middle band to sharpen the distinction between moderate and strong signals. Any such configuration change affects every entry scored against every ensemble that uses the modified settings. This is one of the most consequential Library Settings changes available, and it should require central methodology team approval before being applied.

Stage 3

Demerit point accumulation per entry

For each transaction entry, the platform sums the demerit points produced by all firing control points in the relevant ensemble, broken down by assertion. A single entry can accumulate points across multiple assertions simultaneously, because a single anomalous posting may carry risk signals for more than one audit objective.

The maximum possible demerit points per entry on any given assertion is 2 × the assignment count for that assertion. For Expense Cut-off (9 active CPs) the maximum per entry is 18 points; for Equity Rights and Obligation (1 active CP) the maximum is only 2 points. This ceiling matters because it determines the denominator in the aggregation step — a higher ceiling means more room for the scoring to differentiate between entries.

Cross-assertion accumulation is also significant. A manual journal entry posted on the last working day of the year, with an unusual round-dollar amount, would trigger the following under the mapping, expressed correctly in demerit point terms:

Control point fires	Raw score assumption	Demerit points	Assertions credited
Manual Entry	Fires → assumed 75%+	2 points per assertion	Existence +2, Completeness +2, Valuation +2 (per account category)
Unusual Amount	Fires → assumed 75%+	2 points per assertion	Existence +2, Accuracy/Valuation +2 (per account category)
End of Reporting Period	Fires → assumed 75%+	2 points per assertion	Liabilities Completeness +2, Revenue Cut-off +2 (if applicable)

The resulting demerit profile shows points on Existence, Completeness, Valuation, and Cut-off simultaneously — a differentiated, multi-assertion risk signature that maps to different audit procedures for each assertion. This is the core advantage of the assertion-aware architecture over a single composite risk flag.

Stage 4

Ensemble aggregation

Once entry-level demerit point totals have been calculated for all the entries in every MAC L3 account, the platform aggregates them into a single assertion-level percentage for such account-assertion pair. The current configuration of the platform uses an average-based approach to this aggregation. For each account-assertion pair (e.g. Accounts Receivable-Existence), the average-based method works as follows:

- Calculate the total demerit points across all the entries (N): T
- Divide by N to get the average points per entry: $\mu = T / N$
- Divide by the maximum possible points per entry M (= 2 × assignment count): Score = μ / M

Scenario	T (total points)	N (entries)	M (max per entry)	μ (average)	Assertion Score
Base example	140	30	12 (6 CPs × 2)	4.67	39% — moderate risk
Same entries, higher max	140	30	20 (10 CPs × 2)	4.67	23% — same raw average, lower score because ensemble is larger
Half the total points	70	30	12	2.33	19% — less concentrated risk
Double the total points	280	30	12	9.33	78% — highly concentrated risk

The key trade-off: a small number of very high-risk entries can be averaged out by a large population of low-risk ones. If 5 entries have maximum demerit points and 10,000 entries have zero, the average will be very close to zero even though the 5 genuinely risky entries warrant investigation. Auditors should therefore consider the Demerit Point breakdown alongside the headline percentage, paying particular attention to the highest-scoring individual entries in the detail view.

The risk score distribution chart

To address precisely this trade-off, the platform provides a risk score distribution chart for each assertion. The chart plots all scored entries against their accumulated demerit points, with the Y axis representing either the count or monetary amount of entries and the X axis representing the aggregated demerit points. The distribution chart gives users a visual picture of where entries are concentrated across the demerit point range making it immediately visible whether a headline assertion score is being driven by a broad population of moderately scored entries, a small cluster of very high-scoring entries, or a large mass of near-zero entries pulling the average down.

This visibility is what allows an auditor to contextualise the assertion score percentage rather than take it at face value. A score of 35% on an account with 10,000 entries looks very different depending on whether the distribution shows a roughly even spread across the demerit range, or a concentration of 9,500 entries near zero and 500 entries at maximum demerit points. In the first case, the average is representative; in the second, the average is limited and the high-scoring cluster warrants direct investigation regardless of what the headline percentage suggests.

