DOI: https://doi.org/10.68732/8e745a76f9092813

ISBN: 978-9989-2053-7-8

Evidence-Linked Point Load Test Measurement Uncertainty via Human-Supervised AI

D. Dimitrievski[1], Lj. Dimitrievski[2], T. Dimitrievski[1] & M. Jovancevska[1]
[1] *Geing DOO, Skopje, North Macedonia*
[2] *Ss. Cyril and Methodius University in Skopje, Faculty of Civil Engineering, Skopje, North Macedonia*
dragan@geing.com.mk
0000-0002-3487-3382

Abstract

In rock engineering the point load strength index Is(50) informs rock strength classification, correlation-based estimation of uniaxial compressive strength (UCS), and rock-mass classification systems used in design. The reliability of each such decision depends on the evidence behind the laboratory value. Yet under ISO/IEC 17025:2017 accreditation, the measurement uncertainty of Is(50) and its supporting evidence are frequently fragmented across spreadsheets, certificates and folders, and rarely traceable as a single record. An uncertainty that is unquantified or undocumented does not disappear; it propagates into classification and design as unmanaged risk.

This paper presents a pilot study of the documentation quality of Is(50) inputs. Using real records from an accredited laboratory that took part in an interlaboratory comparison, it reconstructs the Is(50) uncertainty budget from its physical sources, following JCGM 100:2008 and ISO/IEC 17025:2017 Clause 7.6. It then proposes an Evidence Traceability Matrix (ETM) linking each source to its required documentary evidence and clause, and a Documentation Completeness Score (DCS), a weighted clause-anchored indicator of record completeness, supported by a seven-stage human-supervised workflow in which AI produces the provisional candidate drafts requiring engineer approval.

For the case dataset, repeatability driven by specimen heterogeneity dominates, and the expanded uncertainty is of the order of ±1.2 MPa (k=2) on a ~5 MPa index — a magnitude relevant to the strength-class boundaries the index informs.

The contribution is a structured demonstration on real data and a purpose-built AI-assisted workflow for evidence-linked uncertainty documentation in accredited geotechnical laboratories, with preliminary scoring weights requiring multi-laboratory validation as the next step.

Keywords

Point load strength index, Measurement uncertainty, ISO/IEC 17025, Evidence traceability, Human-supervised AI.

1 Introduction

The point load strength index Is(50) is among the most widely used index tests in rock engineering: fast, inexpensive and field-deployable, it feeds directly into rock strength classification, into correlation-based estimation of uniaxial compressive strength (UCS ≈ C·Is(50), with C commonly 20–25 and wide scatter; ISRM 1985), and into the intact-strength inputs of rock-mass classification (Bieniawski 1989). Because these are design inputs, their reliability is a risk-management concern: for near-boundary cases, an uncertainty of the order found here — about ±1.2 MPa on a ~5 MPa Is(50) value — can decide whether a result sits securely within a class or overlaps adjacent strength-class interpretations.

Any measured index carries a measurement uncertainty; where it is unquantified, or where the supporting evidence is fragmented and unlinked, it carries forward into classification and design as unmanaged decision risk. Machine learning and AI are now well established in rock mechanics and geotechnical engineering for prediction, classification, monitoring and decision support (Zhao & Wu 2025; Zhang et al. 2020). These mainly address prediction and interpretation, whereas this study targets a narrower workflow problem: AI-assisted preparation, evidence linking and technical review of measurement-uncertainty records in accredited laboratories, a documentation challenge also emerging in adjacent regulated contexts (Zai et al. 2025). The workflow uses AI for provisional drafting and evidence linking, while engineer approval remains the human-in-the-loop control before the record is used (Amershi et al. 2019; Mosqueira-Rey et al. 2023).

Under ISO/IEC 17025:2017, accredited laboratories are required to evaluate measurement uncertainty (Clause 7.6) and to maintain traceable records. This paper presents an application-specific structured approach comprising an Evidence Traceability Matrix, a proposed Documentation Completeness Score, and a seven-stage human-supervised workflow, demonstrated on real interlaboratory point-load data. The guiding principle is simple: in regulated rock testing, the credibility of a design input is only as strong as the documented, traceable uncertainty behind it.

