



Оригинальное исследование

УДК 371

DOI: 10.18413/2313-8971-2025-11-1-0-4

S. Tobbi 

**What counts as academic rigour? Epistemic politics  
in MA dissertation assessment in an Algerian EFL department**

University of Batna 2  
53 Constantine Rd., Fesdis, Batna 05078, Algeria  
s.tobbi@univ-batna2.dz

*Received on December 31, 2025; accepted on March 17, 2026;  
published on March 31, 2026*

**Abstract:** *Introduction.* Academic rigour is central to graduate assessment but how written rubrics actually translate into examiners' judgements remains under-theorized. *Objective.* This paper investigates how standards of academic rigour are both articulated in written policy and enacted in practice when assessing Master of Arts dissertations. *Materials and Methods.* Drawing on a qualitative, multi-methods study conducted at the English Department, University of Batna 2, this research project employs a purposeful corpus comprising 120 Master's dissertations that were submitted between 1 May 2023 and 30 June 2025. Additionally, it incorporates the examiners' reports and semi-structured interviews with 12 supervisors and 13 examiners. A stratified sub-sample of 36 dissertations was analysed in depth. Data were examined through document analysis, thematic coding and cross-source triangulation to map written criteria against evaluative practice. *Results.* The results show that, although official rubrics supply clear procedural criteria, evaluators frequently rely on tacit interpretive standards so that policy and practice align only partially. Three interrelated mechanisms explain this divergence: methodological legibility (how clearly methodological choices make a thesis readable and defensible), supervisory socialisation (the informal norms supervisors transmit), and internal board composition (the mix of examiners' expertise and expectations). *Conclusion.* We argue that improving fairness and consistency requires calibrated rubrics augmented with annotated exemplars, routine examiner-calibration workshops, and targeted supervisor development to increase analytic transparency. The study's significance lies in offering an empirically grounded account of the policy – practice gap, providing concrete interventions for institutional assessment and quality-assurance, and setting an agenda for comparative and experimental research to evaluate the effectiveness of the proposed measures.

**Keywords:** academic rigour; dissertation assessment; higher education evaluation; examiner guidelines; rubric calibration

**Information for citation:** Saida Tobbi (2026), "What counts as academic rigour? Epistemic politics in MA dissertation assessment in an Algerian EFL department", *Research Result. Pedagogy and Psychology of Education*, 12 (1), 47-59, DOI: 10.18413/2313-8971-2026-12-1-0-4.

Тобби С. 

## Что считается академической строгостью? Эпистемическая политика в оценивании магистерских диссертаций на алжирской кафедре EFL

Университет Батна 2  
ул. Константина, 53, Фесдис, Батна, 05078, Алжир  
s.tobbi@univ-batna2.dz

*Статья поступила 31 декабря 2025; принята 17 марта 2026;  
опубликована 31 марта 2026*

**Аннотация.** *Введение.* Академическая строгость играет центральную роль в оценивании магистерских работ, однако то, как письменные рубрики фактически трансформируются в суждения экзаменаторов, пока недостаточно теоретически осмыслено. *Цель.* В статье исследуется, как стандарты академической строгости формулируются в письменной политике и как они реализуются на практике при оценивании магистерских диссертаций. *Материалы и методы.* Опираясь на качественное многометодное исследование, проведённое на кафедре английского языка Университета Батна 2, работа использует целевой корпус из 120 магистерских диссертаций, защищённых в период с 1 мая 2023 года по 30 июня 2025 года, отчёты экзаменаторов по ним, а также полуструктурированные интервью с 12 научными руководителями и 13 экзаменаторами; углублённому анализу был подвергнут стратифицированный подвыборочный массив из 36 диссертаций. Данные были проанализированы посредством документ-анализа, тематического кодирования и триангуляции по различным источникам с целью сопоставления письменных критериев с оценочной практикой. *Результаты* показывают, что, хотя официальные рубрики содержат чёткие процедурные критерии, оценщики нередко опираются на неявные интерпретативные стандарты, вследствие чего политика и практика совпадают лишь частично. Это расхождение объясняется тремя взаимосвязанными механизмами: методологической читаемостью (тем, насколько ясно методологические решения делают диссертацию понятной и подлежащей защите), социализацией через научное руководство (неформальными нормами, передаваемыми руководителями), и внутренним составом комиссии (сочетанием экспертного опыта и ожиданий экзаменаторов). *Заключение.* Мы утверждаем, что повышение справедливости и согласованности требует калиброванных рубрик, дополненных аннотированными образцами, регулярных семинаров по согласованию оценивания между экзаменаторами и целенаправленного развития компетенций научных руководителей для повышения аналитической прозрачности. Значимость исследования состоит в том, что оно предлагает эмпирически обоснованное описание разрыва между политикой и практикой, предоставляет конкретные меры для институционального оценивания и обеспечения качества, а также задаёт повестку для сравнительных и экспериментальных исследований по проверке эффективности предложенных мер.

