Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Visible Governance, Hidden Routines: A Three-Institution Pilot of Authorship Ethics in Saudi-Affiliated Linguistics

Hashim Asiri

hashim.asiri@gmail.com

King Khalid University

Research Article

Keywords: authorship misappropriation, research integrity, mixed methods, linguistics, Saudi Arabia

Posted Date: August 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10362331/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Visible Governance, Hidden Routines: A Three-Institution Pilot of Authorship Ethics in Saudi-Affiliated Linguistics

Hashim Asiri
Faculty of English and Translation, King Khalid University, Abha, Saudi Arabia
Corresponding author: hashim.asiri@gmail.com
ORCID: https://orcid.org/0000-0001-8924-6151

Approximate word count: 6,000 words, excluding abstract, declarations, references, and figure legends.

Abstract

Background: Institutional governance of research integrity has expanded across higher-education systems, yet authorship remains difficult to regulate because it allocates credit, responsibility, and accountability simultaneously. This study examines what recurrent authorship patterns in Saudi-affiliated linguistics reveal about the limits of publicly visible governance.

Methods: A corpus of 209 papers (2021–2026) from three Saudi institutions was retrieved via OpenAlex. Eighty-six papers received Authorship Misappropriation Diamond (AMD)-informed screening. Inter-coder reliability was assessed on 16 multi-author papers (Cohen's κ moderate to perfect across directly coded dimensions). Eight high-priority cases underwent contextual validation through full-text review, institutional-policy comparison, and author-network analysis.

Results: Five cases retained high residual ethical concern; three retained moderate-high concern. Cases resistant to explanation combined structural recurrence with paper-level textual or funding irregularities and an absence of publicly articulated institutional rules governing the relevant authorship practice.

Conclusions: Findings do not establish misconduct or institutional causation. They identify configurations warranting editorial or qualitative follow-up and show how visible integrity policy may leave author-order and contribution-attribution practices under-specified. The study

contributes a bounded, configuration-level approach while clarifying the inferential limits of publication metadata.

Keywords

authorship misappropriation; research integrity; mixed methods; linguistics; Saudi Arabia

Introduction

Authorship credit, contributorship, and the institutional governance of research integrity have all expanded substantially in recent years, yet authorship ethics remain difficult to study empirically at the institutional level. The growth of team-based publication has made the allocation of credit increasingly consequential [1], while studies continue to document honorary, ghost, and contested authorship [2, 3]. Author count, order, and contribution statements remain field-dependent and imperfect measures of contribution [4, 5]. The present study therefore asks what recurrent authorship patterns reveal about institutional conditions without treating recurrence as misconduct. Restricting the dataset to Saudi-affiliated linguistics reduces some cross-disciplinary noise and makes repeated pair, team, and terminal-author structures more interpretable within a narrower scholarly culture. The goal is not to estimate national prevalence, but to test whether a structured mixed-method design can identify patterns that warrant closer contextual examination.

The analytical anchor for the study is the Authorship Misappropriation Diamond (AMD) framework proposed by Oyenuga et al. [6]. AMD extends the Fraud Diamond proposed by Wolfe and Hermanson [7] by adding normalisation to explain how problematic authorship may become routinised within a publication environment. In this study, AMD is used as a sensitising and screening framework rather than as a mechanism for diagnosing misconduct. It organises the

examination of recurrent structures while requiring each dimension to remain bounded by what the available evidence can support.

The ethical question this study sets out to address is therefore not whether misconduct can be detected from publication metadata (it cannot), but whether the ethical conditions under which authorship operates can be made analytically visible in a setting where institutional governance has become increasingly formalised. Saudi higher education has, in the past decade, expanded its visible integrity infrastructure through national policy, institutional ethics review, and explicit publication-quality incentives. Whether this visible governance reaches the authorship routines that productivity-driven publication environments tend to produce is an ethical question rather than a methodological one. The present study takes one field, Saudi-affiliated linguistics, as a bounded site for asking it.

Three sub-questions structure the analysis: how recurrent structural patterns distribute across institutions in a field-bounded national setting; which configurations survive paper-level contextualisation (full-text review, institutional-policy comparison, author-network analysis) as ethically interpretable rather than dissolving into ordinary collaboration; and what the contrast between surviving and dissolving configurations says about institutional policy visibility and authorship ethics. The study uses the Authorship Misappropriation Diamond (AMD) framework [6] as the conceptual instrument, but the analytic centre is the ethical interpretation of what the framework allows the data to reveal, not the framework itself.

What follows is therefore not merely a methodological pilot, even though it documents its methodological discipline carefully. It is an empirical contribution to the literature on authorship ethics under conditions of visible institutional governance, with implications for how journals,

institutions, and integrity offices might recognise the configurations in which authorship recurrence becomes ethically interpretable. The Saudi context is the bounded site; the contribution is to the broader field. Validation cases are reported through pseudonymous institution and cluster labels rather than direct author names, paper titles, or journal names, to minimise identifiability in a misconduct-adjacent pilot design.

AMD as an Explanatory Framework

First, authorship functions simultaneously as scholarly credit, responsibility, and institutional currency. Priority and recognition have long structured the scientific reward system [8–10], while authorship itself remains unstable because the same byline must allocate intellectual credit, accountability, and career value [11, 12]. In evaluative environments, that credit acquires exchange value: publication records influence employment, promotion, reputation, and access to resources [13–15]. Pressure in the AMD framework is therefore treated here as an institutional condition under which authorship becomes more valuable, not as a directly observed motive of any author.

Second, the allocation of authorship is relational and potentially hierarchical. Accumulated recognition can reproduce further advantage through the Matthew effect [9], while career evaluation systems can make first, last, and shared authorship positions consequential in ways that vary by discipline [16, 17, 5]. Capability is consequently used in this study as a sensitising concept for the structural position from which credit may be influenced. A recurrent terminal-author position is only a proxy for such positioning; it is not evidence that authority was exercised.

Third, formal integrity systems operate through what audit-culture scholarship describes as visible procedures, documentary compliance, and commensurable indicators [18–20]. Authorship standards and contributorship taxonomies make responsibility more explicit [21–25], while

responsible-metrics principles warn against treating indicators as self-interpreting evidence [26]. Formal visibility does not guarantee that local allocation routines are captured by those standards. Normalisation is therefore approached as a possible institutional process suggested by recurrence and procedural regularity, while recognising that repeated collaboration may also reflect legitimate disciplinary or supervisory practice.

Read together, these strands position AMD as a sociological sensitising framework rather than merely a metadata scaffold. Productivity regimes can make authorship increasingly valuable as institutional currency; positional structures can shape how that currency is distributed; and repeated procedures can become ordinary without necessarily becoming ethically justified. The present data can identify configurations consistent with these processes, but cannot demonstrate that the processes occurred. The validation cases therefore test where recurrence remains insufficiently explained after contextual examination, while preserving ordinary collaboration as a competing explanation (Figure 1).