The distribution chart also surfaces the low-population risk described in Section 3.6 — where an account has very few entries, the distribution will show a sparse, discontinuous pattern that signals to the auditor that the average is not statistically robust. This is a visual prompt to apply the population-size judgment that the headline percentage alone does not convey.

Methodology guidance — designing follow-up procedures around the distribution

Firm methodology should explicitly address how auditors are expected to use the distribution chart alongside the headline assertion score. At a minimum, methodology guidance should require auditors to: (a) review the distribution chart for any assertion score that will be used to inform inherent risk assessment; (b) investigate entries in the extreme high demerit point range regardless of the headline percentage, since these represent the most concentrated risk signals in the population; (c) apply heightened judgment where the distribution shows a very low entry count, as the average is least reliable in these cases; and (d) document in working papers which view of the data — headline score, distribution, or individual entry breakdown — informed the risk assessment conclusion and why.

Stage 5

Account population size effects

Account population size (i.e. the number of entries within a MAC L3) is an independent driver of scoring reliability regardless of the aggregation method used. The internal SME review identified a consistent pattern: accounts with thin populations are structurally prone to inflated scores, because many control point algorithms determine what is anomalous by comparing each entry to the distributional characteristics of the account population. In a large account, this produces a robust baseline. in an account with say 38 entries (there is insufficient variance for a reliable baseline).

Account population	Warning to users (as flagged in the UI)
Fewer than 100 entries	Assertion score carries an explicit population-size caveat in working papers. Auditor qualitative judgment is required before any reliance on score for procedure design.

Stage 6

The Library Settings configuration layer

The mapping matrix represents the platform default configuration. The scope document for Phase 1 confirms that Library Settings allow an admin to configure both assertion-account pairs and risk ranges and demerit points.

The defaults mentioned above are a starting point, not a fixed standard.

Admin-configurable parameters

Three categories of Library Settings configuration affect the scoring pipeline directly:

Configurable parameter	Pipeline stage affected	Methodology implication
Assertion-account pairs — which assertions apply to which account categories	Stage 1 (mapping)	Changes what is scored. Any deviation from defaults must be documented in the configuration record and approved by the methodology team.
Three-band bucketing thresholds — where the 0/1/2 point boundaries sit	Stage 2a (bucketing)	Changes how raw CP scores convert to demerit points. Raising the lower boundary suppresses more noise; narrowing the middle band sharpens signal discrimination. High governance risk — affects all entries across all ensembles.
Risk range thresholds — what demerit point totals map to low/medium/high risk bands	Stage 4 (aggregation)	Changes how scores are classified. Can serve as an interim mitigation for thin-ensemble inflation by preventing single-CP scores from reaching the high-risk band.

Configuration documentation

Working papers that rely on assertion-level scores must document the configuration in effect at the time of scoring:

a) whether default assertion-account pairs are in effect or custom pairs have been configured;
b) the three-band bucketing thresholds — default or custom;
c) whether risk range thresholds are default or custom. This is consistent with any validation and configurations made by the firm before roll out.

Auditor-facing features available (for Phase 1)

The Risk Assertion tab provides auditors with the following capabilities in Phase 1:

Auditor capability	Where available	Purpose
View and filter assertion risk scores	Risk summary page	Identify high-risk account-assertion pairs requiring attention
View detailed entry-level information	Risk assertion detail page	Drill down to the individual GL entries driving a score
View Demerit Point breakdown	Risk assertion detail page	See which control points fired, at what raw score, and how many demerit points each contributed — an important verification tool for explainability purposes
Create samples	Risk assertion detail and summary	Generate a sample population from flagged entries for testing
Create tasks	Risk assertion detail and summary	Assign follow-up actions to engagement team members
Export documentation	Risk assertion detail and summary	Export scored data and audit trail for working paper inclusion

The Demerit Point breakdown view is the primary tool by which an auditor can verify that a score is driven by the control points they expect. It shows each control point's raw 0–100% score, the demerit points that score produced under the three-band system, and the entry's running total.

A firm's methodology guidance should direct auditors to review this breakdown for any high-risk score before relying on it to design a procedure — and specifically to check whether the score is driven by multiple independent signals or by a single control point firing consistently across all entries.

Known Limitations and Scoring Guidance

The following scoring characteristics have been identified through the mapping review and internal SME testing. Each has corresponding guidance for methodology teams and engagement auditors.

