Revista Nebrija
de Lingüística Aplicada a la Enseñanza de las Lenguas

Revista Nebrija de Lingüística Aplicada a la Enseñanza de las Lenguas. Vol. 20 Núm. 40 (2026)
ISSN 1699-6569

# What Counts as Academic Rigour? Epistemic Politics in the Assessment of Master of Arts Dissertations in an Algerian English as a Foreign Language Department

## *¿Qué cuenta como rigor académico? Políticas epistémicas en la evaluación de las tesis de maestría en un departamento argelino de Inglés como Lengua Extranjera*

Saida Tobbi
Batna 2 University, Algeria, s.tobbi@univ-batna2.dz

Abstract

Drawing on a qualitative multi-method study conducted at the English Department of the University of Batna 2, this paper investigates how standards of academic rigour are articulated and enacted in the assessment of Master of Arts (MA) dissertations. Data comprise a purposive corpus of 120 dissertations defended between May 2023 and June 2025, along with their associated examiner reports and semi-structured interviews with 12 supervisors and 13 examiners. A stratified sub-sample of 36 dissertations was analysed in depth. Findings reveal that, although official rubrics supply procedural criteria, evaluators also rely on unspoken interpretive standards, resulting in only partial alignment between written policy and actual practice. Three mechanisms mediate this gap: methodological legibility, supervisory socialisation, and internal board composition. The study contends that improving fairness requires a combined approach of calibrated rubrics supplemented by annotated exemplars, examiner calibration workshops, and supervisor development aimed at enhancing analytic transparency. Implications for assessment policy and comparative research are discussed.

Keywords. Academic writing, academic rigour, English as a foreign language assessment, higher education evaluation, Master of Arts dissertations (MA).

*Resumen*

*Lorem ipsum dolor sit amet consectetur adipiscing elit mollis habitasse semper, ante A partir de un estudio cualitativo multimétodo realizado en el Departamento de Inglés de la Universidad de Batna 2, este artículo investiga cómo se articulan y se ponen en práctica los estándares de rigor académico en la evaluación de las tesis de Maestría (MA). Los datos comprenden un corpus intencional de 120 tesis defendidas entre mayo de 2023 y junio de 2025, junto con sus informes de examinadores y entrevistas semiestructuradas con 12 supervisores y 13 examinadores. Se analizó en profundidad una submuestra estratificada de 36 tesis. Los hallazgos revelan que, aunque las rúbricas oficiales suministran criterios procedimentales, los evaluadores también se apoyan en estándares interpretativos tácitos, lo que provoca una alineación sólo parcial entre la política escrita y la práctica real. Tres mecanismos median esta brecha: la legibilidad metodológica, la socialización en la supervisión y la composición interna del tribunal. El estudio sostiene que mejorar la equidad requiere un enfoque combinado de rúbricas calibradas suplementadas con ejemplares anotados, talleres de calibración para examinadores y programas de formación para supervisores orientados a aumentar la transparencia analítica. Se discuten las implicaciones para la política de evaluación y la investigación comparativa.*

*Palabras clave. Escritura académica, rigor académico, evaluación del inglés como lengua extranjera, evaluación en la educación superior, tesis de Maestría (MA).*

DOI: 10.26378/rnlael2040660
Recibido: 11/01/2026 - Aprobado: 12/03/2026

## 1. Introduction

Master's dissertations operate as high-stakes gateways in higher education: they certify independent research capability and function as credentialing instruments for academic and professional advancement. Yet the precise meaning of "rigour" — the core criterion by which dissertations are judged — is rarely unambiguous. Written rubrics and departmental guidelines set out formal expectations, but the judgments that ultimately determine acceptance, revision or failure are produced in situated evaluative practices: in examiner reports, board deliberations, and day-to-day supervisory advice. In Algeria's LMD system, dissertation evaluation carries particularly high epistemic stakes, as language hierarchies (among Arabic, French, and English), local disciplinary traditions, and institutional pressures for standardization all intersect.

A recent internal study in the English Department at the University of Batna 2 (Benbouabdallah & Benmekhlouf, 2023) reported widespread teacher support for a standardized rubric to increase marking consistency and efficiency; that study produced a detailed checklist of dissertation elements (title, originality, structure, methodology, analysis, depth of discussion, references, etc.). While practically useful, a checklist approach does not explain how criteria are interpreted, negotiated, and operationalised in practice. In particular, it leaves unexamined three core issues: (a) the gap between what is written and what is rewarded (how examiner reports and marks align with rubric items), (b) the role of local epistemic hierarchies (how language choices and methodological preferences function as proxies for rigour), and (c) the institutional configuration of gatekeeping (in Batna 2, boards comprise a chairperson, the student's supervisor and an internal examiner — with no external examiners — a structure with important implications for local control over standards).

This study examines these issues through a multi-method qualitative study of the English Department at the University of Batna 2. The empirical core is a corpus of 120 MA dissertations defended between May 2023 and June 2025 (the COVID period was intentionally excluded because emergency assessment practices would distort findings). These dissertations belong to the three departmental options — LLA (Language and Applied Linguistics), LC (Language and Culture), and Didactics — providing a cross-section of disciplinary orientations. The corpus analysis is complemented by a purposive, stratified set of in-depth readings of a sub-sample of dissertations and their examiner reports, and by semi-structured interviews with n = 25 departmental teachers (supervisors and internal examiners) sampled across Algerian academic ranks: Maître assistant (Assistant Lecturer); Maître de Conférences B (Associate Professor); Maître de Conférences A (Senior Associate Professor); and Full Professor.

The study asks the following research questions:

1) How do supervisors and internal examiners in the English Department at Batna 2 articulate the criteria of rigour when assessing MA dissertations?
2) To what extent do written departmental rubrics and guidelines correspond with evaluative practices evident in examiner reports and dissertation outcomes across the corpus of 120 dissertations?

3) What institutional and epistemic factors (e.g., originality, methodological norms, language and citation practices) shape the enactment of standards of rigour?

This study makes a focused empirical and theoretical contribution. Empirically, it provides a corpus-based account of 120 MA dissertations (with 36 close-read cases) that links written rubrics, examiner reports and supervisor practices to concrete outcomes in an Algerian EFL context. Theoretically, it introduces a mechanism-level explanation for evaluative discretion by identifying three mediators — methodological legibility, supervisory socialisation and board composition — that translate written criteria into enacted judgements. Practically, it offers testable, institution-level interventions (calibrated exemplars, examiner calibration, supervisor development) that directly address the documented rubric–practice gap. Together these elements move the scholarly conversation beyond descriptive accounts of inconsistency toward an operational model for reducing evaluative variance.

