Statistical Uncertainty under Extreme Class Imbalance: An Auditable Applied-Statistics Study of Accreditation in Ecuadorian Higher Education

Santiago Solís[1], Daniela Salazar[2]
ORCID: 139-04-139-04
[1] Instituto Superior Tecnológico Nelson Torres, Cayambe, Ecuador
sys.solis@gmail.com ORCID: 139-04-139-04
[2] Departamento de Ciencias Exactas, Universidad de las Fuerzas Armadas ESPE, Sangolquí 171103, Ecuador
dmsalazar10@espe.edu.ec

HOW TO CITE:
Santiago Solís, Daniela Salazar (2026). Statistical Uncertainty under Extreme Class Imbalance: An Auditable Applied-Statistics Study of Accreditation in Ecuadorian Higher Education International Journal of Special Education, 41(21s), 139 - 163

COPYRIGHT STATEMENT:
Copyright: © 2026 Authors. Open access publication under the terms and conditions of the Creative CommonsAttribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

ABSTRACT:
This study examines how extreme class imbalance shapes the interpretation of administrative indicators in higher education quality assurance. The analysis linked 2018 records from 55 Ecuadorian universities and polytechnic schools with document-verified outcomes from the 2019 institutional evaluation process, formally resolved in 2020. The cohort included 52 accredited institutions and three historical non-accreditation outcomes. The analytical strategy combined documentary verification, data-quality checks, medians, median differences, Cliff's delta, and leave-one-minority-case sensitivity analysis; a majority-class reference and original-direction within-sample ROC AUC were retained only as secondary diagnostics. All 55 outcomes and eight exclusion decisions were confirmed, and seven flagged values were verified against the original sources without correction. The majority-class rule reached 94.55% accuracy but yielded 0% sensitivity for historical non-accreditation and 50% balanced accuracy. Effect estimates also changed substantially after removing individual minority observations, revealing strong case-level dependence. The proposed workflow supports institutional self-evaluation, evidence review, and responsible data governance by identifying findings that require documentary, contextual, or qualitative follow-up. Given the scarcity of independent minority outcomes, the results do not justify stable prediction, institutional ranking, operational thresholds, or automated regulatory classification.

Keywords: applied statistics; extreme class imbalance; rare institutional outcomes; sensitivity analysis; administrative data; institutional analytics; higher education quality assurance

1. Introduction

Institutional accreditation is a core component of external quality assurance in higher education. Its role extends beyond certification: it also contributes to public accountability, organisational learning, continuous improvement, and confidence in educational provision. Because university quality encompasses excellence, consistency, fitness for purpose, value added, and transformation, accreditation status cannot be treated as a complete or continuous measure of quality (Harvey & Green, 1993; Martin & Stella, 2007). Nor can it be reduced to a small group of administrative indicators.

Quality-assurance systems generally bring together internal review, external evaluation, documentary evidence, quantitative information, and peer judgement. International frameworks place particular emphasis on transparent criteria, contextual interpretation, and improvement in teaching and learning (European Association for Quality Assurance in Higher Education et al., 2015; International Network for Quality Assurance Agencies in Higher Education, 2016). Administrative indicators can inform institutional decisions, but their meaning depends on the quality of the source data, the operational definitions applied, comparability across institutions, and the setting in which they are interpreted (Beerkens, 2018). Their most defensible role is therefore to support auditable educational review, not to replace comprehensive academic evaluation.

1.1. Higher Education Quality Assurance in Ecuador

External quality assurance in Ecuador is overseen by the Council for Quality Assurance in Higher Education (Consejo de Aseguramiento de la Calidad de la Educación Superior, CACES). The evaluation models implemented by CACES integrate predefined standards with institutional self-assessment, external peer reviews, and expert judgment (Council for Quality Assurance in Higher Education [CACES], 2019, 2023). The present analysis focuses specifically on the evaluations conducted under the 2019 framework, which reached formal resolution in 2020. References to the newer 2023 model serve primarily to place this historical process within the broader context of the country's evolving national policies. Rather than reducing accreditation to a single numerical threshold, the final decision relied on a comprehensive appraisal that weighed qualitative details, institutional context, and documentary proof.

Across Latin America, quality-assurance systems have developed through different national trajectories, but they commonly combine institutional self-evaluation, documentary evidence, quantitative information, external review, and academic judgement. Regional analyses also show that accreditation has moved beyond formal compliance toward institutional learning, accountability, and continuous improvement, although tensions remain between regulatory demands and institutional autonomy (Fernández Lamarra, 2003; Lemaitre, 2017; Rifo & Durán, 2023; Salto, 2018).