Scoring characteristic	Scope	Guidance
Thin-ensemble assertions score high when single CP fires consistently	Equity R&O (count=1), Asset R&O, Liability R&O, Revenue Presentation/Classification, Expense Presentation/Classification (count=4)	Require mandatory Demerit Point breakdown review before relying on scores in these assertion assignments.
Thin-population accounts produce inflated scores	Any account with fewer than 100 entries	Apply the population-size caveat documented in Section 3.6. Auditor qualitative judgment is required before relying on scores from accounts with fewer than 100 entries.
Revenue Cut-off scores are characteristically elevated	Revenue Cut-off (5 active CPs — minimum denominator mechanism does not apply)	Supplement platform scores with independent cut-off procedures. Do not rely on the Revenue Cut-off score alone to scope period-end testing. See Section 5.4.

Scoring characteristic	Scope	Guidance
Equity Rights and Obligation has a single active control point	Equity R&O (count=1, Suspicious Keyword only)	Either exclude this assertion assignment from platform reliance and apply firm-specific procedures, or treat any score as requiring qualitative corroboration before informing procedure design.
COGS entries can wash out non-COGS expense sub-account scores	Expense assertion assignment where COGS and non-COGS accounts share an L3	Configure account-specific assertion-account pairs via Library Settings. Apply auditor qualitative review to high Expense assertion scores on manufacturing or distribution clients.
Material Value control point is unassigned by default	All 25 assertion assignmentss — unassigned in defaults	Activate Material Value for specific account-assertion pairs via Library Settings using account-specific thresholds where materiality-aware scoring adds value.

Completeness and Cut-off — ensemble sensitivity and auditor guidance

Completeness and Cut-off has an average of 7-8 active control points across account categories. The ensemble is sensitive to a wide range of period-end transaction patterns, which is by design: completeness and cut-off is a high-priority assertion area for accounts like revenue, and the ensemble is calibrated to surface period-end clustering, reversal patterns, and timing anomalies that warrant auditor attention.

There is, however, an inherent limitation in any transaction-level approach to completeness assessment that methodology teams must understand and communicate clearly. The platform scores anomalies based on the transactions that are present in the general ledger. A completeness failure — an unrecorded liability, an omitted revenue transaction, a missing accrual — leaves no entry in the ledger and therefore produces no demerit points. Because the platform analyzes recorded ledger transactions, it cannot directly identify transactions that were never recorded.. This is not a gap in the scoring design; it is a structural boundary of ledger-based anomaly detection that applies regardless of ensemble strength or configuration.

For methodology teams, this means Completeness scores should be read as evidence about the pattern and character of the transactions that were recorded — not as an assessment of whether the population of recorded transactions is complete. A low Completeness score does not mean the account is complete; it means the recorded transactions do not exhibit the patterns the ensemble associates with completeness risk. Auditors must supplement platform scores with procedures specifically designed to address unrecorded items — search for unrecorded liabilities, analytical procedures comparing recorded revenue to operational data, and cut-off testing around period boundaries — to address the completeness assertion in a way that the platform, by its nature, cannot.

Presentation and disclosure — the inherent limits of transaction-level detection

Presentation and disclosure is an assertion where the audit evidence that matters most sits above the transaction level. Whether an entity has correctly classified revenue between streams, properly presented financial instrument gains and losses, or met the specific disclosure requirements of the applicable financial reporting framework or GAAP — is a judgment about accounting policy application and financial statement presentation, not about whether individual transactions look anomalous in the ledger. Transaction-level anomaly detection can surface signals that warrant further investigation, but it cannot assess whether the presentation and disclosure choices made by management comply with the specific requirements of the standard.

The platform's Presentation/Classification ensemble — Complex Instrument, Complex Structure, Reversed, and Suspicious Keyword — is therefore best understood as a screening layer: it flags transactions that may carry classification or disclosure risk, prompting the auditor to look more closely. The substantive assessment of whether presentation is appropriate requires auditor procedures that go beyond the ledger — reading financial statement notes, assessing accounting policy disclosures, and evaluating whether the classification of significant transactions reflects their economic substance under the applicable framework.