**Ключевые слова:** академическая строгость; оценивание диссертации; оценивание в высшем образовании; рекомендации для экзаменаторов; калибровка рубрики.

**Информация для цитирования:** С. Тобби (2026), “Что считается академической строгостью? Эпистемическая политика в оценивании магистерских диссертаций на кафедре EFL в Алжире”, *Научный результат. Педагогика и психология образования*, 12(1), 3-15, DOI: 10.18413/2313-8971-2026-12-1-0-4.

**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 addresses these issues through a multi-method qualitative investigation 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?

In response to these questions, the primary objective of this study is to investigate how standards of academic rigour are articulated in

departmental rubrics and enacted in examiner and supervisory practices, and to identify the institutional and epistemic mechanisms that mediate the translation of written criteria into evaluative outcomes.

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.

**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. 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. 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).

**Methodology and methods. 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.

**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.

#### **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.

Supervisors and internal examiners associated with the close-read theses 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.

#### **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.

**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 (see [Appendix A](#)); 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 (see [Appendix B](#)). 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 (see [Appendix C](#)).

**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 (see [Appendix D](#)).

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 (see [Appendix E](#)).

**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) (see [Appendix F](#)). These matrices are the primary analytic devices used to answer RQ2 and to illustrate correspondences or divergences between stated criteria and enacted practice.

**Trustworthiness and reflexivity.** Trustworthiness was enhanced through triangulation across documents and interviews, thick description and rich case vignettes, inter-coder checks on purposive subsets of close-read cases, and an audit trail that recorded sampling decisions and coding adjustments. Member

checking was offered to a limited set of participants in the form of anonymised summaries of emergent findings. Reflexive memos documented the researcher's positionality and analytic decisions throughout the project.

**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.

**Research Results and Discussion.** This section presents the empirical analysis first and then interprets those findings. 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.

Triangulation across corpus analysis, examiner reports and interviews reveals that rigour is articulated and enacted through two interlocking registers: procedural compliance

and epistemic contribution. Departmental templates and written guidelines supply a formal vocabulary centred on explicit research questions, methodological transparency and structural completeness, and examiner reports frequently invoke these elements as justificatory resources in accept or revise decisions. However, both examiner discourse and interview accounts show that these procedural criteria function primarily as instruments of defensibility in board deliberations, while substantive judgements often hinge on tacit normative assessments such as analytical depth, conceptual engagement and intellectual contribution. The convergence of corpus-level templates, rubric-consistent examiner language and explicit examiner reflections confirms that procedural criteria provide the formal scaffold of evaluation, but tacit interpretive judgements ultimately shape how rigour is enacted in practice.