Even with this holistic approach, the administrative databases managed by SENESCYT/SIIES remain highly practical tools. They enable universities and quality-assurance agencies to systematise their evidence, verify the chronological consistency of data, and flag specific records for in-depth audits. For clarity throughout this paper, the Spanish term docentes has been rendered as academic staff. We use metrics such as the percentage of academic staff holding doctoral degrees, the percentage of academic staff employed full time, total enrolment, and indexed article records to reflect specific dimensions of university capacity. However, it is essential to note that these quantitative indicators serve as proxies; they cannot directly evaluate the actual quality of teaching and learning, the successful execution of an institution's unique mission, or the nuanced, context-specific assessments formulated by external reviewers. These data are therefore used as inputs for self-evaluation and policy learning rather than as substitutes for formal accreditation procedures.

1.2. Administrative Data and Institutional Analytics

Institutional analytics brings together information on organisational structures, resources, activities, and outcomes to support decision-making. It is distinct from learning analytics, which more often focuses on students and learning processes, although both fields raise similar concerns about transparency, privacy, fairness, accountability, and human oversight (Cerratto Pargman & McGrath, 2021; Ferguson, 2012; Guan et al., 2023; Khalil et al., 2023; Siemens & Long, 2011; Slade & Prinsloo, 2013; Viberg et al., 2022). Research on university performance indicators and data-driven organisations also warns against overlooking data quality and governance or reducing complex educational processes to simplified metrics (Komljenovic et al., 2025; Sarrico, 2022; Sharvashidze et al., 2023). The analysis in this article is grounded in applied statistics; it does not propose a machine-learning system or an automated accreditation procedure. Reviews of educational data mining and algorithmic fairness are included only as broader methodological context (Kalita et al., 2025; Pham et al., 2025).

National quality-assurance frameworks illustrate why administrative indicators should be interpreted as components of a broader evaluative process. The Ecuadorian model combines standards, documentary evidence, institutional self-assessment, external evaluation, and expert judgement, while the Colombian, Chilean, and Brazilian systems similarly integrate self-evaluation, external review, institutional context, and multiple sources of evidence. These frameworks support the use of quantitative indicators for verification and improvement, but do not treat them as independent substitutes for comprehensive evaluation (CACES, 2023; Consejo Nacional de Educación Superior, 2020; Consejo Nacional de Acreditación, 2022; Comisión Nacional de Acreditación, 2021; Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira [INEP], 2004).

1.3. Research Gap and Rationale for the 2018 Administrative Cut

Ecuador combines a national external-evaluation system led by CACES with publicly available administrative datasets produced by SENESCYT through the Integrated Higher Education Information System. The 2019 model defined the historical evaluation considered in this study, and its institutional outcomes were formally resolved in 2020. While the 2023 model is mentioned, it functions strictly as a point of contextual comparison (CACES, 2019, 2023). Public registries detailing academic staff, student enrolment, and indexed publications provided the necessary data. Although these metrics act as verifiable proxies for certain institutional characteristics, they inherently cannot replicate the complexity of the official evaluation framework. An assessment of source coverage confirmed that 2018 was the most recent period offering complete, comparable data across all three administrative databases prior to the accreditation decisions. Selecting this specific year safeguards the chronological sequence and minimises the risk of introducing post-resolution bias.

In the context of Ecuadorian higher education, few studies have explicitly examined the methodological consequences of analysing accreditation outcomes under extreme class imbalance. In particular, limited attention has been given to temporal linkage, documentary provenance, institutional-name harmonisation, publication-record deduplication, and case-level sensitivity within the same analytical workflow.

Developed within the applied-statistics research line of the Department of Exact Sciences, this study contributes primarily to the analysis of administrative data when the target outcome contains very few independent minority observations. Ecuadorian higher education accreditation provides a substantive educational context for examining data quality, metric instability, conditional uncertainty, and the leverage of individual observations. The contribution to educational research lies in an auditable framework for institutional analytics and data governance, rather than in an operational model for predicting accreditation status.

The methodological problem examined in this study may also arise in other Latin American systems that combine administrative evidence with accreditation decisions. Nevertheless, quality-assurance arrangements differ in their institutional design, regulatory objectives, evidence requirements, and relationships between evaluation and institutional improvement. The proposed workflow is therefore not presented as a uniform regional model, but as an adaptable analytical structure whose application requires renewed verification of data definitions, temporal coverage, linkage procedures, and local evaluation rules (Fernández Lamarra, 2003; Lemaitre, 2017; Rifo & Durán, 2023).