Rights and Obligation — a contractual assertion that extends beyond the ledger

Rights and Obligation is, at its core, a contractual assertion. The question it asks: does the entity have the rights to the assets it reports, and does it have the obligations for the liabilities it recognises, is answered by examining the legal and contractual arrangements underpinning the balance sheet, not by analysing the pattern of individual ledger entries.

The platform's Rights and Obligation ensembles surface keyword signals and structural patterns that may indicate the presence of complex contractual arrangements — unusual financing structures, “bill-and hold” arrangements, related party obligations, or instruments with embedded features. These are useful prompts for directing auditor attention. But the assertion itself requires procedures that go to the contracts: inspection of agreements, confirmation of legal title, review of security registers, and assessment of whether the entity's contractual obligations are completely and accurately reflected in the financial statements. Methodology teams should treat Rights and Obligation platform scores as one input into that process, not as a substitute for it.

Firm-level validation for these assertion areas

MindBridge is actively developing deeper control point refinements and coverage for Completeness, Rights and Obligation, and Presentation/Classification — assertions where, as described above, the inherent nature of the audit evidence extends beyond what transaction-level detection can fully reach.

In the interim, the most effective path for methodology teams is to run their own validation on these assertion areas using their own client datasets, applying the same back-testing approach described in the section below. The purpose of that validation is to determine the appropriate level of reliance for each firm's specific client profile and methodology.

Some firms may find that the platform's signals for these assertions align well with their prior risk assessments on client types, in which case, a calibrated level of reliance with documented caveats is supportable. Others may find the alignment is weaker for their portfolio, pointing to a configuration adjustment, use as supplementary signal only or even turning them off. Either outcome is a valid conclusion from a well-conducted validation, what matters is that the decision is evidence-based and documented, rather than assumed.

Preparing for a Defensible Rollout

Sections above have explained how assertion-level scores are constructed, where the known limitations sit, and how Library Settings allow firms to customise the configuration.

This section translates all of that into a structured validation programme — a sequence of steps a methodology team can follow to test the feature against their own data, make informed configuration decisions, and achieve a sign-off position that will withstand regulatory and quality review scrutiny. Each stage produces a concrete output that contributes to the final sign-off package.

Understanding how assertion-level scores are constructed — and what the validation process established about their behaviour — puts a methodology team in a strong position to sign off on their use. This section describes what defensible rollout looks like at the principles level: what sign-off means for a tool of this kind, the three things that must be in place before engagement teams can rely on platform scores, and how to maintain that position as the platform and the firm's methodology evolve together.

Validate before you roll out

ISA 315, SAS 145, AS 2110 requires auditors to identify and assess risks of material misstatement through an understanding of the entity and its environment, including the entity's internal controls. Critically, the standard introduces a more structured approach to inherent risk — requiring auditors to assess inherent risk factors such as the complexity of transactions, subjectivity in measurement, and susceptibility to management bias.

Assertion-level risk scoring is designed to support exactly this kind of data-driven inherent risk assessment: it surfaces transaction-level signals at the assertion level, providing quantitative evidence to inform where inherent risk is elevated and where it is not.

For scores to be used in this way, as evidence underpinning inherent risk factor assessments under ISA 315, the methodology team must be able to demonstrate that the scoring engine behaves consistently with auditor judgment on real engagement data. Validation is how that demonstration is built. It is not a technical proof-of-concept; it is a professional standards exercise. The question is whether the platform's assertion-level output is sufficiently aligned with experienced auditor risk assessments to be relied upon as a data source in the ISA 315 risk identification process.

A meaningful validation selects a small number of prior-period engagements where the risk assessment outcome is already known — where the engagement team identified specific areas of elevated inherent risk, applied extended procedures, or documented significant findings. Running those engagements through the platform and comparing the assertion-level output against what the prior team actually concluded answers the central question: do the scores point to the same inherent risk factors the auditors identified, and at the same assertions?

The validation should assess this alignment across at least three (3) dimensions.

First, directional consistency: high-scoring account-assertion pairs should correspond to the areas where the prior team assessed inherent risk as elevated — the platform should not be concentrating scores on assertions the engagement team considered routine.