Recent scholarship has emphasised that specification alone (rubrics, checklists) is insufficient unless accompanied by social processes that produce shared interpretive work among examiners and supervisors. Examiner calibration and social moderation practices have been proposed as practical complements to rubric specification, because they help translate procedural criteria into shared reading practices and reduce local arbitrariness in judgement (O'Donovan et al., 2024; Tan, 2024). This paper therefore positions its institutional recommendations (calibrated exemplars, examiner workshops, supervisor development) not merely as administrative fixes but as socially embedded instruments for shifting shared interpretive repertoires in departmental cultures.

## 2. Literature review

Research on postgraduate dissertation assessment treats rigour not as a single, self-evident property but as a multidimensional achievement produced in situated evaluative practice. Across the literature, scholars converge on several interdependent dimensions that examiners and committees mobilise: methodological soundness (robust design and transparent analytic procedures), theoretical and conceptual depth, analytical coherence, and trustworthiness/ethical reporting — the latter expressed in discipline-appropriate terms (e.g. validity/reliability or credibility/dependability/transferability). These dimensions operate less as neutral checkboxes than as normative axes that actors selectively invoke to justify judgements. (Goodman et al., 2020; Morse, 2015; Mullins & Kiley, 2002; Varela et al., 2021; Yadav, 2021.)

A recurrent finding is the coexistence of two evaluative registers. One is procedural — checklist-like rubric language that notes the presence/absence of required elements (research questions, method chapter, bibliographic conventions) and offers administrative defensibility. The other is tacit and interpretive — idioms of "analytical depth," "intellectual contribution," and "conceptual engagement" that rubrics do not fully capture. Empirical analyses of examiner reports and defence interactions show that committees use procedural language instrumentally while substantive decisions often depend on tacit interpretive labour. This duality explains why formally compliant theses may still be asked for substantial revision, and why papers with strong

conceptual claims can succeed despite presentation weaknesses. (Mullins & Kiley, 2002; Holbrook et al., 2004; Man et al., 2020.)

Institutional responses commonly emphasise rubrics and QA frameworks because these instruments improve clarity and drafting (Hsiao, 2024). Yet scholarship warns that specification alone can privilege particular epistemic forms and leave substantive discretion intact: rubrics provide vocabularies and scaffolds but must be calibrated to be reliably determinative (Homer, 2026). Studies that compare written rubrics with enacted practice repeatedly find that rubrics function as justificatory covers for discretionary judgement unless accompanied by exemplification and shared interpretive work (Bukhari et al., 2021; Belcher et al., 2016; Reddy & Andrade, 2010).

Supervision is central to how standards are realised in practice. Supervisors translate tacit expectations into manuscripts by advising on structure, method transparency, and presentation; this editorial labour can standardise theses and advantage candidates whose supervisors possess stronger genre knowledge and networks. At the same time, supervisory mediation produces inequality when supervisory capacity is uneven, supporting calls for formal supervisor development (Lee, 2018; Bastola & Hu, 2020; Chugh et al., 2021).

Methodological heterogeneity further shapes legibility. Quantitative designs tend to yield discursively visible chains of evidence (sampling frames, tables, statistical summaries) that boards find straightforward to evaluate; qualitative traditions require explicit analytic transparency (coding procedures, audit trails, reflexivity) to achieve equivalent credibility. Where exemplars and discipline-sensitive guidance are absent, qualitative work risks being read as anecdotal or under-analysed, thereby creating pressure to mimic quantitative legibility or to provide supplementary documentation (Morse, 2015; Varela et al., 2021; Crowe et al., 2024).

Institutional configuration and committee composition matter: the presence (or absence) of external examiners, reputational relations, and local disciplinary norms influence interpretive frames and gatekeeping dynamics. Internal-only boards can amplify local epistemic hierarchies; external examiners may introduce alternative perspectives and reduce insularity (Mullins & Kiley, 2002; Mafora & Lessing, 2016; Stigmar, 2018).

In EFL and international candidate contexts, language and rhetorical fluency interact with substantive assessment. Examiners sometimes conflate presentation and analytic substance, risking epistemic exclusion where language proficiency is taken as a proxy for scholarly merit. Interventions that scaffold disciplinary writing and separate language from epistemic contribution are therefore vital in multilingual contexts (Othman & Lo, 2023; Man et al., 2020; Tiwari, 2023).

These strands of research converge on a pragmatic conclusion: to make academic rigour more transparent and equitable, specification (rubrics) must be paired with social processes that render tacit norms explicit — notably calibrated rubrics with annotated exemplars, examiner calibration workshops, and supervisor development focused on analytic transparency. This combined strategy respects methodological plurality while reducing arbitrary local epistemic effects (Belcher et al., 2016; Bukhari et al. 2021; Kumar & Stracke, 2011).

## 3. Methods

### 3.1. Design

This study adopts a qualitative multi-method interpretive design with convergent triangulation to examine how standards of academic rigour are articulated, operationalised and legitimised in MA dissertation assessment. The design is appropriate because rigour is not a fixed or directly observable attribute, but a socially constructed judgement produced through discourse, institutional routines and professional interpretation. Qualitative methods are therefore required to capture examiners' reasoning, the discursive work of assessment texts, and the mechanisms—such as supervisory mediation, methodological legibility and board dynamics—that shape evaluative outcomes. Multiple qualitative data sources are analysed in parallel and integrated through triangulation to explain divergences between written criteria and enacted practice. Limited descriptive quantification is used only to contextualise the corpus; the study's explanatory force rests on qualitative interpretation and case-level integration of evidence.

### 3.2. Setting and corpus construction

The empirical setting is the Department of English, University of Batna 2. The documentary corpus for analysis is a purposive sample of 120 MA dissertations drawn from the larger set of theses submitted to the department between May 2023 and June 2025. The sampling frame for this corpus was constructed as follows. First, the departmental registry of MA submissions for the period May 2023–June 2025 was consulted and used to identify candidate files. Second, electronic copies of the identified dissertations were retrieved from the department repository or produced by scanning official printed copies. Third, each file was inspected for completeness (title page, abstract, chapters, bibliography and final board decision) and assigned an anonymised identifier. Fourth, associated artefacts (available internal examiner reports or marking sheets and the departmental guidelines/rubrics in force during the period) were collected and linked to the corresponding dissertation records.