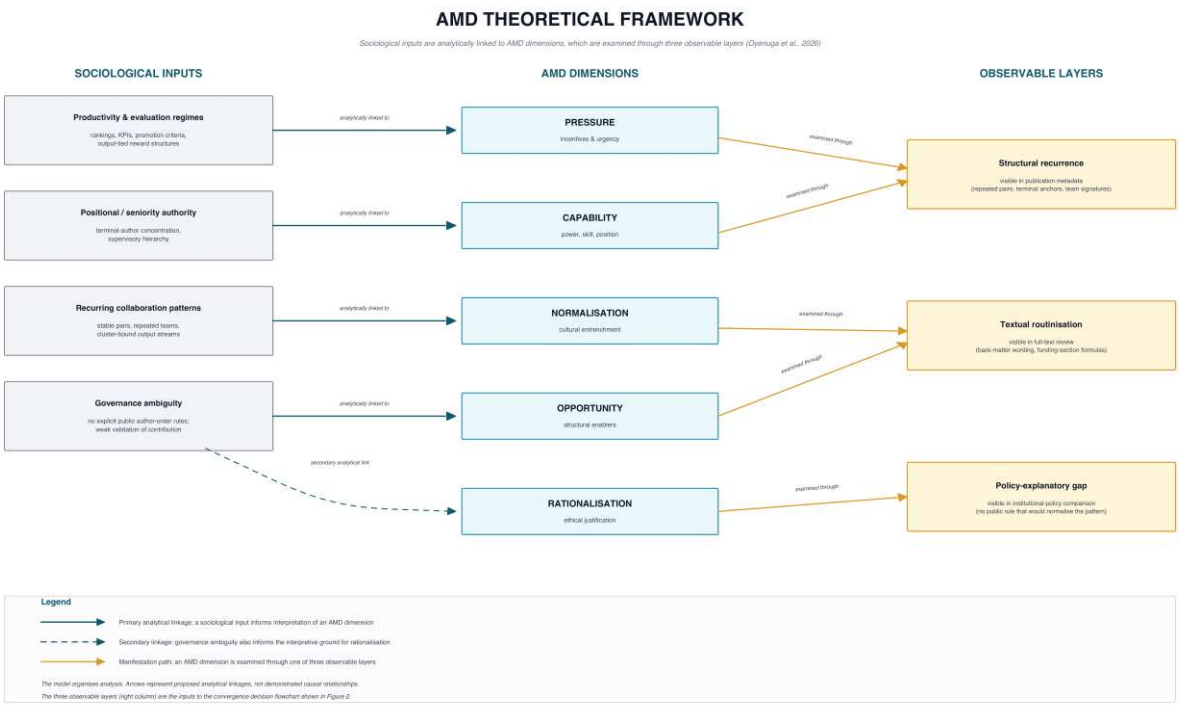


Figure 1. The AMD framework as a sensitising model. Sociological inputs are analytically linked to AMD dimensions (centre, following Oyenuga et al. [6]), which guide interpretation across three observable layers (right). Arrows represent proposed analytical linkages and interactions rather than demonstrated causal relationships. The model organises the analysis and does not establish that these processes occurred in any individual case.

Saudi Integrity and Governance Context

This study analyses authorship patterns in Saudi-affiliated linguistics publications using the AMD framework. It is not a retrospective audit of compliance with any specific national policy. The Research, Development and Innovation Authority's 2024 national framework [27] is used as contextual evidence relevant to the publication environment, not as a compliance instrument for individual papers. Saudi ethics infrastructure includes expanding institutional review arrangements and researcher-facing governance systems [28, 29], but those developments do not establish how authorship was negotiated in the publications examined here.

The integrity-framework expansion sits within Saudi higher education's broader transformation toward research productivity and global ranking targets. Vision 2030 [30] and the National Transformation Program [31] set explicit ranking-competition goals; Smith and Abouammoh [32],

Onsman [33], and Allmnakrah and Evers [34] document the trajectory from teaching-oriented to output-evaluative, with substantial institutional incentive pressure on publication output [35–37].

The institutional consequences of ranking-driven productivity targets are well-documented [38, 39]. Espeland and Sauder's [40] reactivity argument applies particularly to settings where ranking aspirations are policy-level: institutions exposed to a public ranking instrument adapt their behaviour to its measurement criteria. The Saudi case operationalises this with ranking-visibility goals embedded in national policy rather than left to institutional choice [34, 32].

The broader Arab-region publication culture shapes venue, team composition, and language. Hanafi [41] frames the structural dilemma as a choice between publishing globally and perishing locally and publishing locally and perishing globally. Hanafi and Arvanitis [42, 43] extend this to language and regional knowledge production. Regional empirical work on authorship credit [44] shows that honorary-authorship perceptions are part of the working repertoire of Middle Eastern academic culture.

Saudi-affiliated linguistics is therefore not simply a bounded site that happens to be analytically convenient. It is a setting in which several sociological pressures intersect in unusually visible form: a nationally articulated productivity-and-ranking regime; an integrity-governance infrastructure that has formalised rapidly in the last few years; a regional publication culture in which the publish-globally-and-perish-locally dilemma shapes author behaviour at the venue and team level; and a recent policy framework that emphasises scientific integrity without explicitly addressing the author-order and contribution-attribution practices through which the integrity claim is operationalised. The combination is what makes the case ethically informative beyond Saudi Arabia: the same configuration, which combines visible integrity policy, structurally

constrained productivity incentives, and a field-bounded publication culture, exists in many higher-education systems undergoing similar transformations. What this pilot can show about authorship ethics in this setting is therefore relevant to the broader question of how authorship integrity is constituted, governed, and rendered ethically interpretable under conditions of rapidly expanding research output.

Methods

The methods are organised around study scope, retrieval and field screening, structural enrichment, threshold coding, reliability assessment, and contextual validation.

Study Design

This study used a sequential mixed-method screening and contextual-validation design. The first phase retrieved, screened, deduplicated, and enriched Saudi-affiliated linguistics records with authorship-structure indicators. The second phase applied AMD-informed threshold coding to a mechanically prioritised subset. The third phase examined eight high-priority cases through full-text review, public-policy comparison, and author-network analysis. Integration occurred in an evidence matrix that recorded whether each additional layer weakened, sustained, or complicated the initial signal. Metadata and recurrence were treated as screening indicators, not as proof of authorship misappropriation.

Institutional Scope and Time Window

The pilot was restricted to three Saudi universities, reported as Institution X, Institution Y, and Institution Z, and to publications from 2021 through April 2026. The institutions were selected as analytically useful anchors because together they produced a manageable and substantively varied

linguistics corpus. This purposive scope was designed for a field-bounded pilot and does not represent Saudi higher education or Saudi linguistics as a whole.

Search Strategy and Retrieval

Records were retrieved through the OpenAlex API using institution-ID-based queries anchored to the OpenAlex institution identifiers for Institutions X-Z. Retrieval was completed in 2026-04 using a snapshot of the OpenAlex database represented in the project retrieval files. The institution IDs used for retrieval are recorded in the supplementary materials. Queries were restricted to the 2021-2026 publication window and to journal-article-equivalent records. No language restriction was applied at retrieval. Following the logic of PRISMA-style reporting [45], the flow was documented as follows: 360 raw retrieval hits, 281 records after first-pass deduplication, 217 records after linguistics-inclusion screening, and 209 final included papers after DOI- and exact-title-based deduplication. Supplementary File S1 lists the institution-level retrieval identifiers and the work-level OpenAlex identifiers used for the 209 included records; one included record remained unresolved at the work-ID enrichment stage and is left blank in that file.

Retrieval Logic

The retrieval strategy was deliberately broad at first pass. Institution-ID queries were followed by screening of titles, abstracts, source titles, and journal scope. A record was included only when language, linguistics, discourse, pragmatics, sociolinguistics, corpus linguistics, translation studies, or language teaching and learning was the central object of inquiry. Records were excluded when language appeared only as a measurement tool, participant characteristic, application setting, or incidental keyword in engineering, computing, health, or educational-technology research. Borderline records were retained only when the abstract and source scope together established a

substantive linguistics contribution. This rule was applied before exact-title and DOI deduplication; the screening criteria are reproduced in the supplementary material.

Linguistics Inclusion, Deduplication, and Corpus Construction

Candidate records were screened into include, exclude, and borderline categories using a field codebook. Records were retained only when language, linguistics, discourse, pragmatics, sociolinguistics, corpus linguistics, translation studies, or language teaching and learning formed the central object of study. After screening, exact-title and DOI-based deduplication was applied to remove duplicate records across overlapping query streams. This produced the final included corpus of 209 papers.

Metadata Enrichment and Structural Indicators

The included corpus was then enriched with work-level metadata from OpenAlex, including publication date, source title, open-access status, author count, first author, last author, corresponding author, institution diversity, and country diversity. From these fields, a set of structural indicators was derived, including repeated first-last pairs, repeated unordered coauthor pairs, repeated exact teams, recurrent last-author positions, and a composite recurrent-structure flag. These indicators were used to support staged screening and coding, not to prove misconduct.