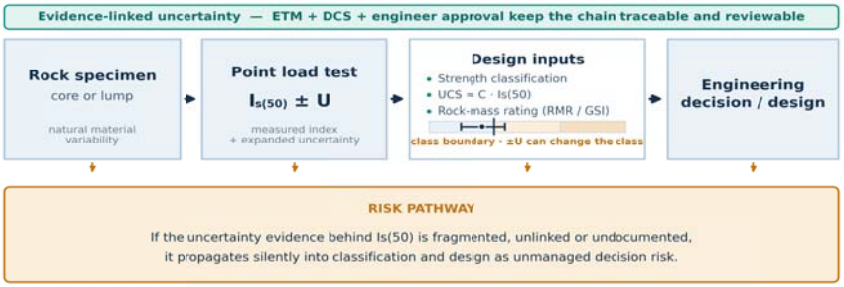


Fig. 1 Rock-engineering consequence chain: Is(50) + U → strength class / UCS correlation → RMR or GSI → design decision. Fragmented evidence turns U into hidden decision risk.

2 Current documentation practice

In many accredited geotechnical laboratories, uncertainty budgets are maintained in spreadsheets, while calibration certificates, interlaboratory results, environmental logs and approvals are stored separately. The evidence chain is therefore reconstructable, but only through manual search, and gaps often become visible immediately before an accreditation assessment. This study addresses that documentation problem by making the uncertainty evidence package explicit, linked and reviewable.

3 Proposed workflow and tools

3.1 Evidence Traceability Matrix (ETM)

The ETM is a method-specific reference table that maps each uncertainty source applicable to a test to the minimum documentary evidence required to establish its traceability, and to the relevant ISO/IEC 17025:2017 clause. Its structure is derived from the source-identification requirements of JCGM 100:2008 Section 4 and the traceability requirements of ISO/IEC 17025:2017 Clause 7.6 (Table 1). For rock point-load testing the matrix makes the load and the specimen-dimension sources explicit, and omits a generic ‘reference-standard traceability’ source, which is not applicable to rock specimens; traceability is instead established through the force and dimensional measurement instruments. Here, u denotes a standard uncertainty component; for example, u_P is the standard uncertainty associated with applied load P, while expanded uncertainty U denotes the combined uncertainty multiplied by the coverage factor k.

Table 1. Evidence Traceability Matrix (ETM) for the point load strength index Is(50). Relative sensitivity coefficients follow $Is(50) = P/De^2 \times (De/50)^{0.45}$ with $De^2 = (4/\pi)\cdot W\cdot D$.

Uncertainty source	Symbol	Type	Relative sensitivity	Required documentary evidence	Clause / guidance	Status (case data)
Applied load (force)	u_P	B	+1	Loading-frame calibration certificate (expanded U)	ISO 17025 Cl.6.4.6	**Present — 24/0674F**
Specimen width / diameter	u_W, u_D	B	−0.78 ea.	Caliper calibration certificate / resolution	GUM Cl.4.3.7	**Present — LM24LAB-0541**
Size-correction factor (De/50)^0.45	u_F	B	model	Method definition (size-correction)	ISRM SM	Requires verification
Repeatability / specimen scatter	u_rep	A	s/√n	Test dataset, n ≥ 10	ISO 17025 Cl.7.6.3	**Present — dominant**
Reproducibility (interlaboratory)	u_repr	A	external	Interlaboratory / PT comparison record	ISO 17025 Cl.7.7	Available; incorp. deferred
Specimen heterogeneity / representativeness	u_samp	B	—	Geological / representativeness assessment	ISRM method note	Largely within u_rep; documented
Environmental (temperature)	u_env	B	minor	Temperature log + sensitivity coefficient	ISO 17025 Cl.6.3	Ambient log available; sensitivity treatment pending

3.2 Proposed Documentation Completeness Score (DCS)

The Documentation Completeness Score (DCS) is a provisional weighted completeness index (Table 2): a checklist-style score showing how complete the uncertainty evidence package is before and after technical review. It is reported as a raw fractional score with a normalized percentage, e.g. DCS = 12/14 (86%). The 14-point weighting reflects the expected importance of each evidence category to the traceability chain in this case: force calibration and repeatability carry higher weights because they directly affect the numerical uncertainty result, whereas environmental monitoring and coverage-factor reporting are lower-weight completeness checks. The DCS is an internal review aid, not a measure of conformity, and requires validation across laboratories, methods, and assessors before broader use.

Table 2. Proposed Documentation Completeness Score (DCS) for the Is(50) case: a raw weighted evidence score with normalized percentage. Weights are preliminary and require multi-laboratory validation.