The sample of 120 dissertations comprises 40 dissertations from each of the department's three principal options: LLA, LC, and Didactics. This balanced sampling across options supports comparative reading across the department's main programme orientations.

### 3.3. Corpus and Sub-sample Construction

Because the selected corpus was large for intensive interpretive reading, the analysis proceeded in two complementary tiers. Tier 1 applied concise objective coding across the full corpus of 120 dissertations to produce contextual descriptions that supported interpretive claims, while Tier 2 employed a stratified purposive sub-sample for close qualitative reading and case-level triangulation.

Tier 1 (corpus-wide coding) used a short coding sheet applied to every dissertation to capture essential interpretation-relevant features while deliberately avoiding heavy quantification. Tier 2 (close reading) selected a moderate sub-sample for in-depth analysis: a 36-case sub-sample with equal representation by option (12 per option) was drawn. Within each option, cases were stratified by year (2023, 2024, 2025),

methodological orientation (qualitative, quantitative, mixed, theoretical) and grade band; within strata, individual theses were selected by random draw. The selection algorithm and the justification for any purposive inclusions were recorded in the Methods appendix to ensure transparency and reproducibility.

Tier 1 descriptive coding (applied to all 120 dissertations) was not only contextual background but also a source of inductive patterning used to inform Tier 2 causal claims. For example, the predominance of quantitative designs (50%) and the distribution of grade bands (10% low, 60% middle, 30% upper) informed targeted comparisons that probed whether evidentiary legibility (e.g., presence of sampling frames, tables) predicted fewer revision requests in examiner reports. These descriptive patterns were used instrumentally to select negative cases and to triangulate mechanism claims (see Appendix G).

Supervisors and internal examiners associated with the close-read these were prioritised for interview to enable case-level triangulation; where direct linkage to a sampled thesis was not possible because an individual was unavailable, interview recruitment was broadened purposively to maintain analytic breadth. Supervisor rank was recorded for contextual purposes but was not used as a selection criterion for the sub-sample.

### 3.4. Participants and recruitment

Interviews were conducted with departmental teachers who acted as supervisors or internal examiners during the study period. The target interview sample comprises approximately 25 teachers: 12 supervisors and about 13 examiners. Participants were purposively recruited on the basis of active supervisory or examining experience between May 2023 and June 2025 and with attention to capturing a range of research profiles and experience levels. While academic rank was recorded, it did not influence participant selection. Participants were recruited via an initial email invitation containing an information sheet. Interviews, conducted individually in English, were scheduled at each participant's convenience.

### 3.5. Data sources and instruments

Data sources comprise:

- The corpus of 120 selected dissertations (full files as collected and anonymised);
- The internal examiner reports;
- The departmental guidelines and any formal rubrics in force during the study period; and
- Semi-structured interviews with the purposive sample of departmental teachers

The study employs two complementary coding instruments. The corpus coding sheet, applied to all 120 dissertations, records anonymised ID, year of submission, declared option (LLA / LC / Didactics), anonymised supervisor and internal examiner names, methodology type (qualitative / quantitative / mixed / theoretical), presence of explicit research question(s), presence of an explicit theoretical framework, approximate reference count band, grade band on the department’s 0–20 scale (observed values in

the sample fall between 10 and 17), and a binary flag indicating whether the examiner report records substantive concerns requiring revision. The close-read coding frame used for the sub-sample was derived from Batna 2 University's English department rubric and expanded with inductive epistemic codes that capture conceptual clarity, theoretical engagement, methodological justification, data quality, analytic rigour, originality, literature use, citation practice patterns, academic writing quality and explicit examiner criticisms.

### 3.6. Procedures and data handling

All dissertation files and examiner reports were anonymised at intake: student and staff names were replaced by coded identifiers and a secure, encrypted key linking codes to identities was stored separately and only accessible to the principal investigator. The corpus coding sheet was piloted on a small set of sample files to refine definitions and coding bands; inter-coder checks were applied to a purposive 10% subset before full corpus coding to ensure consistency.

Interviews were semi-structured, lasting approximately 45–75 minutes, audio-recorded with participant consent and professionally transcribed. Transcripts were anonymised and stored on encrypted drives. Interview guides began with broad questions about participants' definitions of rigour and proceeded to request concrete, anonymised examples from their supervisory or examining practice; participants were given the option to respond using composite or masked examples to protect confidentiality.

### 3.7. Analysis

Analysis proceeded iteratively and comparatively. Corpus coding produced a concise descriptive backdrop that was reported sparingly and only to contextualise interpretive claims. Thematic analysis (reflexive approach following Braun & Clarke, 2006) was applied to interview transcripts and to the open interpretive segments of examiner reports. Critical discourse analysis (CDA) was applied to examiner reports and departmental guidelines to reveal how evaluative language constructs authority and frames standards of rigour. For each close-read thesis a triangulation matrix was produced that aligned rubric items and guideline statements (what is written) with examiner comments, thesis features and the supervisor/examiner's interview claims (what is enacted). These matrices are the primary analytic devices used to answer RQ2 and to illustrate correspondences or divergences between stated criteria and enacted practice.

### 3.8. Trustworthiness and reflexivity

Trustworthiness was strengthened through multi-source triangulation (examiner reports, interviews, and close-read case matrices), detailed case vignettes, and a documented audit trail of sampling and coding decisions. Intercoder and interpretive checks were expanded beyond binary coding by implementing a reflexive codebook process and periodic consensus meetings (pilot coding was conducted on 10% of the corpus; co-coding and reconciliation meetings were convened; disputed items were re-coded). For interpretive thematic coding of interview transcripts and open-text examiner comments, a combined procedure of reflexive thematic analysis and coder cross-checks was applied: (1) initial independent coding was conducted by two analysts;

(2) code application was compared and discrepant interpretations were discussed; (3) analytic memos documenting interpretive decisions were produced; and (4) final consensus coding was completed. Raw agreement for the corpus-level binary/ordinal fields was 90%, and Cohen's κ was 0.78 for the pilot co-coded sample; after reconciliation and final consensus coding, raw agreement increased to 94% and Cohen's κ increased to 0.82. Qualitative evidence of interpretive depth is provided through analytic memos and extracted analytic examples that document how discrepant readings were resolved. This procedure is consistent with current guidance on reporting intercoder procedures for interpretive qualitative work (Cheung, 2023; Cofie et al., 2022).