Coding Logic

Manual AMD-informed coding was carried out on a subset of 86 papers selected progressively from the structurally enriched corpus. The goal of coding was not to estimate prevalence across the full 209-paper corpus, but to distinguish analytically unremarkable papers from papers that warranted closer validation. For multi-author papers, the coding scheme recorded manual authorship concern, pressure, opportunity, capability, normalisation, rationalisation evidence, and follow-up priority. Rationalisation was intentionally fixed as not inferable from metadata. For solo-author papers, the coding protocol treated the team-based AMD dimensions as structurally inapplicable. In the final coding set, 10 solo-author papers were explicitly coded as n/a (sole author) for authorship concern, while the remaining 23 solo-author papers were retained in a pressure-oriented contrast strand and received low concern labels because they were coded for productivity-related screening rather than team-based authorship interpretation. Methods, tables, and prose were reconciled to this explicit rule.

Selection of the Coded Subset

Of the 209 papers in the retrieved dataset, 86 received manual AMD-informed coding at the current stage of analysis. The selection was deliberately non-random and followed a mechanical priority

queue documented in the supplementary master workbook. The queue assigned each paper a Round-Score derived from its structural-metadata profile. Higher scores reflect higher Structural Indicator Counts (a count of structural triggers present, including non-zero Repeated Team Count, non-zero First-Last Pair Count, and high Last-Author or Max-Author Dataset Counts), and prioritise multi-author teams ahead of sole-author cases and primary AMD-search candidates ahead of secondary-search candidates.

Papers were assigned to one of three queue types: AMD-core multi-author cases (papers whose structural metadata met AMD-priority criteria), team-secondary multi-author cases (papers meeting team-based criteria but ranking below the AMD-core threshold), and solo-pressure cases (sole-author papers retained in a productivity-pressure contrast strand). At the current stage, all 52 AMD-core multi-author cases plus one team-secondary case plus 33 solo-pressure cases (n = 86) have been coded; 54 team-secondary and 69 solo-pressure cases remain in the queue for subsequent coding rounds.

The priority-queue logic is mechanical: a paper's queue position is computable from its structural-metadata profile without coder discretion. The selection is therefore transparent and reproducible. It is also non-random by design: the coded subset is structurally weighted toward higher-priority cases. This is a deliberate efficiency choice in which coder effort is allocated to the most analytically informative cases first; the trade-off is documented in the Limitations section. The coded subset is therefore not a probability sample of the 209-paper corpus; it is a structured priority subset whose composition is auditable from the supplementary materials. Round-by-round queue logs, per-paper Round-Scores, and queue-type assignments are provided in the supplementary master workbook.

Threshold Protocol

The v3 coding protocol operationalised each AMD dimension through explicit bands anchored to observed structural-metadata attributes. Pressure, Capability, Opportunity, and Normalisation were coded directly from documented metadata; Rationalisation was fixed as not inferable at this stage. Table 1 reproduces the operational bands. Each band is defined by an enumerated range of observed attributes, so that a coding decision reduces to reading the paper's attributes against the enumerated range. The v3 wording is stricter than the v2 wording used during the reliability exercise, but the thresholds and substantive meanings are unchanged; the full codebook, including boundary examples for each band, is deposited on Figshare (see Data Availability). The composite Authorship Concern and Follow-up Priority variables are derived from the four directly coded dimensions through the explicit rules described in the next subsection.

Table 1. AMD-dimension operational bands (v3 protocol).

Dimension	None evident	Weak	Moderate
Pressure	Institutional productivity indicator below field median; no productivity-cluster involvement	Institutional productivity indicator in the upper quartile, or one recurrent last-author position	Institutional productivity indicator in the top decile combined with a recurrent last-author position, or high recurrent output rate
Capability	No terminal-author concentration; no persistent supervisory-role indicator	One recurrent terminal-author position, or one supervisory-role indicator	Multiple recurrent terminal-author positions and a documented supervisory-role indicator
Opportunity	No structural enabler observed (isolated paper; no public-rule gap)	One structural enabler present (repeated coauthor pair, or under-specified public authorship rule)	Multiple concurrent structural enablers (repeated pair and public-rule gap)
Normalisation	No recurrent back-matter formula; no cluster-bound output stream	One recurrent feature (formula or cluster-bound stream)	Multiple recurrent features (formula and stream and team-signature persistence)
Rationalisation	Fixed as N/A: not inferable from publication metadata in this design	—	—

Note. The full v3 codebook, including boundary examples for each band, is deposited on Figshare (see Data Availability).

AMD Coding Pipeline and Priority Derivation

The AMD coding procedure has three sequential stages that transform observed metadata for each coded paper into a shortlist decision (Figure 2). This Follow-up Priority is distinct from the Round-Score priority described in the Selection of the Coded Subset subsection above: the Round-Score determined which of the 209 corpus papers received manual coding first, whereas the Follow-up Priority determined which of the 86 coded papers advanced to contextual validation.

The first stage codes each of the four directly coded dimensions independently, using the operational bands in Table 1. Each paper receives Pressure, Capability, Opportunity, and Normalisation codes; Rationalisation is fixed as not inferable at this stage. The second stage derives the composite Authorship Concern from Pressure, Capability, and Normalisation using a pre-specified aggregation rule: None = all three dimensions are none evident; Low = at least one Weak score and no Moderate score; Medium = one Moderate score; High = two or more Moderate scores. Opportunity is not included in this concern score; it is treated separately as a modifier in the third stage. The third stage derives Follow-up Priority by combining Authorship Concern with Opportunity: a paper is prioritised as High only when Concern is Medium or greater AND Opportunity is Weak or greater; papers with elevated Concern but no observed Opportunity receive Medium priority; papers with neither elevated Concern nor Opportunity receive Low priority. Given a paper's four dimensional codes, the resulting Concern and Priority are computable without further coder discretion. The eight cases that advanced to contextual validation were those whose Follow-up Priority was High under this derivation.

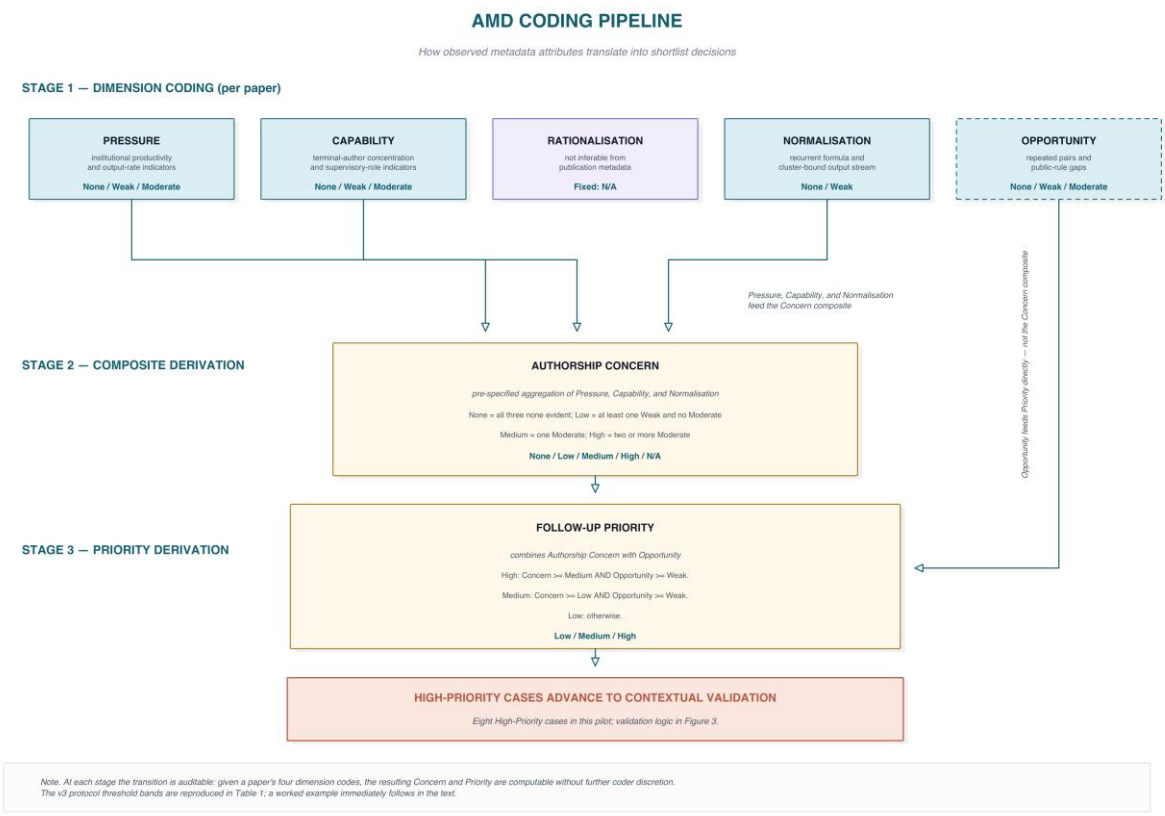


Figure 2. AMD coding pipeline. The figure shows the three sequential stages of the AMD-informed coding procedure. Stage 1 codes each of the four directly coded dimensions (Pressure, Capability, Opportunity, Normalisation) against the operational bands in Table 1; Rationalisation is fixed as not inferable at this stage. Stage 2 derives the composite Authorship Concern through the pre-specified aggregation rule across Pressure, Capability, and Normalisation. Stage 3 derives Follow-up Priority by combining Authorship Concern with Opportunity. High-priority cases advance to the contextual-validation phase shown in Figure 3. At each stage the transition is auditable: given a paper's dimensional codes, the resulting Concern and Priority are computable without further coder discretion.