Evidence criterion	Max	Initial review	After review	Basis
Force / loading-frame calibration evidence	3	2	**3**	ISO 17025 Cl.6.4.6
Repeatability dataset (n ≥ method minimum)	3	3	**3**	ISO 17025 Cl.7.6.3
MLS / interlaboratory performance evidence	2	1	**2**	ISO 17025 Cl.7.7
Specimen heterogeneity / representativeness	2	0	**1**	ISRM method note
Dimensional measurement / caliper evidence	1	0	**1**	ISO 17025 Cl.6.4.6
Environmental monitoring evidence	1	0	**0 (log available; incorp. pending)**	ISO 17025 Cl.6.3
Coverage factor & confidence level stated	1	1	**1**	JCGM 100:2008
Independent engineer review / approval	1	0	**1**	ISO 17025 Cl.7.6
Total	**14**	**7/14 (50%)**	**12/14 (86%)**	—

3.3 Seven-stage human-supervised workflow

The workflow comprises seven stages: upload, AI-assisted extraction, ETM cross-referencing, DCS scoring, provisional uncertainty-budget drafting, engineer verification and assessor-facing export. Stages 2–5 produce provisional drafts only. The record can be used only after engineer approval at Stage 6.

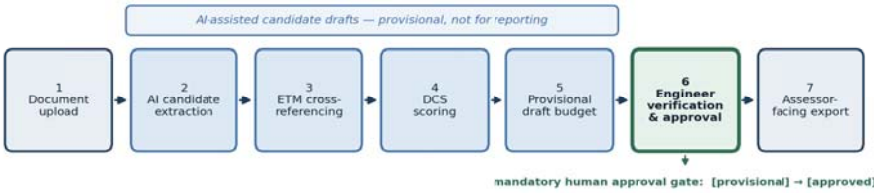


Fig. 2 Seven-stage human-supervised workflow. Stages 2–5 are AI-assisted and produce provisional drafts; Stage 6 is a mandatory engineer-approval gate. Application to the case data is given in Table 3.



Fig. 3 AI Agent prototype (Engeron project) for the Is(50) case. Indicators: DCS = 12/14 (86%) after review (initial 7/14 (50%)); U = ±1.2 MPa (k=2, u_rep dominant); 5 of 7 ETM sources linked, three open items. AI-assisted outputs are provisional until engineer approval; the tool issues no compliance verdict.

4 Case study: interlaboratory point-load test comparison

The case data come from an accredited interlaboratory point-load test comparison organised by the Faculty of Civil Engineering, Skopje, in April 2024, with GEING participating, on rock from the Brazda quarry (Čučer Sandevo). The comparison involved four coded participating laboratories; this laboratory took part as laboratory code 3. It tested at three measurement positions (record sheets MLS-3/1V, MLS-3/2V and MLS-3/3V), each of ten specimens, recording specimen width and diameter, the equivalent core diameter De, the applied load P, and the indices Is and Is(50).

The historical workbooks and the comparison record reference ASTM D5731-16 (withdrawn 2025, no replacement); the methodological background follows the ISRM Suggested Method (ISRM 1985). No standard text is reproduced here.

Applying the workflow, the evidence chain for the reported Is(50) is reconstructed from the loading-frame certificate (24/0674F), the caliper certificate (LM24LAB-0541), the point-load dataset, the interlaboratory record, the environmental record (ambient log available; incorporation pending), and the engineer-approval metadata (Table 3). The laboratory's deviation coefficients satisfied the acceptance criterion ($|Z| \leq 2$; ISO 13528:2022) in all rounds; numerical incorporation of this external evidence is deferred pending method-specific treatment.



Fig. 4 Point-load test evidence at the participating laboratory: (a) Matest A125 press, tag ITJ01; (b) loading cone with calibration-certificate sticker 24/0674F; (c) nameplate, serial A125/AA/0057 — documenting instrument traceability.

Table 3. Seven-stage human-supervised workflow on the interlaboratory case data. AI-assisted outputs (Stages 2–5) are provisional; the record reaches assessor-facing status only after engineer approval (Stage 6).

Stage	Action and outcome on the case data
1 Document upload	Six records uploaded: method reference; loading-frame certificate 24/0674F; caliper certificate LM24LAB-0541; point-load dataset (MLS-3); interlaboratory comparison report; ambient-conditions log (incorporation pending).
2 AI-assisted candidate extraction	Instrument identities, calibration statements and dataset values extracted as a candidate draft for engineer confirmation.
3 ETM cross-referencing	Each Is(50) source matched to its evidence; the dimensional source added, a generic reference-standard source removed as non-applicable; environmental evidence pending.
4 DCS scoring	Provisional DCS = 7/14 (50%); items requiring engineer action listed (caliper linkage, interlaboratory linkage, specimen assessment, approval).
5 Provisional draft budget	Candidate budget generated, all values marked provisional; the certificate statement '±0.50%' flagged for interpretation; u_repr not numerically incorporated.
6 Engineer verification & approval	Confirms that ±0.50% is the expanded uncertainty at k=2, giving u_P = 0.25% relative; links the caliper certificate; records the interlaboratory result ($|Z| \leq 2$); documents the specimen-heterogeneity judgement; approves the record for assessor review. DCS = 12/14 (86%); ambient-conditions record available, numerical incorporation pending.
7 Assessor-facing export	Exported with evidence links: U = ±1.2 MPa (k=2, dataset MLS-3/1V), DCS = 12/14 (86%), approval metadata, and an open-item register (environmental and u_repr incorporation pending).