### 3.9. Ethical considerations

Ethical approval was obtained before data collection. Informed written consent was secured for all interviews. Dissertation files and examiner reports were anonymised on intake; publications use anonymised citations and redacted quotations where necessary to prevent identification. Sensitive examiner comments are only quoted with explicit consent or presented in heavily anonymised form. All raw data are stored in encrypted media with restricted access.

## 4. Results

This section answers the study's three research questions by moving from a concise corpus description to patterns visible in examiner reports, to themes derived from interviews, and then to an integrated synthesis that triangulates across sources. The analytic procedures combined reflexive thematic analysis of interview transcripts and open-text examiner comments with CDA of examiner reports and guidelines; case-level triangulation used a documented matrix that aligned dissertation features, examiner comments and interview claims for each close-read case.

At the corpus level, the distribution is balanced across the department's three options (40 dissertations in LLA, 40 in LC, and 40 in Didactics). Methodologically, the corpus is dominated by quantitative research: 60 dissertations (50%) employ quantitative designs, 36 (30%) use mixed methods, and 24 (20%) rely primarily on qualitative approaches. An explicit research question is stated in 96 dissertations (80%) and an explicit theoretical framework is visible in 78 ones (65%). Examiner reports flagged substantive methodological or reporting concerns in 24 dissertations (20% of the corpus). Grade bands cluster in the mid-range: approximately 10% of dissertations fall in the lowest band (grades 10–11), 60% in the middle band (12–14) and 30% in the upper band (15–17). These descriptive figures contextualise the interpretive claims that follow and are reported sparingly so as not to displace the study's qualitative emphasis.

Critical discourse analysis of examiner reports and departmental guidelines reveals a recurrent rhetorical double register. Examiner texts routinely invoke rubric-consistent language — explicit mentions of "research questions," "methodological clarity" and "formatting requirements" appear in a substantial number of reports — and these formal items are mobilised as explicit justificatory resources in accept/revise decisions. At the same time, examiner reports also regularly use tacit evaluative idioms, such as "analytical depth," "intellectual contribution" and "conceptual engagement," which do not map neatly onto checklist items. The CDA shows that rubric language is used

instrumentally: it legitimates administrative decision-making, while tacit idioms articulate the department's implicit standards of scholarly value.

Interviews with supervisors and examiners illuminate how these discursive registers are lived and operationalised. Participants articulate rigour in a dual mode: as procedural defensibility and as interpretive contribution. A supervising senior associate professor captured this duality succinctly: "We ask for a method chapter that is readable and auditable; that gives the board something concrete to point to. But when it comes to awarding merit, we are looking for a thesis that actually argues — not just reports." [Sup-11] An examiner explicitly described the evaluative asymmetry between methodological forms: "Quantitative work presents chains of evidence in a way boards like to see; qualitative work must build an equivalent chain — coding steps, traces of interpretation — otherwise we ask for more detail." [Exam-09] Interviews also describe supervisory labour as a mechanism of alignment: supervisors routinely advise on the format and presentation that boards recognise, and in many cases assist in revising method chapters and results prior to submission.

From these sources four integrated themes answer the research questions directly:

### 4.1. Dual register of rigour

The data show that evaluative work in this department operates through two interlocking registers. The procedural register (rubric language: presence of research questions, method chapter components, referencing) functions as a discursive instrument of defensibility in boards' written reports and deliberations: invocation of rubric items provides a publicly legible rationale for decisions. The epistemic register (idioms such as "analytical depth" or "intellectual contribution") governs substantive valuation. Mechanistically, the procedural register reduces cognitive and reputational risk for examiners and the board (it is a legalistic script to justify outcomes), while the epistemic register enables evaluators to apply tacit disciplinary hierarchies when assigning merit. These registers therefore perform different functional work: one secures procedural defensibility; the other adjudicates scholarly worth. Triangulation matrices show repeated cases where the rubric was quoted in the written report while the decision rationale in interviews invoked tacit evaluative language. This co-occurrence supports a mechanism in which rubric invocation operates instrumentally to legitimate decisions that are substantively driven by tacit epistemic judgments.

### 4.2. Methodological hierarchies and evidentiary legibility

The corpus distribution (50% quantitative; 30% mixed; 20% qualitative) alone does not fully explain outcomes. Crucially, examiners privileging of "visible chains of evidence" creates a structural incentive: artefacts that render claims auditable (sampling frames, tables, code logs) act as institutional currency because they materially reduce the interpretive labour required by boards. Mechanistically, then, method forms that produce legible artefacts are advantaged because they minimize the need for deep interpretive work during deliberations. Case-level triangulation supports this: quantitative theses with clearly presented sampling and statistical tables frequently received fewer revision requests, even when theoretical novelty was modest; conversely, qualitative theses demonstrating conceptual insight but lacking documented analytic trails were asked to supply coding logs or audit trails before being

accepted. This pattern is consistent with the hypothesis that legibility, not only methodological correctness, mediates evaluative outcomes.

### 4.3. Supervisory socialisation as de facto standardization

Supervisors in this department function as gatekeepers who translate tacit departmental expectations into manuscript practice. The mechanism here is socialisation: supervisors transmit interpretive repertoires (what counts as a readable methods section; how to package results) through repeated editorial and pedagogic interventions. Evidence of supervisor edits, timestamps, and supervisor interview claims (e.g., advising on format and method transparency) show supervisor work converting tacit norms into legible products. This produces distributed inequalities: candidates whose supervisors possess stronger genre knowledge and experience are more likely to produce theses that align with board expectations. That pattern is visible in the close-read sub-sample where supervisor-mediated reworking correlated with fewer examiner revision demands.

### 4.4. Rubric–reality gap and instrumental invocation

Written rubrics are often cited in reports, but their application is inconsistent in ways that favour particular forms of scholarship. In the stratified close-read sample, rubric–report alignment was observed in 19 of 36 cases (53%) while in 17 cases (47%) reports foregrounded tacit criteria that did not appear explicitly in the rubric. Interviewees described routine use of rubric language as a defensive device for borderline decisions: “We quote the guideline to be transparent; sometimes it is our shield,” one examiner admitted. The CDA makes visible how rubrc text and tacit evaluative idioms co-exist and how the former is mobilised instrumentally.

### 4.5. Extended discourse-analytic excerpts and coding

To strengthen the discourse-analytic evidence, a set of anonymised and systematically analysed excerpts from examiner reports and interview transcripts is included. Quotations are anonymised to protect participants.