Worked Example: Coding a Multi-Author Case

To make the AMD coding procedure concrete, this subsection walks through the coding decisions for one illustrative case drawn from the Institution X portion of the coded subset. The paper is presented under the same pseudonymised institutional and cluster labels used throughout this manuscript. It is a four-author linguistics paper published in 2024 in a subdomain where team-based collaboration is common.

Observed structural-metadata attributes for this paper included the following. First, the paper's last author held a recurrent terminal-author position across a substantial share of the Institution X

linguistics papers coded in this study, meeting the operational definition of a recurrent last-author position. Second, the first-last authorship pair recurred in three other papers within the coded corpus, meeting the operational definition of a repeated pair. Third, the paper's back-matter authorship-role text used a formula ("all authors contributed equally to conceptualisation, methodology, data curation, and writing") that appeared verbatim in several other Institution X papers in the coded corpus, meeting the operational definition of a recurrent back-matter formula. Fourth, the paper's funding statement was internally consistent and cited a specific institutional grant. Fifth, no public Institution X policy addresses author-order priority or contribution-attribution rules for linguistics work, meeting the operational definition of a public-rule gap.

Applying the operational bands in Table 1, Pressure was coded Moderate: the institutional productivity indicator was in the top decile and the last author's recurrent terminal-author position was documented. Capability was coded Moderate: the recurrent terminal-author position combined with a documented supervisory-role indicator crossed the Moderate band. Opportunity was coded Weak: a repeated first-last pair was present, but the paper stood alone in terms of concurrent public-rule gaps at the paper level. Normalisation was coded Moderate: the recurrent back-matter formula plus the cluster-bound output stream met the Moderate band. Rationalisation was fixed at N/A per protocol.

Applying the composite rules, Authorship Concern was High: two of the three team-based dimensions were Moderate, satisfying the "two or more Moderate" rule for High. Follow-up Priority was High: Concern was Medium or greater AND Opportunity was Weak or greater. The paper was therefore shortlisted for contextual validation.

At validation (Figure 3), the paper's Cluster A membership was confirmed through the author-network comparison; the recurrent back-matter formula and the absence of a public-policy explanation were confirmed in full-text review; the funding-statement irregularity was not observed for this case, so this evidence stream did not contribute to residual concern. The paper was classified as Convergent because structural recurrence, paper-level textual routinisation (the back-matter formula), and the public-policy explanatory gap converged; residual ethical concern was rated High.

This worked example illustrates that AMD-informed coding is neither a diagnostic algorithm for detecting misconduct nor a subjective judgement about specific authors: it is a structured mapping from observed metadata attributes onto operational bands, then onto composite classifications through explicit derivation rules. The resulting classification identifies where the metadata plus policy environment cannot together explain the observed authorship configuration; it does not identify authorship misappropriation.

Use of AI Tools in Manuscript Preparation

During manuscript preparation, OpenAI's ChatGPT was used only for language editing and formatting support. It was not used to generate or analyse the dataset, assign AMD codes, determine validation classifications, or produce substantive interpretations. The author reviewed and remains responsible for all content, claims, references, and revisions in this manuscript.

Coder, Audit Trail, and Reliability

All AMD-informed manual coding was conducted by a single primary coder under a pre-specified threshold protocol. The v2 protocol governed the initial coding and reliability exercise. Following that exercise, boundary wording was rewritten as v3 by enumerating the integers in each band; the

thresholds and their substantive meanings were unchanged. Pressure, Capability, Normalisation, and the Opportunity composite were derived from documented structural metadata, while Rationalisation was fixed as not inferable from metadata.

Inter-coder reliability was assessed on 16 multi-author papers. The papers were selected with Python's random.sample function using seed 42 from the 33 multi-author cases completed when the reliability exercise began. They therefore constitute 48.5% of that then-completed set and 30.2% of the final 53 multi-author coded papers, but they were not a random sample of the final set because 20 cases were coded later. An applied linguist at a Saudi public university, unaffiliated with the project or corpus institutions, independently applied the threshold protocol without access to the first coder's decisions or the manuscript.

Agreement was perfect for Pressure and Opportunity (Cohen's kappa = 1.000 for each), substantial for Normalisation (kappa = 0.619), moderate for Capability (kappa = 0.602), and moderate for the derived Authorship Concern and Follow-up Priority composites (kappa = 0.474 for each), using the descriptive bands of Landis and Koch [46]. The lower composite agreement propagated from upstream Capability and Normalisation disagreements because Concern and Priority were derived through the pre-specified aggregation rule described above; it nevertheless indicates uncertainty in the resulting classifications and is treated as a limitation.

The six directly coded disagreements occurred at threshold boundaries: four in Capability and two in Normalisation. The protocol was subsequently rewritten to enumerate every integer explicitly. This clarification does not retroactively change the reliability estimate under the wording used in the exercise. The second coder's location within the Saudi higher-education system is an additional independence caveat.

Validation Procedure

The contextual-validation phase focused on eight cases drawn from the integrated follow-up matrix. Selection considered the strength of the paper-level signal, whether publicly available institutional rules offered a plausible explanation, and whether the paper formed part of a repeated pair or broader collaboration cluster. Structured case memos and cross-case comparison, informed by Braun and Clarke [47], examined recurrent authorship language, funding consistency, author order, governance context, and alternative ordinary explanations. Because selection deliberately favoured strong signals, the phase was designed to interrogate and contextualise high-priority cases rather than provide an independent validation sample.

Validation drew on three substantive evidence streams. First, full-text review examined author-order structure, corresponding-author position, contribution architecture, ethics-statement phrasing, and the internal consistency of funding-related sections. Second, policy comparison mapped each case against what the two focal institutional contexts publicly formalised in areas such as author-position incentives, corresponding-author practice, co-author approval, affiliation requirements, and funding-statement wording. Third, author-network comparison assessed whether a focal paper was isolated, part of a repeated pair, or embedded in a broader cluster. The resulting integrated matrix recorded full-text signal strength, policy-fit strength, network-fit strength, residual concern after integration, and the next-best evidence source.