Second, assertion-level specificity: because ISA 315/AS 2110 requires risk assessment at the assertion level, scores that surface Occurrence risk separately from Completeness risk, or Valuation risk separately from Existence, are providing genuinely useful disaggregated evidence — the validation should confirm this specificity holds in practice.

Third, inherent risk factor mapping: where the platform surfaces a high score, the Demerit Point breakdown should be traceable to control points that correspond to recognisable inherent risk factors — complexity, unusual transaction patterns, period-end concentration, susceptibility to manipulation — rather than data artefacts that have no audit rationale.

The output of this exercise should be a short validation report that forms part of the methodology sign-off package. Where assertion-level scores align with prior inherent risk assessments, that alignment is the evidentiary basis for relying on the platform within the ISA 315/AS 2110 risk identification process. Where they diverge, that divergence should be understood — either as a configuration decision to make, or as a known characteristic to document as a caveat on reliance. The Demerit Point breakdown is the tool for tracing any divergence to its cause.

Selecting validation datasets

Choose a reasonable number of prior-period engagements (firm judgement) that meet the following criteria:

- Industries the firm is auditing (e.g., manufacturing or distribution clients) to test the COGS wash-out behavior described above against real data.

- An engagement where revenue recognition was a significant risk area, to validate the Revenue Occurrence ensemble against known findings.
- An engagement with a small-population account (fewer than 100 entries in a material account), to observe score inflation behavior in a real context.
- An engagement from a sector the firm serves frequently, to establish a baseline for what 'normal' scores look like for that client profile.
- An engagement where a material misstatement or significant finding was identified, so the validation can test whether the platform would have flagged the relevant account-assertion pairs at elevated risk.

What to look for in validation

Run each selected engagement through the platform with mapping default settings and document the following for each:

Validation check	What to examine	Pass condition
High-risk alignment	Do the account-assertion pairs scoring above 60% correspond to the areas where the prior engagement identified elevated risk or applied extended procedures?	Material alignment — not exact match, but high-risk scores should not be concentrated in areas the prior team assessed as low risk
Thin ensemble behavior	What scores are produced and are they driven by a single control point?	Scores from thin ensembles should be reviewed alongside the Demerit Point breakdown; if Suspicious Keyword alone is driving a 100% Equity R&O score, apply minimum risk range configuration before reliance
Population size effect	For accounts with fewer than 100 entries, are scores elevated relative to accounts with similar risk profiles but larger populations?	If thin-population accounts consistently outscore comparable larger accounts, document this as a known pattern and apply the population-size caveat to all such accounts

Validation check	What to examine	Pass condition
COGS wash-out (if applicable)	For manufacturing clients, compare Expense assertion scores for COGS sub-accounts vs non-COGS sub-accounts at the same L3	If non-COGS expense sub-accounts score anomalously high relative to their actual risk, the wash-out effect is confirmed and sector-specific Library Settings configuration should be considered
Control point explainability	For the five highest-scoring account-assertion pairs, open the Demerit Point breakdown and review which control points fired	Auditors should be able to form a plausible audit rationale for why the flagged control points are relevant to the assertion; escalate to the methodology team if a score is dominated by a single control point without an obvious assertion connection
Prior-period stability	If prior-period GL data is available, compare current and prior period scores for the same accounts	Significant score shifts without corresponding changes in business activity or risk profile should be investigated; stable accounts should show broadly stable scores unless the underlying risk genuinely changed

Configure deliberately, not reactively

Library Settings give firms meaningful control over the assertion-account pairs, band thresholds, and risk range classifications that govern scoring. That flexibility is valuable, but it needs to be exercised deliberately before rollout, not in response to individual engagement surprises.

The right approach is to use validation findings to make a small number of principled configuration decisions that apply consistently across the firm's engagements. Configuration that varies engagement by engagement, or that is adjusted informally to manage a particular score, undermines both the comparability of scores and the defensibility of the methodology team's sign-off position.

Every configuration decision, including the decision to keep the default configuration unchanged, should be recorded with its rationale and the approval it received. The configuration in effect at the time of scoring should be identifiable from the working papers. This is what makes a score traceable: not just the number, but the framework that produced it.