“The methodology chapter follows the required structure and provides a clear chronology of fieldwork and data collection, yet no coding log or analytic trace was supplied that would allow a reader to verify how interpretive moves were produced from the raw materials. The report stated that ‘interpretive claims are asserted but not shown’ and requested an appendix with coding frames, exemplar coded extracts, and an explanatory note on analytic procedures; the board’s written guidance concluded that, in the absence of an audit trail, the interpretive claims could not be treated as reproducible.” (Source: anonymised internal examiner report.) This passage is coded as procedural invocation, legibility demand and evidentiary insufficiency. It is interpreted as an instantiation of the legibility mechanism in which artefacts such as coding logs and exemplar extracts are required to render interpretive claims externally verifiable; consequently, the burden of proof is shifted onto the candidate, and an incentive structure is produced that favours methods yielding auditable traces.

Following that procedural emphasis, evaluative language shifts the focus to conceptual sufficiency and conditional endorsement. “Chapters 2 and 5 advance a promising theoretical line and propose several novel links between the literature and the dataset,

yet the conceptual exposition remains under-developed and insufficiently integrated with the empirical examples. The report recommended strengthening the conceptual scaffolding, tightening theoretical definitions, and demonstrating more clearly how selected empirical extracts substantiate the proposed claims; guarded praise for originality was offered, but a higher grade was made contingent on substantial elaboration." (Source: anonymised internal examiner report.) This statement is coded as epistemic valuation, conditional endorsement and rhetorical hedging. It is interpreted as evidence that the epistemic register acknowledges conceptual merit while making acceptance contingent through hedging devices; thus, a parallel evaluative pathway is created in which substantive judgement is negotiated independently of, yet intersecting with, procedural checks.

Bridging the procedural and epistemic registers, supervisory practice is invoked as the mechanism that converts tacit expectations into the legible artefacts demanded by examiners. "Advice was routinely provided on how to prepare the methods chapter so that board members could follow the argument without needing to reconstruct analytic steps from raw materials. Typical guidance included inserting sample tables, providing an ordered list of analytic steps in an appendix, adding short exemplar extracts showing how codes were applied, and noting data cleaning procedures. It was reported that theses incorporating these features tended to face fewer procedural revision requests during board consideration, even when theoretical claims were ambitious." (Source: anonymised supervisor interview.) This passage is coded as supervisory socialisation, transmission of interpretive repertoire and packaging for legibility. It is interpreted as direct evidence that supervision functions as an intermediary mechanism that translates tacit departmental norms into concrete documentation practices, thereby increasing manuscript legibility and reducing the interpretive labour required by examining bodies.

Collectively, the excerpts are read as mutually reinforcing: the first establishes the procedural requirements that make interpretive claims auditable; the second shows how epistemic endorsement is frequently made conditional by those procedural expectations; and the third demonstrates how supervisory action is deployed to produce the artefacts that satisfy both sets of expectations. The analytic readings paired with each excerpt serve to make explicit the rhetorical moves and the institutional effects that constitute the mechanisms identified in the Results.

To illustrate how these dynamics operate in concrete decisions, two case vignettes are presented below that exemplify divergent outcomes produced by differences in evidentiary legibility and supervisory mediation. In one instance (T12-LLA-2024), a qualitative dissertation advanced a novel interpretive framing and a richly argued conceptual claim, but the methods chapter did not include a documented coding trail or a systematic appendix. The internal examiner's written report observed that "the claims are interesting and theoretically suggestive, yet a clear coding trail is absent; without exemplar coded extracts or a coding frame the interpretive steps remain opaque," and a request for appendices documenting coding procedures was issued. In a subsequent interview, the supervising lecturer stated that editorial attention had been focused on developing argument and conceptual coherence and that the student had

not anticipated the board's demand for explicit analytic documentation (Source: internal examiner report and supervisor interview).

By contrast, a different instance (T07-DID-2023) presented a thesis in which methodological transparency was foregrounded: a clearly stated sampling frame, tabulated descriptive summaries, and an appendix containing coding notes or statistical output were provided. The examiner report commended the evidence base, noting that "a robust chain of evidence is visible from sampling to results" and recommended acceptance with minor revisions despite limited theoretical novelty. During interview, an internal examiner explained that visible artefacts of evidentiary legibility frequently reduce the need for protracted deliberation and allow boards to foreground substantive claims more readily (Source: examiner report and examiner interview). These two vignettes illustrate the pragmatic trade-offs that frequently underlie assessment decisions: conceptual innovation may be disadvantaged when procedural artefacts are absent, whereas evidentiary visibility can mitigate limited theoretical distinctiveness.

Analytically, these findings converge on a central conclusion: written rubrics and guidelines structure departmental expectations and supply a shared vocabulary, but enacted standards of rigour are produced in practice through interactions among methodological legibility, supervisory mediation and board-level interpretive habits. The triangulated evidence demonstrates not only where rubric and practice align, but also the mechanisms that explain divergence — local expectations about evidentiary form, supervisory editorial labour, and the internal composition of boards that privileges reputational influence.

Inter-coder reliability checks of the corpus coding process showed 90% raw agreement on binary items (explicit research questions, explicit theoretical framework, rubric–report alignment) and a Cohen's kappa of 0.78 for the rubric–report alignment variable, indicating substantial agreement; discrepancies were resolved through consensus coding and refinement of the codebook. These reliability measures increase confidence that the patterns reported above reflect systematic features of the corpus and not idiosyncratic coding decisions.

In sum, the analysis answers the research questions by showing how supervisors and examiners articulate rigour in dual registers, how written criteria correspond with enacted practice only partially, and how institutional and epistemic factors mediate which forms of scholarship are rewarded.

## 6. Discussion

This study reframes MA dissertation evaluation as an interpretive practice enacted through two registers — procedural defensibility and epistemic valuation — that interact through three mediating mechanisms (methodological legibility; supervisory socialisation; and board composition). Rather than merely documenting inconsistency, the dual-register framework suggests how and why rubrics are often mobilised instrumentally and why specification alone fails to eliminate variance: unless specification is accompanied by shared interpretive calibration (examiner workshops, exemplars, supervisor development), tacit evaluative repertoires continue to govern substantive judgments. The framework therefore foregrounds social and procedural

pairings as the route to greater fairness, a claim supported by case matrices and CDA of examiner reports.