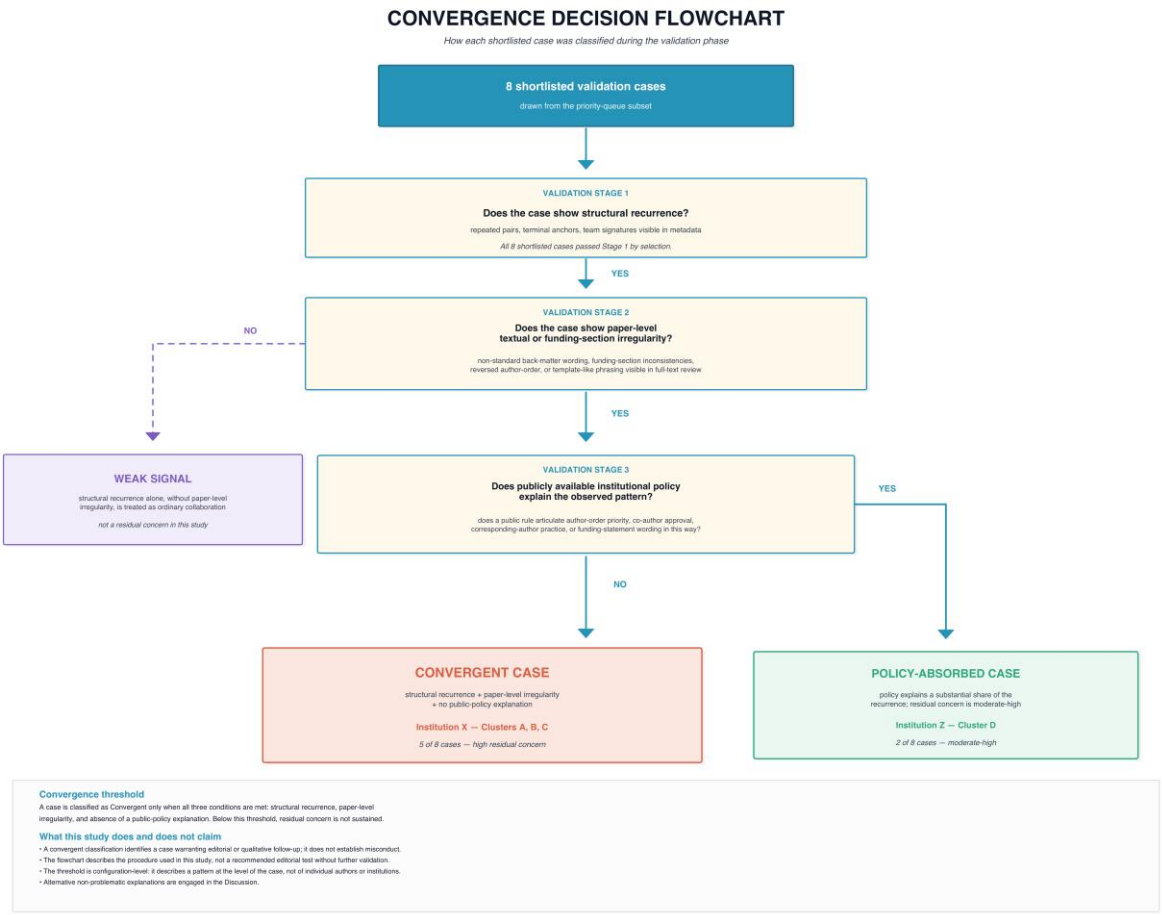


Figure 3. Convergence decision flowchart used in the validation phase. Each shortlisted case was classified by sequential consideration of three conditions: structural recurrence (visible in metadata), paper-level textual or funding irregularity (visible in full-text review), and the absence of a public-policy explanation. The procedure is a stylised representation of an interpretive analytical process; it is not a deterministic algorithm and is not proposed as a routine editorial test without further validation.

Trustworthiness, Reflexivity, and Limits

The staged design sought interpretive trustworthiness through progressive contextualisation. Metadata screening identified recurrent structures; threshold coding made the screening logic auditable; and full-text, policy, and network evidence tested whether high-priority signals weakened, persisted, or acquired plausible alternative explanations. A retained residual-concern classification means only that the available evidence did not adequately explain the configuration. It does not adjudicate intent or misconduct.

Anonymisation and Research Ethics

This study uses publicly available bibliographic and full-text data and does not involve human subjects in its current phase. A planned subsequent interview phase will be subject to formal institutional review-board approval before any data collection begins.

Because the topic is misconduct-adjacent, the manuscript reports validation findings through pseudonymous institution and cluster labels and omits author names, paper titles, and journal names. Public supplementary files contain aggregated or pseudonymous analytical outputs. Identifiable source mappings are not required to understand the reported findings and should be supplied only to the editor, confidentially, if requested and ethically appropriate.

The study makes no claim of authorship misappropriation against any individual or institution. Recurrent metadata and full-text patterns are reported as analytically meaningful screening signals that warrant further qualitative investigation, not as evidence of misconduct.

Results

The results first describe the retrieved and coded corpus and then report the cross-case findings from the contextual-validation stage.

Metadata Screening and Manual Coding

The pilot dataset comprised 209 Saudi-affiliated linguistics papers after title- and DOI-based deduplication. The corpus was distributed across the three pilot institutions labeled Institution X, Institution Y, and Institution Z.

Across the 208 papers with resolved authorship metadata, the mean author count was 2.04 and the median was 2. Publication years were distributed as follows: 2021 (52), 2022 (49), 2023 (41), 2024 (32), 2025 (30), and January-April 2026 (5). The corpus spanned 117 source titles. Of the 208

resolved records, 107 (51.4%) were multi-author and 101 (48.6%) were sole-author. Open-access status was available for all 209 included records; 183 (87.6%) were marked open access and 26 (12.4%) were not.

Table 2 summarizes the institutional distribution of the coded subset and the high-priority follow-up cases.

Table 2. Institutional distribution of the coded subset

Institution	Coded papers	Multi-author coded	Solo-pressure coded	Medium concern	High follow-up priority
Institution X	31	20	11	11	11
Institution Y	29	15	14	4	4
Institution Z	26	18	8	9	9
Total coded subset	86	53	33	24	24

Manual AMD-informed coding was completed for 86 papers. This left 123 included papers uncoded at the current stage and reflects the pilot's staged design rather than an abandoned sampling logic. Of the coded set, 53 papers were multi-author cases and 33 were solo-pressure cases. Within the coded subset, 52 papers were classified as AMD-core and one as team-secondary, while the remaining cases belonged to the solo-pressure contrast strand.

The coding phase was not intended to produce a prevalence estimate of problematic authorship across the full pilot corpus. Instead, it was designed to identify which recurrent multi-author patterns remained analytically important once the field had been narrowed to a more information-rich subset. Across all coded papers, manual authorship concern was distributed as 15 none, 37 low, 24 medium, and 10 n/a (sole author). No coded paper was classified as high concern at the metadata stage alone.

The solo-author cases reinforce this point by functioning primarily as a contrastive strand. Where authorship was genuinely solo, the team-based AMD dimensions could not be applied in the same way. In the final coding scheme, 10 solo-author papers were explicitly coded as n/a (sole author) for authorship concern, while 23 solo-author papers remained in the pressure-oriented contrast strand and were assigned low concern labels because they were retained for productivity-pattern screening rather than for direct team-based authorship interpretation.

Within the 53 multi-author papers, pressure was coded predominantly as moderate (36 cases), followed by weak (14) and none evident (3). Opportunity was coded as none evident in 21 cases, weak in 18, and moderate in 14. Capability showed the same upper-end pattern: 18 none evident, 21 weak, and 14 moderate. Normalisation was the least visible AMD dimension at the metadata stage, with 45 none-evident and 8 weak classifications. Rationalisation was not inferable from metadata in any coded case.

Figure 4 consolidates these distributions across the six AMD-related screening variables and makes the distributional shape immediately visible.