5 Corrected uncertainty budget

The budget below is for the representative set MLS-3/1V (n = 10), with mean Is(50) 5.15 MPa, sample standard deviation 1.90 MPa (coefficient of variation 37%). With Is(50) = P/De² × (De/50)^0.45 and De² = (4/π)·W·D, the relative sensitivity to the applied load is +1 and to each specimen dimension is −0.78 after the size correction. The Type B inputs come from the calibration certificates: the loading-frame expanded uncertainty is ±0.50% at k=2 (standard uncertainty 0.25%), and the vernier-caliper dimensional uncertainty rests on its calibration certificate (LM24LAB-0541) and a resolution/MPE of 0.05 mm (Table 4).

The coverage factor k=2 is supported by pooling the within-position repeatability variances of the three GEING measurement positions (MLS-3/1V, MLS-3/2V and MLS-3/3V; n=10 each), which are sufficiently homogeneous for this purpose (variance ratio 2.15). This gives a pooled repeatability standard deviation of 1.82 MPa on approximately ν=27 effective degrees of freedom, for which $t_{0.975,27}$=2.05, close to the conventional coverage factor k=2. A single-position treatment of MLS-3/1V alone (ν=9, t=2.26) would give U ≈ ±1.4 MPa; therefore the reported U ≈ ±1.2 MPa is robust to the repeatability treatment. Table 4 therefore retains the MLS-3/1V single-position repeatability estimate for the representative dataset, while the pooled within-position data are used to establish the effective degrees of freedom supporting the coverage factor.

Table 4. Corrected Is(50) uncertainty budget for dataset MLS-3/1V. Assumptions are marked; the size-correction term requires verification and the environmental term's numerical incorporation is pending (ambient log available). The expanded uncertainty is dataset-specific and dominated by specimen scatter. The reported k=2 is supported by pooled within-position repeatability across MLS-3/1V, MLS-3/2V and MLS-3/3V, giving approximately ν=27.

Component	Source value	Distribution / divisor	u (relative)
u_P (applied load)	±0.50% expanded (k=2), cert 24/0674F	÷ 2	0.25%
u_dim (width, diameter)	caliper cert. LM24LAB-0541; res./MPE 0.05 mm	÷√3, ×0.78	≈ 0.07%
u_F (size-correction factor)	model term	—	≈ 1% (requires verification)
u_rep (specimen scatter)	s/√n = 0.60 MPa (n=10)	Type A	**11.7% (dominant)**
u_env (temperature)	ambient log; cert H240549	—	sensitivity term pending
u_repr (reproducibility)	interlaboratory, |Z| ≤ 2	—	deferred
u_c (combined)	**—**	**—**	**≈ 11.8%**
U = 2·u_c	**on mean 5.15 MPa**	**k=2 (≈95%)**	**≈ 23.5% → ±1.2 MPa**

Repeatability driven by specimen heterogeneity dominates so strongly that the corrected Type B terms scarcely affect the result: for this dataset the uncertainty of Is(50) is governed by material variability rather than the instrument (Fig. 5), which is why it must be documented rather than assumed small.

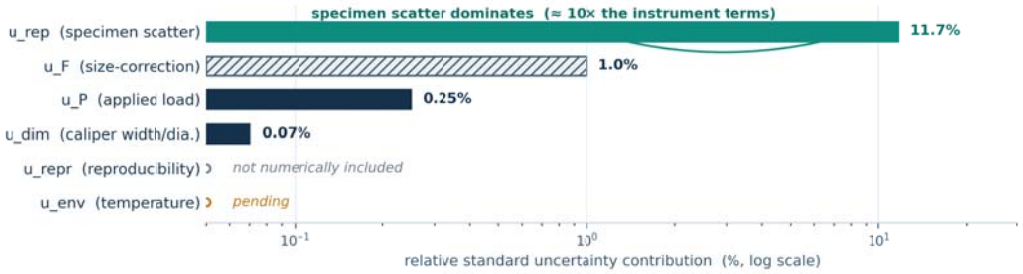


Fig. 5 Relative standard-uncertainty contributions for MLS-3/1V (log scale); specimen scatter (u_rep) dominates the instrument terms by ~an order of magnitude. Dataset-specific values.