From this conceptual vantage, three analytic insights follow. First, methodological legibility is not a neutral technical issue but an institutional currency: artifacts that reduce interpretive effort (tables, sampling frames, coding logs) become de facto tokens of rigour because they allow boards to adjudicate with low cognitive cost. Second, supervisory socialisation functions as a distributive mechanism: supervisors who convert tacit expectations into legible manuscript practices effectively confer evaluative advantage on their candidates. Third, board composition shapes interpretive latitude: internal-only configurations amplify local norms and reputational dynamics, whereas external perspectives tend to broaden interpretive repertoires. These insights are not additive descriptions; together they specify how the dual registers are realised in everyday assessment work.

The theoretical payoff of this account is practical as well as analytic. If rigour is produced through interacting registers and mediating mechanisms, then interventions that target only one surface (e.g., more detailed rubrics) will have limited effect. Instead, the model argues for paired reforms that simultaneously alter documentary expectations and shared interpretive practices: calibrated rubrics accompanied by annotated exemplars, regular examiner calibration sessions using anonymised exemplars, and structured supervisor development focused on documenting analytic processes. These interventions flow directly from the mechanisms identified by the dual-register framework, and they are testable in departmental or inter-departmental pilot studies.

Methodologically, the study illustrates the value of triangulating corpus-level description, close reading, and interview data to trace how discursive practices (examiner reports, supervisory edits) instantiate evaluative registers. Analytically, the dual-register framework can be applied beyond this single department: it provides a heuristic for comparative work that seeks to map how differing institutional architectures (e.g., use of external examiners, national QA regimes) shift the balance between procedural and interpretive registers.

The study is an in-depth single-department qualitative investigation, and our findings therefore reflect institutional configurations particular to the University of Batna 2 (internal-only boards, local language ecologies, supervisory practices). As a result, claims of broad generalisability are limited: the dual-register framework should be regarded as a theoretically informed heuristic that is transferable rather than directly generalizable. Transferability is achieved through detailed case vignettes, triangulation matrices, and explicit description of sampling strategies so that other researchers can assess fit with their contexts. Future comparative tests (multi-department or cross-national) are needed to evaluate whether the mechanisms operate similarly where external examiners are routine or where different quality-assurance regimes obtain.

In sum, this study's novel contribution is not only empirical description but conceptual translation: it turns the commonplace recognition that "rubrics are not everything" into a precise framework that explains how and why rubrics are incomplete, and it points

to interventions that address the root mechanisms by which evaluative judgements are produced.

## 7. Conclusion

This study reframes MA dissertation assessment as a process of epistemic adjudication in which written criteria, local practices, and interpersonal dynamics jointly determine what counts as rigour. Its core contribution is conceptual: by showing that evaluative judgements are enacted through distributed interpretive work, the study shifts the analytic focus from whether rubrics exist to how institutional arrangements and everyday practices translate those rubrics into outcomes. These reframing foregrounds the politics of interpretation rather than treating evaluation as a merely technical exercise.

Practically, the findings point to institutional reforms that operate at the level of shared interpretation and capacity rather than only at the level of paperwork. Departments seeking fairer, more transparent assessment should therefore prioritise measures that make tacit expectations explicit and that build collective reading practices among supervisors and examiners. Finally, while the single-department design limits claims of broad generalisability, the argument produces clear, testable propositions for comparative and experimental work—most pressingly, whether exemplar-based calibration and targeted supervisor development measurably reduce evaluative variance and unequal student burdens.

For policymakers, the study suggests that fairness requires both rule specification and capacity building: regulatory instruments (rubrics, examiner guidelines) should be coupled with funded examiner calibration and supervisor training. Implementing such paired reforms at departmental and national levels will be the clearest route to aligning written expectations with enacted judgements.

## References

Bastola, N., & Hu, G. (2020). Supervisory feedback across disciplines: Does it meet students' expectations? *Assessment and Evaluation in Higher Education*, *46*, 407–423. https://doi.org/10.1080/02602938.2020.1780562

Belcher, B., Rasmussen, K., Kemshaw, M., & Zornes, D. (2016). Defining and assessing research quality in a transdisciplinary context. *Research Evaluation*, *25*(1), 1–17. https://doi.org/10.1093/reseval/rvv025

Benbouabdallah, H., & Benmekhlouf, I. (2023). Teachers' opinions regarding the main standards for evaluating a master thesis: The case of EFL teachers at the Department of English, Batna 2 University. [Unpublished Master's dissertation, University of Batna 2, Batna, Algeria].

Bourdieu, P. (1988). *Homo academicus*. Stanford, United States: Stanford University Press.

Bourke, S., & Holbrook, A. (2013). Examining PhD and research masters theses. *Assessment and Evaluation in Higher Education*, *38*(4), 407–416. https://doi.org/10.1080/02602938.2011.638738

Bukhari, N., Jamal, J., Ismail, A., & Shamsuddin, J. (2021). Assessment rubric for research report writing: A tool for supervision. *Malaysian Journal of Learning and Instruction*, *18*(2), 1–43. https://doi.org/10.32890/mjli2021.18.2.1

Cheung, K. K. C. (2023). The use of intercoder reliability in qualitative interview data analysis in science education. International Journal of Science Education. https://doi.org/10.1080/02635143.2021.1993179

Chugh, R., Macht, S., & Harreveld, B. (2021). Supervisory feedback to postgraduate research students: A literature review. *Assessment and Evaluation in Higher Education*, *47*(5), 683–697. https://doi.org/10.1080/02602938.2021.1955241

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Crowe, M., Slater, P., & McKenna, H. 32(3). (2024). Demonstrating research quality. *Journal of Psychiatric and Mental Health Nursing*, *32*(3), 686-688. https://doi.org/10.1111/jpm.13145

Goodman, P., Robert, R., & Johnson, J. (2020). Rigor in PhD dissertation research. *Nursing Forum*, *55*(4). https://doi.org/10.1111/nuf.12477

Holbrook, A., Bourke, S., Lovat, T., & Dally, K. (2004). Investigating PhD thesis examination reports. *International Journal of Educational Research*, *41*, 98–120.

Homer, M., & Ababei, V. (2026). Evidencing improvement in examiner calibration in OSCEs. *Medical teacher*, 1–11. Advance online publication. https://doi.org/10.1080/0142159X.2026.2621959

Hsiao, Y. P. A. (2024). Ensuring bachelor's thesis assessment quality: A case study at a Dutch technical university. *Higher Education Evaluation & Development*, 18(1), 2–16. https://doi.org/10.1108/HEED-08-2022-0033

Knorr-Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge*. Cambridge, United States: Harvard University Press.