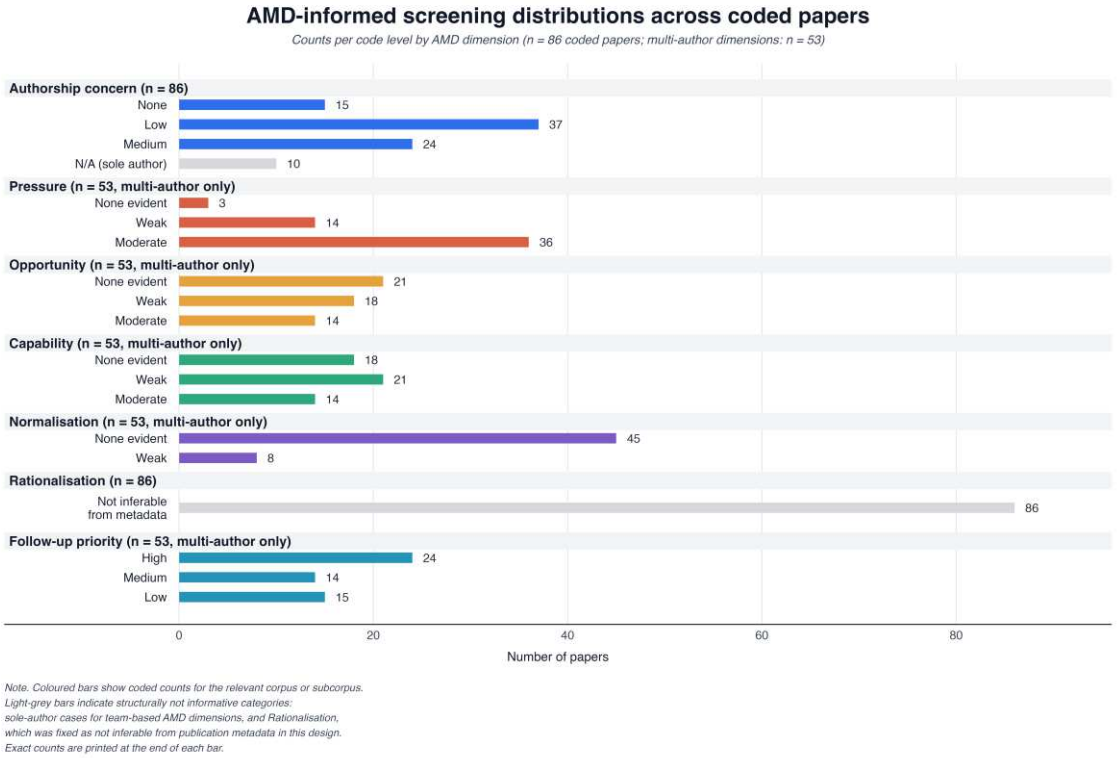


Figure 4. AMD-informed screening distributions across coded papers. Counts per code level for six AMD-related screening variables. The full coded subset is n = 86; for team-based AMD dimensions (Pressure, Capability, Opportunity, Normalisation, Follow-up priority) the analytical denominator is the multi-author subset (n = 53), with sole-author cases excluded as structurally not informative for those dimensions and shown as a light-grey N/A bar where relevant. Rationalisation is not inferable from publication metadata in this design and is shown as a fixed N/A row. Exact counts are printed at the end of each bar.

Taken together, the quantitative findings show that the metadata and coding stages produced a mixed field of none, low, and medium concern cases rather than a sharp split between ordinary and evidently problematic papers. Medium-concern multi-author cases clustered around recurrent coauthorship structures, repeated pairings, or stable last-author positions, but did not supply direct evidence of intent. The quantitative stage therefore functioned as a screening and triage mechanism: it identified where recurrent structure appeared analytically important and where metadata alone had reached its interpretive limit.

Validation-Phase Findings

Table 3 presents the integrated validation summary for the eight shortlisted cases.

458 **Table 3.** Integrated validation summary for shortlisted cases

Case	Institution	Cluster label	Full-text effect	Residual concern	Interview priority	Main residual issue
1	Institution X	Cluster A	Substantially stronger than metadata	High	High	Non-standard wording in authorship-related back matter
2	Institution X	Cluster A	Substantially stronger than metadata	High	High	Repeated non-standard wording in authorship-related back matter
3	Institution X	Cluster B	Substantially stronger than metadata	High	High	Internal inconsistency between funding-related sections requiring author clarification
4	Institution X	Cluster B	Stronger than metadata	Moderate-high	Medium-high	Small-team overview-style paper with limited role differentiation
5	Institution X	Cluster C	Stronger than metadata	High	High	Reversed first/last authorship order across a related pair of papers
6	Institution X	Cluster C	Stronger than metadata	High	High	Repeated author-order irregularity across a related pair of papers

Case	Institution	Cluster label	Full-text effect	Residual concern	Interview priority	Main residual issue
7	Institution Z	Cluster D	Stronger than metadata	Moderate-high	High	Broad equal-role contribution architecture across paired empirical papers
8	Institution Z	Cluster D	Stronger than metadata	Moderate-high	High	Repeated contribution formula across distinct study designs

Cross-Case Validation Results

The contextual-validation stage differentiated the eight high-priority cases rather than treating all recurrence as equivalent. Five cases retained high residual concern after the three evidence streams were integrated, while three retained moderate-high concern. These classifications describe unresolved explanatory difficulty within a deliberately enriched set of cases; they are not prevalence estimates and do not indicate that misconduct occurred.

At Institution X, Cluster A combined recurrent terminal-author positioning with unusual wording in authorship-related back matter across two papers. Cluster B combined a shifting small collaboration network with inconsistencies between funding-related sections in two papers. Cluster C comprised two related outputs in which first- and last-author order was reversed. These observations strengthened the case for author or editorial clarification, but each also retained a plausible non-problematic explanation, including journal templating, editorial error, co-principal-investigator rotation, disciplinary convention, or genuine changes in contribution. The available documents could not adjudicate among those explanations.

Cluster D at Institution Z showed recurrent contribution wording and broad equal-role statements across two empirical papers. Publicly available institutional rules governing author-position priority, co-author approval, corresponding-author practice, and funding wording offered a more plausible administrative context for that recurrence than was available for the Institution X cases. The policy evidence did not establish that the authors followed those rules, nor that the rules caused the observed patterns; it reduced, but did not eliminate, the explanatory gap.

Across cases, recurrence alone was insufficient to sustain concern. Residual concern increased only when recurrence converged with a second paper-level signal and when the public policy layer did not offer a plausible account of the practice. Because the cases were selected for strong initial signals, this pattern should be interpreted as a within-case result requiring comparison with lower-signal cases in future work.

Discussion

The findings of this pilot can be read at three levels: the empirical level (what the validation cases show), the theoretical level (how the AMD dimensions operate together in the surviving configurations), and the ethical level (what the configurations imply for authorship integrity under conditions of visible institutional governance). The three levels are connected, and the analytical contribution of the pilot lies in their integration.

Interpretation of Cross-Case Patterns

The cross-case results suggest that explanatory convergence, rather than any single AMD indicator, is the manuscript's central analytical object. The following model interprets how the dimensions could interact while distinguishing that theoretical interpretation from what the documentary evidence directly establishes.

AMD as a Dynamic Explanatory Model

Across the configurations that retained concern, the AMD dimensions offer a relational interpretation rather than independent causal findings. Productivity-and-evaluation pressure may increase the operational value of authorship; under-specified attribution rules may create opportunity structures; positional authority may shape how credit is allocated; and repeated collaboration may make an allocation routine appear ordinary. Governance ambiguity is the context in which those processes could interact. The present evidence is consistent with that model but does not demonstrate any actor's pressure, authority, or rationalisation.

Cluster A is consistent with a configuration in which recurrent terminal-author positioning coincides with unusual back-matter wording that public policy does not address. Cluster B locates the explanatory gap in funding-attribution consistency, while Cluster C locates it in reciprocal author-order changes. Cluster D provides a contrast because Institution Z's public rules offer a plausible administrative account of part of the recurrence. These interpretations remain provisional because journal production, contribution changes, co-investigator rotation, and disciplinary practice could also produce the observed patterns.

Read together, the strongest cases are better treated as unresolved configurations than as signals of problematic behaviour by identifiable authors or institutions. They show where direct clarification would be ethically and analytically useful; they do not establish that publicly visible governance failed in practice or that any individual benefited improperly.

Ethical Contribution and Implications for Institutional Responsibility

The pilot's ethical contribution is to propose a configuration-level threshold for further ethical inquiry. That threshold is the convergence of structural recurrence, a second paper-level inconsistency or irregularity, and the absence of a plausible public-policy explanation. Below the

threshold, recurrence remains readily compatible with ordinary collaboration. Above it, the appropriate response is clarification or qualitative follow-up rather than a presumption of misconduct.

The institutional implications are practical but provisional. Integrity policies may be strengthened by articulating author-order expectations, contribution-attribution procedures, corresponding-author responsibilities, and routes for resolving authorship disagreement. Journals may use the configuration vocabulary as an editorial triage aid, while institutions may use it to identify areas in which public guidance is silent. Neither use should convert screening indicators into misconduct determinations.

The next analytical step is direct qualitative inquiry with researchers and relevant institutional actors. Interviews could examine how authors understand contribution, authority, policy, and authorship order, and could activate the Rationalisation dimension that publication records cannot access. Such work would require prior ethics approval and safeguards against reputational harm.