Govern configuration as you would govern methodology

Because Library Settings can change what assertions are scored, how signals are classified, and how thresholds are applied, access to them should be governed with the same seriousness as access to audit methodology. Configuration changes that affect assertion-account pairs are equivalent to methodology changes — they should require central approval and be communicated to engagement teams before they take effect.

A practical governance model restricts admin access to a small number of designated individuals, distinguishes between changes that require methodology partner sign-off and those that can be delegated, and treats the configuration log as a quality control document rather than a technical record. Scores produced before and after a configuration change are not directly comparable — which means mid-engagement changes should be avoided, and any change to the firm default should trigger notification to all active engagement teams.

Governance should also be periodic, not just reactive. The active configuration should be reviewed whenever the platform is updated, when a new ensemble version is released, or at minimum annually — to confirm that firm-specific choices remain aligned with how the methodology has evolved.

Documenting validation outcomes

Validation findings should be compiled into a concise validation report covering: which engagements were tested; the overall alignment between platform scores and prior-year risk assessments; specific instances where scores did not align and the likely explanation; and a recommended set of Library Settings adjustments to address observed issues before rollout. This validation report forms part of the methodology sign-off package.

Appendix A — Control Point–to–Assertion Mapping Matrix (Default)

The table below reproduces the default mapping. A filled circle (●) indicates the control point is active for that account-assertion pair. Empty cells indicate no active assignment. Control points shown in grey carry no active assignments.

Control point	Asset					Liability					Equity					Revenue					Expense				
	Exist.	Compl.	R&O	Val./ Acc.	Present.	Exist.	Compl.	R&O	Val.	Present.	Exist.	Compl.	R&O	Val.	Present.	Occur.	Compl.	Accur.	Cut-off	Pres./ Class.	Occur.	Compl.	Accur.	Cut-off	Pres./ Class.
2 Digit Benford				●		●										●		●			●				
Analysis Period Adjustment																	●								
Cash Expenditure		●																							
Cash to Bad Debt Conversion				●																					
Change in Account Trend		●		●			●					●			●		●					●		●	
Complex Instrument				●	●			●	●	●			●	●	●										●
Complex Structure				●	●				●					●				●		●			●			
Duplicate Transaction	●					●					●					●					●				
Empty Text Field								●																	
End of Analysis Period																									
End of Reporting Period																			●					●	
Expense Flurry																					●			●	
Expert Score																									
Flow Analysis																									
High Monetary Value			●						●																
Interfund Transfer														●	●										
Last N Digits									●		●					●		●			●				
Manual Entry	●		●			●		●						●		●		●			●		●	●	
Material Value																									

Control point	Asset					Liability					Equity					Revenue					Expense				
	Exist.	Compl.	R&O	Val./ Acc.	Present.	Exist.	Compl.	R&O	Val.	Present.	Exist.	Compl.	R&O	Val.	Present.	Occur.	Compl.	Accur.	Cut-off	Pres./ Class.	Occur.	Compl.	Accur.	Cut-off	Pres./ Class.
Outlier Anomaly	●			●	●	●				●	●			●	●	●		●	●		●		●		●
POP Change in Risk Profile																									
POP Change in Transaction Frequencies		●					●	●									●		●			●		●	
POP Change in Transaction Values		●		●								●		●			●	●				●			
Rare Flow	●		●			●					●			●		●			●						
Reporting Period Adjustment							●					●							●						
Reversal		●					●			●		●					●					●		●	
Reversed		●					●			●	●	●					●					●		●	
Sequence Gap		●					●					●					●					●			
Split Expense (Single Transaction)																									
Split Expense (Multiple Transactions)																									
Start of Analysis Period																									
Start of Reporting Period			●																					●	
Suspicious Keyword	●			●	●	●	●			●				●	●	●				●	●				●
Unbalanced Debits and Credits		●										●										●			
Unusual Amount																									
Unusual Digit Combination																									
Unusual Digit Sequence											●	●											●		
Weekend Post																								●	
Zero Entry																				●					
Active CP count	**6**	**8**	**4**	**9**	**5**	**8**	**7**	**4**	**5**	**5**	**8**	**8**	**1**	**8**	**5**	**8**	**7**	**8**	**5**	**4**	**8**	**7**	**5**	**9**	**4**