Kumar, V., & Stracke, E. (2011). Examiners' reports on theses: Feedback or assessment? *Journal of English for Academic Purposes*, *10*, 211–222. https://doi.org/10.1016/j.jeap.2011.06.001

Lee, A. (2018). How can we develop supervisors for the modern doctorate? *Studies in Higher Education*, *43*, 878–890. https://doi.org/10.1080/03075079.2018.1438116

Mafora, P., & Lessing, A. (2016). The voice of the external examiner: Experiences from South African higher education. *South African Journal of Higher Education*, *28*, 1295–1314.

Man, D., Xu, Y., Chau, M., O'Toole, J., & Shunmugam, K. (2020). Assessment feedback in examiner reports on master's dissertations in translation studies. *Studies in Educational Evaluation*, *64*, 100823. https://doi.org/10.1016/j.stueduc.2019.100823

Morse, J. M. (2015). Critical analysis of strategies for determining rigor in qualitative inquiry. *Qualitative Health Research*, *25*, 1212–1222.

Mullins, G., & Kiley, M. (2002). "It's a PhD, not a Nobel Prize": How experienced examiners assess research theses. *Studies in Higher Education*, *27*(3), 369–386. https://doi.org/10.1080/0307507022000011507

O'Donovan, B., Sadler, I., & Reimann, N. (2024). Social moderation and calibration versus codification: a way forward for academic standards in higher education? Studies in Higher Education, 49(12), 2693–2706. https://doi.org/10.1080/03075079.2024.2321504

Othman, J., & Lo, Y. (2023). Constructing academic identity through critical argumentation: A narrative inquiry of Chinese EFL doctoral students' experiences. *SAGE Open*, *13*. https://doi.org/10.1177/21582440231218811

Phuong, H., Phan, Q., & Le, T. (2023). The effects of using analytical rubrics in peer and self-assessment on EFL students' writing proficiency: A Vietnamese contextual study. *Language Testing in Asia, 13*. https://doi.org/10.1186/s40468-023-00256-y

Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment and Evaluation in Higher Education*. 35(4), 435–448. https://doi.org/10.1080/02602930902862859

Sadler, D. R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment and Evaluation in Higher Education*, *34*(2), 159–179. https://doi.org/10.1080/02602930801956059

Stigmar, M. (2018). Learning from reasons given for rejected doctorates: Drawing on some Swedish cases from 1984 to 2017. *Higher Education, 77*, 1031–1045. https://doi.org/10.1007/s10734-018-0318-2

Tan, W. C. (2024). Empowering examiners to develop doctoral assessment literacy: A situated learning perspective. *Frontiers in Education, 9*. https://doi.org/10.3389/feduc.2024.1345661

Tiwari, H. (2024). Behind the curtain: External Examiners' Experiences about Thesis Evaluation. *Shanti Journal*, *4*(1). https://doi.org/10.3126/shantij.v4i1.70529

Varela, M., Lopes, P., & Rodrigues, R. (2021). Rigour in the management case study method: A study on master's dissertations. *The Electronic Journal of Business Research Methods*, *19*, 1–13.

Vita, G., & Begley, J. (2023). A framework of 'doctorateness' for the social sciences and postgraduate researchers' perceptions of key attributes of an excellent PhD thesis. *Studies in Higher Education*, *49*, 1884–1899. https://doi.org/10.1080/03075079.2023.2281540

Yadav, D. (2021). Criteria for good qualitative research: A comprehensive review. *The Asia-Pacific Education Researcher, 31*, 679–689. https://doi.org/10.1007/s40299-021-00619-0

## Appendices

### Appendix A: Corpus Coding Sheet (Tier 1: corpus-level coding)

| FIELD | VALUES / FORMAT | NOTES |
|---|---|---|
| **ANONYMISED ID** | e.g., T001 – T120 | Unique identifier |
| **SUBMISSION YEAR** | 2023 / 2024 / 2025 | |
| **DEPARTMENTAL OPTION** | LLA / LC / Didactics | |
| **METHODOLOGY TYPE** | Qualitative / Quantitative /Mixed / Theoretical | Choose best fit |
| **EXPLICIT RESEARCH QUESTION (S)** | Yes / No | Binary flag |
| **EXPLICIT THEORETICAL FRAMEWORK** | Yes / No | Binary flag |
| **NUMBER OF REFERENCES (BAND)** | 0–49 / 50–99 / 100+ | Banding only |
| **GRADE BAND** | 10–11 / 12–14 / 15–17 | Department scale 0–20 |
| **SUPERVISOR EDITS VISIBLE** | Yes / No / Unknown | Tracked changes, marginalia, document properties |

### Appendix B: Close-Read Coding Frame (Tier 2: qualitative close reading)

1. **Case ID**
2. **Methodological orientation** — Qual / Quant / Mixed / Theoretical
3. **Clarity of research questions** — None / Weak / Clear / Strong (brief justification)
4. **Theoretical engagement** — Absent / Descriptive / Moderate / Strong (examples of conceptual moves)
5. **Analytic transparency (Qualitative)**— None / Minimal / Adequate / Exemplary (are coding steps documented?)
6. **Data quality (quantitative)** — Poor / Adequate / Robust (sampling frame, response rate)

7. **Evidence legibility** — Low / Medium / High (presence of tables, appendices, logs)
8. **Originality/contribution** — None / Modest / Clear / Strong (brief memo)
9. **Citation practice** — Problematic / Adequate / Exemplary (consistency, up-to-date sources)
10. **Writing quality** — Poor / Acceptable / Good / Excellent (clarity, argument flow)
11. **Examiner comments: main concerns** — free text (extract key phrasing)
12. **Supervisor-mediated edits visible** — Yes / No / Unclear (evidence and type)
13. **Rubric–report alignment** — Aligned / Partially aligned / Misaligned (brief memo)
14. **Interpretive memo** — 3–6 sentences synthesising how features map to outcomes

**Appendix C: Interview Guide (Supervisors and Examiners)**

**Introductory text (read aloud):**
"Thank you for participating. I will ask about your experiences supervising/examining MA dissertations at Batna 2 University's English Department. Your responses will be anonymised. You may decline to answer any question. With your permission I will record this interview."
**Questions:**

1. How do you define "rigour" when assessing an MA thesis in your department? (Prompt: methodological, theoretical, analytical, ethical dimensions)
2. What written criteria or rubrics do you refer to when assessing a thesis? How useful are they in practice?
3. Can you describe a recent thesis you examined or supervised that you regarded as rigorous? What features led you to that judgement?
4. Can you describe a recent thesis you felt lacked rigour? What specifically was missing or unclear?
5. How do issues of language (EFL) affect your assessment of substance vs. presentation? Do you separate language proficiency from epistemic contribution? How?
6. What role does supervisory work play in preparing theses for examination? Can you give examples of specific editorial or pedagogic interventions you provide?
7. How often do you cite the rubric in your written report? In what circumstances do you rely on tacit judgement instead?
8. Do you think internal-only boards shape how you evaluate theses? If yes, how?
9. Would annotated exemplars or examiner calibration sessions be useful? Why or why not?
10. Is there anything else you would like to add about standards of rigour, fairness, or improvements for thesis assessment?