In the Saudi higher-education context, this configuration is especially important because research productivity, international visibility, and ranking-oriented institutional performance have become increasingly salient markers of academic value [30, 31, 38–40]. The findings do not show that these pressures caused any specific authorship arrangement, but they suggest that productivity-oriented environments may increase the ethical significance of otherwise ordinary authorship routines when public contribution-attribution rules remain under-specified. In this sense, the Saudi case is not treated as exceptional; rather, it illustrates a broader integrity problem faced by rapidly internationalising higher-education systems in which visible governance expands faster than the local authorship practices it is expected to regulate.

Limitations

First, the AMD variables are metadata-based proxies rather than direct observations of pressure, opportunity, capability, normalisation, or rationalisation. High productivity does not itself demonstrate pressure; recurrent terminal authorship does not establish positional authority; and repeated teams do not establish normalisation. The framework therefore functions as a sensitising and screening architecture. Causal claims about the operation of its dimensions require interviews, contribution records, or other direct evidence.

Second, one researcher coded the full set. A second coder independently assessed 16 papers selected with a fixed random seed from the 33 multi-author cases completed at that stage. The sample represents 48.5% of the then-completed cases and 30.2% of the final 53 multi-author cases, but it was not randomly drawn from the final set because the remaining 20 cases had not yet been coded. Agreement was perfect for Pressure and Opportunity, moderate-to-substantial for Capability and Normalisation, and moderate for the derived composites. The reliability exercise therefore supports the reproducibility of several threshold rules but does not remove single-coder dependence from the full dataset.

Third, the eight contextual-validation cases were intentionally selected for the strongest combined signals. Their residual concern is consequently partly built into the design, and the validation stage should be understood as contextual interrogation of high-priority cases rather than an independent test of the screening results. The absence of randomly sampled or matched lower-signal cases limits claims about the distinctiveness of the observed configurations.

Fourth, the coded subset comprises 86 of 209 retrieved papers and was selected through a mechanical but non-probability priority queue. Institution-level contrasts, including the Institution

X and Institution Z comparison, are configuration-level findings from a structurally enriched subset rather than population-level institutional comparisons. Stratified random sampling or exhaustive coding of all multi-author papers would be required to estimate institutional distributions.

Fifth, the study reaches structural recurrence but not authorship intent. Publication metadata identifies repeated pairs, stable terminal-author positions, and team signatures, but it does not reveal how authors negotiated credit. Rationalisation remains inaccessible without direct fieldwork, and several apparently unusual patterns may have ordinary explanations grounded in contribution changes, journal production, disciplinary norms, or supervisory arrangements.

Sixth, the institutional-policy layer relies on publicly available materials. Internal workflows, unpublished guidance, and informal attribution conventions are not represented. Public policy can therefore offer a plausible explanatory context but cannot establish what occurred in any specific case. To reduce deductive-identification risk, public supplementary materials report only pseudonymous or aggregated analytical outputs; identifiable source mappings are reserved for confidential editorial review where necessary.

Conclusion

This study contributes a bounded configuration-level approach to examining recurrent authorship patterns under visible integrity governance. In this three-institution pilot, residual concern was sustained not by recurrence alone but when recurrence converged with a paper-level inconsistency and an unresolved policy-explanatory gap. The finding identifies circumstances in which editorial or qualitative clarification is warranted.

Contextual validation reduced the risk of treating recurrent bylines as self-explanatory. Five cases retained high residual concern and three retained moderate-high concern, but every configuration also admitted ordinary alternative explanations that the available records could not exclude. Institution Z's public rules supplied a stronger administrative context for its repeated pair than was available for the Institution X clusters.

The study's contribution is methodological as well as substantive. It shows how metadata screening can be made more transparent through pre-specified thresholds, reliability assessment, and progressive contextualisation, while also demonstrating the limits of the AMD dimensions in documentary data. Pressure, Capability, and Normalisation remain proxy-based interpretations, and Rationalisation remains inaccessible without direct fieldwork.

Accordingly, the strongest next step is targeted qualitative follow-up combined with exhaustive or probability-based coding of the remaining corpus. Recurrent structures in this pilot should be treated neither as innocuous by default nor as evidence of misconduct by default, but as patterns whose ethical meaning depends on progressively stronger evidence.

List of abbreviations

AMD: Authorship Misappropriation Diamond

API: Application Programming Interface

COPE: Committee on Publication Ethics

DOI: Digital Object Identifier

IRB: Institutional Review Board

LLM: Large Language Model

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

RDIA: Research, Development and Innovation Authority (Saudi Arabia)

SRQR: Standards for Reporting Qualitative Research

Declarations

Ethics approval and consent to participate

Not applicable. This study analysed publicly available bibliographic metadata and full-text content of already-published academic papers retrieved from the OpenAlex database. No human participants were directly involved at this stage; no biological samples, interviews, or interventional procedures were used. Findings are reported in pseudonymous form, with institutional identifiers replaced by anonymous labels (Institution X, Y, Z) and cluster identifiers replaced by anonymous labels (Cluster A, B, C, D). The study does not adjudicate misconduct and does not identify individual authors. Ethical approval for the planned follow-up interview phase will be obtained separately from the relevant institutional review board prior to interview-stage data collection.

Consent for publication

Not applicable. No identifiable individual data are presented in this manuscript; institutional and cluster identifiers are pseudonymised throughout.

Availability of data and materials

The datasets supporting the conclusions of this article are available in the Figshare repository under the persistent identifier https://doi.org/10.6084/m9.figshare.32112022 [48]. The deposit contains aggregated coding distributions, the integrated validation matrix in pseudonymous form, the

institution-level retrieval identifiers, the OpenAlex work identifiers for the 209 included records, the project Codebook (v3), the paired inter-coder reliability ratings, and the reproducible Cohen's kappa computation script. Individual paper-level analyst notes that could enable de-anonymisation of cluster labels are not included in this public deposit.

Competing interests

The author declares that they have no competing interests with respect to the research, authorship, or publication of this article.

Funding

The author received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors for this research.

Authors' contributions

HA is the sole author. HA conceived and designed the study, performed the manual AMD-informed coding, conducted the contextual validation, wrote the manuscript, and takes full responsibility for its content.

Acknowledgements

The author thanks the second coder for the inter-coder reliability check, the colleagues who reviewed earlier drafts of this manuscript, and the participating institutional communities whose published outputs form the corpus of this study.

Use of AI / Generative AI

The author used OpenAI's ChatGPT solely for language editing during manuscript preparation. The tool was not used to generate or analyse the empirical dataset, to make AMD-aligned coding decisions, to determine the contextual validation classifications, or to draft substantive content. All

claims, interpretations, references, and revisions were produced and reviewed by the author, who takes full responsibility for the manuscript.

References

1. Wuchty S, Jones BF, Uzzi B. The increasing dominance of teams in production of knowledge. Science. 2007;316(5827):1036–9. https://doi.org/10.1126/science.1136099.

2. Savchenko E, Rosenfeld A. Authorship conflicts in academia: an international cross-discipline survey. Scientometrics. 2024;129:2101–21. https://doi.org/10.1007/s11192-024-04972-x.

3. Wislar JS, Flanagin A, Fontanarosa PB, DeAngelis CD. Honorary and ghost authorship in high impact biomedical journals: a cross-sectional survey. BMJ. 2011;343:d6128. https://doi.org/10.1136/bmj.d6128.

4. Jakab M, Kittl E, Kiesslich T. How many authors are (too) many? A retrospective, descriptive analysis of authorship in biomedical publications. Scientometrics. 2024;129:1299–328. https://doi.org/10.1007/s11192-024-04928-1.

5. Sundling P. Author contributions and allocation of authorship credit: testing the validity of different counting methods in the field of chemical biology. Scientometrics. 2023;128:2737–62. https://doi.org/10.1007/s11192-023-04680-y.

6. Oyenuga MO, Apata SB, Oladele TO, et al. You Put my name, I Put your name: exploring unethical practices in academic publishing using the authorship misappropriation Diamond framework. Ethics Behav. 2026. Epub ahead of print. https://doi.org/10.1080/10508422.2026.2661701.