6 Discussion: from lab uncertainty to design risk

An undocumented expanded uncertainty of ±1.2 MPa on a ~5 MPa index can shift a near-boundary result into an adjacent strength class (Bieniawski 1989); because UCS is estimated as a multiple of Is(50) with wide scatter, this uncertainty compounds and propagates into RMR and GSI. The ETM and DCS make the evidence reviewable. The reusable element is not the specific 14-point score but the structure — source identification, evidence linkage, completeness review and human approval — which extends to other accredited rock tests (e.g. UCS, triaxial, Schmidt-hammer).

7 Limitations

This is a pilot demonstration based on one laboratory, one method and one dataset; it is not a formal software-validation study or a substitute for accreditation assessment. The DCS is a provisional internal review aid, not an accreditation score, compliance percentage, maturity model, or substitute for competent technical judgement. Reproducibility evidence is deferred pending method-specific treatment, and the environmental component is not yet numerically incorporated (an ambient-conditions log for the period is available). Generalisation requires multi-laboratory, multi-method and multi-assessor evaluation.

8 Conclusion

Point load testing is fast, inexpensive and widely used, which is exactly why the uncertainty of its output deserves explicit documentation: even within one accredited laboratory on nominally similar rock, the scatter is large. For the case dataset, specimen-driven repeatability dominates and yields an expanded uncertainty of about ±1.2 MPa (k=2, supported by pooled within-position repeatability). Where this uncertainty is neither quantified nor traceable, the classification and design decisions built on it inherit unmanaged risk. The Evidence Traceability Matrix and Documentation Completeness Score are an initial step toward making that evidence explicit for assessor review, while the seven-stage human-supervised workflow keeps every AI-assisted output provisional until engineer approval. In regulated rock testing, the credibility of a design input is only as strong as the documented, traceable uncertainty behind it.

Acknowledgements

The authors thank the laboratory team at Geing DOO and the Faculty of Civil Engineering, Ss. Cyril and Methodius University in Skopje, including the laboratory that organised the comparison, for the records used in this study. The AI Agent supports preparation and review of uncertainty records and is not a certification authority or an accreditation-assessment substitute.

References

Amershi S, Begel A, Bird C et al. (2019) Software engineering for machine learning: a case study. Proc. 41st IEEE/ACM Int. Conf. on Software Engineering: Software Engineering in Practice (ICSE-SEIP), Montreal: 291–300. DOI: 10.1109/ICSE-SEIP.2019.00042

ASTM D5731-16 (withdrawn 2025) Standard Test Method for Determination of the Point Load Strength Index of Rock and Application to Rock Strength Classifications. ASTM International, West Conshohocken, PA.

Bieniawski ZT (1989) Engineering Rock Mass Classifications. John Wiley & Sons, New York.

ISO 13528:2022 Statistical methods for use in proficiency testing by interlaboratory comparison. International Organization for Standardization, Geneva.

ISO/IEC 17025:2017 General requirements for the competence of testing and calibration laboratories. International Organization for Standardization, Geneva.

ISRM (1985) Suggested Method for Determining Point Load Strength. International Journal of Rock Mechanics and Mining Sciences and Geomechanics Abstracts 22(2): 51–60.

JCGM 100:2008 Evaluation of measurement data — Guide to the expression of uncertainty in measurement (GUM). Joint Committee for Guides in Metrology, BIPM, Sèvres.

Mosqueira-Rey E, Hernández-Pereira E, Alonso-Ríos D, Bobes-Bascarán J, Fernández-Leal Á (2023) Human-in-the-loop machine learning: a state of the art. Artificial Intelligence Review 56(4):3005–3054. DOI: 10.1007/s10462-022-10246-w

Zai K, Mukti AP, Notty GS, Niswah SU, Refliandi, Oktivendra F (2025) Artificial intelligence in the non-clinical laboratory: enhancing good laboratory and documentation practices. International Journal of Pharmaceutics 685:126266. DOI: 10.1016/j.ijpharm.2025.126266

Zhang W, Zhang R, Wu C et al. (2020) State-of-the-art review of soft computing applications in underground excavations. Geoscience Frontiers 11(4): 1095–1106. DOI: 10.1016/j.gsf.2019.12.003

Zhao G-F, Wu Y (2025) Artificial intelligence in rock mechanics. International Journal of Rock Mechanics and Mining Sciences 195:106245. DOI: 10.1016/j.ijrmms.2025.106245