**Closing:** Thank participant, remind about anonymisation and offer summary of findings.

**Appendix D: Rubric (Department guideline extract and annotated exemplar template)**
**Example condensed rubric (short form)**

| CRITERION | EXCELLENT (16–20) | SATISFACTORY (12–15) | REVISION REQUIRED (10–11) |
|---|---|---|---|
| **RESEARCH QUESTIONS & OBJECTIVES** | Clear, novel, well-justified | Clear but limited in novelty | Absent or vague |
| **THEORETICAL FRAME WORK** | Sophisticated integration, critical engagement | Present but descriptive | Absent or superficial |
| **METHODOLOGY & ANALYTIC TRANSPARENCY** | Appropriate, fully documented (appendices/ coding logs) | Adequate Description but some gaps | Major omissions, unclear procedures |

| | | | |
|---|---|---|---|
| **DATA & EVIDENCE** | Robustly presented (tables/ figures), logically connected to claims | Sufficient evidence, occasional gaps | Weak or missing evidence |
| **ARGUMENT & CONTRIBUTION** | Coherent, persuasive, clearly situated in literature | Reasonable argument, limited contribution | Fragmented, descriptive |
| **WRITING & PRESENTATION** | Excellent academic writing, accurate referencing | Acceptable, minor language issues | Major language/ Presentation issues |

**Appendix E: Triangulation Matrix Template (case-level)**

| RUBRIC ITEM | THESIS EVIDENCE | EXAMINER COMMENT (QUOTE) | SUPERVISOR CLAIM (INTERVIEW EXTRACT) | INTERPRETATION (HOW EVIDENCE + CLAIMS EXPLAIN OUTCOME) |
|---|---|---|---|---|
| **RESEARCH QUESTION CLARITY** | e.g., Chapter 1, p. 3: "..." | "RQ unclear" | "We focused on framing, not RQ" | e.g., RQ absent; supervisor prioritized framing; led to revision request |
| **THEORETICAL ENGAGEMENT** | ... | ... | ... | ... |
| **METHODS TRANSPARENCY** | ... | ... | ... | ... |
| **EVIDENCE PRESENTATION** | ... | ... | ... | ... |
| **OVERALL ALIGNMENT WITH RUBRIC** | Aligned / Partially aligned / Misaligned | ... | ... | Summary |

**Appendix F: Codebook extract and intercoder reliability protocol**

**Codebook extract:**

- **Code:** Analytic transparency (qualitative)
- **Coding rules:** 0 = none; 1 = minimal (mentions coding but no detail); 2 = adequate (describes steps and offers one example); 3 = exemplary (codebook + examples + audit trail).
- **Code:** Evidence legibility
- **Definition:** Presence of artefacts that render claims directly verifiable (tables, appendices, code logs). 0–3 as above.

**Inter-coder reliability protocol:**

1. Pilot coding on 10% of the corpus (12 cases).
2. Two coders independently code pilot set.
3. Calculate raw agreement and Cohen's kappa for key binary/ordinal variables.
4. Convene meeting to resolve discrepancies and refine code definitions.
5. Re-code disputed items and finalize codebook.
6. Proceed to full coding with periodic cross-checks on 10% random sample.
7. Report statistics (raw agreement; Cohen's kappa) in Methods appendix.

**Appendix G: Descriptive tables and codebook**

**Distribution of methodological orientation by presence of examiner revision requests (n = 120)**

*Counts and column percentages are presented. "Revision requested — Yes" denotes any examiner or board request for revision prior to final acceptance.*

| METHODOLOGY | REVISION REQUESTED — YES (N, %) | REVISION REQUESTED — NO (N, %) | TOTAL (N) |
|---|---|---|---|
| QUANTITATIVE | 20 (33.3%) | 40 (66.7%) | 60 |
| MIXED METHODS | 22 (61.1%) | 14 (38.9%) | 36 |
| QUALITATIVE | 18 (75.0%) | 6 (25.0%) | 24 |
| TOTAL | 60 (50.0%) | 60 (50.0%) | 120 |

**Cross-tabulation of methodological orientation and grade band (Low / Middle / Upper) (n = 120)**

*Counts and row percentages are presented. Grade bands are defined according to final board-assigned categories recorded in repository metadata.*

| METHODOLOGY | LOW | MIDDLE | UPPER | TOTAL |
|---|---|---|---|---|
| QUANTITATIVE | 6 (10%) | 36 (60%) | 18 (30%) | 60 |
| MIXED METHODS | 4 (11.1%) | 22 (61.1%) | 10 (27.8%) | 36 |
| QUALITATIVE | 6 (25%) | 10 (41.7%) | 8 (33.3%) | 24 |
| TOTAL | 16 (13.3%) | 68 (56.7%) | 36 (30.0%) | 120 |

**Presence of key evidentiary features by methodological orientation (n = 120)**

*Counts and column percentages are presented. Features were coded as present/absent according to the codebook.*

| FEATURE / METHOD | QUANTITATIVE (N, %) | MIXED (N, %) | QUALITATIVE (N, %) | TOTAL (N, %) |
|---|---|---|---|---|
| AUDIT TRAIL / ANALYTIC LOG PRESENT | 38 (63.3%) | 18 (50.0%) | 6 (25.0%) | 62 (51.7%) |
| CODEBOOK / CODING APPENDIX PRESENT | 8 (13.3%) | 6 (16.7%) | 4 (16.7%) | 18 (15.0%) |
| CLEAR SAMPLING FRAME / TABLE | 50 (83.3%) | 24 (66.7%) | 4 (16.7%) | 78 (65.0%) |
| STATISTICAL OUTPUTS / DETAILED TABLES | 56 (93.3%) | 30 (83.3%) | 2 (8.3%) | 88 (73.3%) |