7. Wolfe DT, Hermanson DR. The Fraud Diamond: considering the four elements of fraud. CPA J. 2004;74(12):38–42.

8. Merton RK. Priorities in scientific discovery: a chapter in the sociology of science. Am Sociol Rev. 1957;22(6):635–59. https://doi.org/10.2307/2089193.

9. Merton RK. The Matthew Effect in science. Science. 1968;159(3810):56–63. https://doi.org/10.1126/science.159.3810.56.

10. Merton RK. The Sociology of Science: Theoretical and Empirical Investigations. Chicago, IL: University of Chicago Press; 1973.

11. Biagioli M. The instability of authorship: credit and responsibility in contemporary biomedicine. FASEB J. 1998;12(1):3–16. https://doi.org/10.1096/fasebj.12.1.3.

12. Biagioli M, Galison P, editors. Scientific Authorship: Credit and Intellectual Property in Science. New York: Routledge; 2003.

13. Bourdieu P. Science of Science and Reflexivity. Chicago, IL: University of Chicago Press; 2004.

14. Fochler M. Variants of epistemic capitalism: knowledge production and the accumulation of worth in commercial biotechnology and the academic life sciences. Sci Technol Human Values. 2016;41(5):922–48. https://doi.org/10.1177/0162243916652224.

15. Fochler M, Felt U, Müller R. Unsustainable growth, hyper-competition and worth in life science research: narrowing evaluative repertoires in doctoral and postdoctoral scientists' work and lives. Minerva. 2016;54:175–200. https://doi.org/10.1007/s11024-016-9292-y.

16. Decius J, Schilbach M. Fair credit? The impact of shared first authorship on academic career evaluation. Scientometrics. 2025;130:1731–50. https://doi.org/10.1007/s11192-025-05262-w.

17. Donner P, Korytkowski P. Alphabetical author order and co-author contributions in mathematics. Scientometrics. 2025;130:4141–56. https://doi.org/10.1007/s11192-025-05369-0.

18. Espeland WN, Stevens ML. Commensuration as a social process. Annu Rev Sociol. 1998;24:313–43. https://doi.org/10.1146/annurev.soc.24.1.313.

19. Power M. The Audit Society: Rituals of Verification. Oxford: Oxford University Press; 1997.

20. Strathern M, editor. Audit Cultures: Anthropological Studies in Accountability, Ethics and the Academy. London: Routledge; 2000.

21. Committee on Publication Ethics. Authorship, contributorship, who is doing what and what do we need? 2013. https://publicationethics.org/files/u661/Forum_discussion_summary_on_Authorship.pdf. Accessed 27 June 2026.

22. Contributor Roles Taxonomy. CRediT. 2022. https://credit.niso.org/. Accessed 27 June 2026.

23. International Committee of Medical Journal Editors. Recommendations for the conduct, reporting, editing and publication of scholarly work in medical journals. 2024. https://www.icmje.org/recommendations/. Accessed 27 June 2026.

24. McNutt MK, Bradford M, Drazen JM, et al. Transparency in authors' contributions and responsibilities to promote integrity in scientific publication. Proc Natl Acad Sci U S A. 2018;115(11):2557–60. https://doi.org/10.1073/pnas.1715374115.

25. National Information Standards Organization. ANSI/NISO Z39.104-2022, CRediT, Contributor Roles Taxonomy. Baltimore, MD: NISO; 2022. https://doi.org/10.3789/ansi.niso.z39.104-2022.

26. Hicks D, Wouters P, Waltman L, de Rijcke S, Rafols I. The Leiden Manifesto for research metrics. Nature. 2015;520(7548):429–31. https://doi.org/10.1038/520429a.

27. Research, Development and Innovation Authority. National policy on scientific integrity and research ethics. 2024. https://istitlaa.ncc.gov.sa/en/education/rdia/nationalpolicyforscientificintegrity/Pages/default.aspx. Accessed 27 June 2026.

28. AlFattani A, AlBedah N, AlShahrani A, et al. Institutional review boards in Saudi Arabia: the first survey-based report on their functions and operations. BMC Med Ethics. 2023;24:50. https://doi.org/10.1186/s12910-023-00928-7.

29. AlFattani A, AlShahrani A, AlBedah N, et al. Exploring researchers' perspectives on institutional review boards functions in Saudi Arabia: a survey utilizing the IRB-RAT tool. BMC Med Ethics. 2025;26:26. https://doi.org/10.1186/s12910-025-01179-4.

30. Kingdom of Saudi Arabia. Saudi Vision 2030. 2016. https://www.vision2030.gov.sa/en/. Accessed 27 June 2026.

31. Council of Economic and Development Affairs. National Transformation Program. 2021. https://www.vision2030.gov.sa/en/vision-2030/vrps/ntp/. Accessed 27 June 2026.

729 32. Smith L, Abouammoh A, editors. Higher Education in Saudi Arabia: Achievements,
730 Challenges and Opportunities. Dordrecht: Springer; 2013. https://doi.org/10.1007/978-94-007-
731 6321-0.

732 33. Onsman A. It is better to light a candle than to ban the darkness: government led academic
733 development in Saudi Arabian universities. High Educ. 2011;62(4):519–32.
734 https://doi.org/10.1007/s10734-010-9402-y.

735 34. Allmnakrah A, Evers C. The need for a fundamental shift in the Saudi education system:
736 implementing the Saudi Arabian economic vision 2030. Res Educ. 2020;106(1):22–40.
737 https://doi.org/10.1177/0034523719851534.

738 35. Fanelli D. Do pressures to publish increase scientists' bias? An empirical support from US
739 states data. PLoS One. 2010;5(4):e10271. https://doi.org/10.1371/journal.pone.0010271.

740 36. Müller R. Racing for what? Anticipation and acceleration in the work and career practices of
741 academic life science postdocs. Forum Qual Sozialforschung / Forum Qual Soc Res.
742 2014;15(3):Art 15. https://doi.org/10.17169/fqs-15.3.2245.

743 37. Ozturk O, Taskin Z. How metric-based performance evaluation systems fuel the growth of
744 questionable publications? Scientometrics. 2024;129:2729–48. https://doi.org/10.1007/s11192-
745 024-04991-8.

746 38. Hazelkorn E. Rankings and the Reshaping of Higher Education: The Battle for World-Class
747 Excellence. 2nd ed. Basingstoke: Palgrave Macmillan; 2015.
748 https://doi.org/10.1057/9781137446671.

749 39. Marginson S. University rankings and social science. Eur J Educ. 2014;49(1):45–59.
750 https://doi.org/10.1111/ejed.12061.

751 40. Espeland WN, Sauder M. Rankings and reactivity: how public measures recreate social worlds.
752 Am J Sociol. 2007;113(1):1–40. https://doi.org/10.1086/517897.

753 41. Hanafi S. University systems in the Arab East: publish globally and perish locally vs publish
754 locally and perish globally. Curr Sociol. 2011;59(3):291–309.
755 https://doi.org/10.1177/0011392111400782.

756 42. Hanafi S, Arvanitis R. The marginalization of the Arab language in social science: structural
757 constraints and dependency by choice. Curr Sociol. 2014;62(5):723–42.
758 https://doi.org/10.1177/0011392114523665.

759 43. Hanafi S, Arvanitis R. Knowledge Production in the Arab World: The Impossible Promise.
760 Abingdon: Routledge; 2016.

761 44. Aldughmi M, Qutaishat D, Karasneh R. Knowledge and perceptions of honorary authorship
762 among health care researchers: online cross-sectional survey data from the Middle East. Sci Eng
763 Ethics. 2021;27(3):39. https://doi.org/10.1007/s11948-021-00317-6.

764 45. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline
765 for reporting systematic reviews. BMJ. 2021;372:n71. https://doi.org/10.1136/bmj.n71.

766 46. Landis JR, Koch GG. The measurement of observer agreement for categorical data.
767 Biometrics. 1977;33(1):159–74. https://doi.org/10.2307/2529310.

768 47. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol. 2006;3(2):77–
769 101. https://doi.org/10.1191/1478088706qp063oa.

48. Asiri H. Supplementary data for “Visible Governance, Hidden Routines: A Three-Institution Pilot of Authorship Ethics in Saudi-Affiliated Linguistics” [dataset]. figshare. 2026. https://doi.org/10.6084/m9.figshare.32112022.