These evaluative dynamics are closely linked to the evidentiary legibility of methodological choices. While the corpus distribution indicates that quantitative designs are prevalent (50%, with a further 30% using mixed methods), close analysis of examiner reports reveals that methodological form alone does not guarantee evaluative security. In fact, 56% of quantitative dissertations were explicitly asked to provide additional procedural detail, including clearer sampling procedures, analytic steps and reliability measures. Examiner interviews clarify the mechanism underlying this pattern: boards privilege forms of evidence that are visibly traceable and auditable, such as statistical tables, coding frameworks and explicit procedural chains. Methodological rigour, therefore, is not evaluated solely on epistemological grounds but also on how clearly the evidentiary process is rendered visible and defensible to evaluators. Evidentiary legibility emerges as a key currency in assessment, shaping how methodological choices are interpreted and valued.

Supervisory practice plays a central role in mediating this legibility by translating tacit

departmental expectations into concrete manuscript features. In the close-read subsample, 15 of 36 dissertations (42%) showed clear evidence of supervisor-driven restructuring prior to submission, including reorganised method sections, clarified analytic procedures and documented supervisory annotations. Interview data corroborate that supervisors actively socialise students into the implicit evaluative norms of the department. As one senior supervisor explained, “Part of supervision is showing students how to package their work so it reads as research that can be defended” [Sup-04]. Through this process, supervisors effectively convert implicit evaluative criteria into explicit textual features, thereby shaping how dissertations are subsequently interpreted by examination boards. Supervisory intervention thus functions as an informal standardisation mechanism, influencing whether a dissertation will be read charitably or critically.

At the same time, analysis reveals a persistent gap between written rubric criteria and the interpretive standards that guide actual evaluation. In the stratified close-read sample, explicit alignment between rubric criteria and examiner justification was observed in only 19 of 36 cases (53%), while in the remaining 17 cases (47%), examiner reports foregrounded evaluative criteria not explicitly specified in the rubric. Interviewees acknowledged that rubric language is often invoked strategically, particularly in borderline decisions, to provide procedural legitimacy. As one examiner noted, “We quote the guideline to be transparent; sometimes it is our shield”. Critical discourse analysis shows that rubric references and tacit evaluative idioms coexist within examiner reports, but serve different functions: rubric language legitimates administrative decisions, while tacit criteria articulate substantive scholarly value. This pattern demonstrates that rubrics do not fully constrain evaluation but operate alongside interpretive practices shaped by local epistemic norms.

Two case vignettes illustrate how these mechanisms operate in practice. Dissertation

T12-LLA-2024 presented a novel interpretive argument but lacked detailed documentation of coding procedures. The examiner report acknowledged the conceptual originality but recommended revisions because the evidentiary chain was insufficiently visible, noting that the “claims are interesting but lack a clear coding trail”. The supervising lecturer confirmed that supervision had prioritised argumentation and framing, and that the student had not anticipated the board’s demand for explicit procedural traceability. In contrast, dissertation T07-DID-2023 presented a clearly structured quantitative design with explicit sampling and transparent analytic steps. Examiner reports praised the “robust chain of evidence” despite limited theoretical originality, and interview commentary confirmed that evidentiary visibility facilitated confident evaluation. Taken together, these cases demonstrate that rigour is not determined solely by conceptual innovation or methodological choice, but by the extent to which scholarly claims are rendered procedurally legible within the department’s evaluative framework.

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.

This study reframes MA dissertation evaluation as an active interpretive process in which written instruments, local practices, and interpersonal dynamics collectively determine what counts as rigour. Rather than simply documenting inconsistency between rubrics and practice, the paper offers a new explanatory lens: the dual-register framework. This framework – which distinguishes a procedural, rubric-oriented register from a tacit, interpretive register and shows how they interact as mechanisms of justification and exclusion – is the paper's principal conceptual contribution.

The originality of the dual-register claim lies in three moves. First, it formalises an analytic distinction that prior work has only observed discursively: we do not merely note that discretion exists; we specify the registers through which discretion is enacted and the functional roles those registers play in adjudication (defensibility vs. substantive valuation). Second, it links those registers to concrete mechanisms – methodological legibility, supervisory socialisation, and board composition – thereby moving from description to causal explanation. Third, it shows empirically how the registers are invoked strategically (for example, rubrics are cited as administrative cover even when tacit evaluative judgments determine the substantive outcome), which explains persistent rubric–reality divergence. In short, the dual-register framework translates a vague observation about “discretion” into an operational model for understanding and changing evaluative practice.

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.

Limitations are modest but important: the single-department focus constrains claims of generalisability, and the absence of external examiner data in this context shapes the specific configurations observed. Nonetheless, the explanatory scope of the dual-register framework suggests clear pathways for future research – experimental evaluations of exemplar-based calibration, longitudinal ethnographies of supervisory socialisation, and comparative studies across institutional types – each of which can test whether changing interpretive practice reduces evaluative variance.

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.

**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. This 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. and Hu, G. (2020), “Supervisory feedback across disciplines: Does it meet students’ expectations?”, *Assessment & Evaluation in Higher Education*, 46, 407-423. <https://doi.org/10.1080/02602938.2020.1780562>. (In UK).
- Belcher, B., Rasmussen, K., Kemshaw, M. and Zornes, D. (2016), “Defining and assessing research quality in a transdisciplinary context”, *Research Evaluation*, 25(1), 1-17. <https://doi.org/10.1093/reseval/rvv0.25>. (In UK).
- Benbouabdallah, H. and 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 University Press, Stanford, United States.
- Bourke, S. and Holbrook, A. (2013), “Examining PhD and research masters theses”, *Assessment & Evaluation in Higher Education*, 38(4), 407-416. <https://doi.org/10.1080/02602938.2011.638738>. (In UK).
- Bukhari, N., Jamal, J., Ismail, A. and 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>. (In Malaysia).
- Chugh, R., Macht, S. and Harreveld, B. (2021), “Supervisory feedback to postgraduate research students: A literature review”, *Assessment & Evaluation in Higher Education*, 47(5), 683-697.

<https://doi.org/10.1080/02602938.2021.1955241>. (In UK).

Braun, V. and Clarke, V. (2006), "Using thematic analysis in psychology", *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706qp063oa>. (In UK).

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

Goodman, P., Robert, R. and Johnson, J. (2020), "Rigor in PhD dissertation research", *Nursing Forum*, 55(4). <https://doi.org/10.1111/nuf.12477>. (In UK/In USA)

Holbrook, A., Bourke, S., Lovat, T. and Dally, K. (2004), "Investigating PhD thesis examination reports", *International Journal of Educational Research*, 41, 98-120. <https://doi.org/10.1016/j.ijer.2005.04.008>. (In Netherlands).

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

Kumar, V. and 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>. (In UK).

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>. (In UK).

Mafora, P. and Lessing, A. (2016), "The voice of the external examiner: Experiences from South African higher education", *South African Journal of Higher Education*, 28, 1295-1314. <https://doi.org/10.20853/28-4-389>. (In South Africa).

Man, D., Xu, Y., Chau, M., O'Toole, J. and 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>. (In UK/In Netherlands).

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

1222. <https://doi.org/10.1177/1049732315588501>. (In USA).

Mullins, G. and 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>. (In UK).

Othman, J. and 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>. (In USA).

Phuong, H., Phan, Q. and 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>. (In UK).

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

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

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

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

Varela, M., Lopes, P. and 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. (In UK).

Vita, G. and 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>. (In UK).

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>. (In Singapore/In Philippines).

**Информация о конфликте интересов:** автор не имеет конфликта интересов для декларации.

**Conflicts of Interest:** the author has no conflict of interest to declare.

**About the author:**

**S. Tobbi**, Associate Professor, PhD, Batna 2 University, Faculty of Letters and Foreign Languages, Department of English Language and Literature, Algeria.

**Тобби С.**, доцент, кандидат наук, Университет Батна 2, факультет литературы и иностранных языков, кафедра английского языка и литературы, Алжир.