From an institutional standpoint, simply identifying historical differences between outcome groups offers limited practical value. The real challenge lies in interpreting these variations responsibly, particularly during self-assessment phases or when preparing for external audits. By adopting an auditable workflow, universities can expose underlying data inconsistencies and draw a clear line between descriptive evidence and predictive assumptions. Ultimately, this approach highlights exactly when a numerical finding requires further documentary, qualitative, or contextual scrutiny to be truly meaningful.

1.4. Research Question, Objective, and Contributions

The study addressed the following research questions:

RQ1. How did selected 2018 administrative indicators differ descriptively between the document-verified outcomes of the 2019 Ecuadorian institutional evaluation process, formally resolved in 2020?

RQ2. How did the presence of only three historical non-accreditation outcomes affect the stability and responsible interpretation of the rank-based statistical summaries?

The objective was to develop and evaluate an integrated, auditable applied-statistics workflow for analysing administrative indicators linked to a rare institutional outcome. The workflow combined temporal linkage, documentary provenance, data-quality control, descriptive group comparisons, rank-based effect sizes, and leave-one-minority-case sensitivity analysis. Secondary rank-ordering metrics were retained only to illustrate the consequences of extreme imbalance.

The resulting workflow has six stages: documentary source verification, institutional linkage, temporal ordering, data-quality control, primary descriptive and sensitivity analysis, and responsible interpretation. Accreditation in Ecuador provides the application, while the broader methodological issue is how to analyse administrative datasets that contain very few independent minority outcomes. Within higher education, the workflow is intended to support self-evaluation, preparation for quality assurance, and governance of institutional evidence.

2. Materials and Methods

2.1. Study Design and Unit of Analysis

A quantitative, observational, retrospective, and exploratory design was used, with the university or polytechnic school as the unit of analysis. The analytical dataset linked administrative indicators from 2018 to a document-verified binary outcome from the 2019 institutional evaluation process, formally resolved in 2020. The analysis focused on data quality, descriptive effect sizes, metric behaviour, conditional uncertainty, and the influence of individual observations. The design does not permit causal or individual-level inference, prospective prediction, calibration, institutional ranking, or operational classification.

2.2. Data Sources, Historical Outcome, and Temporal Window

Official records from CACES and SENESCYT/SIIES were integrated into a consolidated institutional workbook. Historical outcomes were checked against CACES documentation, including institution-specific 2020 resolutions referenced in Resolution No. 216-SO-29-CACES-2025 and the outcome pages maintained by CACES (Council for Quality Assurance in Higher Education, 2025, n.d.-a, n.d.-b). The administrative indicators came from

SENESCYT/SIIES datasets on academic staff, enrolment, and indexed publications (Secretariat of Higher Education, Science, Technology and Innovation, n.d.-a, n.d.-b, n.d.-c). Of the 55 institutional links, 53 were exact name matches, one required a normalised alias, and one contained both exact and alias evidence. Matching institutional names and verifying the documentary outcome were conducted as separate tasks.

For every included institution, the audit recorded the historical resolution or official results source, the issuing body and date, document type, source reference, relevant page or section, access date, verification status, outcome code, and reason for inclusion. The same audit documented why eight institutions in the current registry were not part of the linked historical dataset and whether their core 2018 data were available. Uncertain name equivalence and documentary discrepancies were resolved through manual review rather than assumption, and current accreditation status was never used as a substitute for the historical outcome.

The audit confirmed all 55 analytical outcomes and documented the eight exclusions. Four excluded institutions belonged to a later evaluation process and were accredited in 2024, so they had no comparable outcome from the 2019 cohort. The other four lacked both a comparable 2019-process outcome and a complete core record for 2018. The review left the analytical distribution unchanged at 52 accredited institutions and three historical non-accreditation outcomes. Because those three institutions may be indirectly identifiable, detailed audit files remain under controlled access. Authorised editors or reviewers may receive restricted access under documented confidentiality conditions for the limited purpose of checking outcome coding, linkage, and exclusion decisions.

Although the source files span 2015 to 2022 or 2023, the main analysis was restricted to 2018. This was the latest complete and comparable year before the historical outcome was formally resolved. The choice provided a common cross-sectional reference, maintained temporal precedence, and reduced information leakage. The wider panel was retained for traceability and possible later longitudinal work, but observations after 2019 were not treated as antecedent indicators.

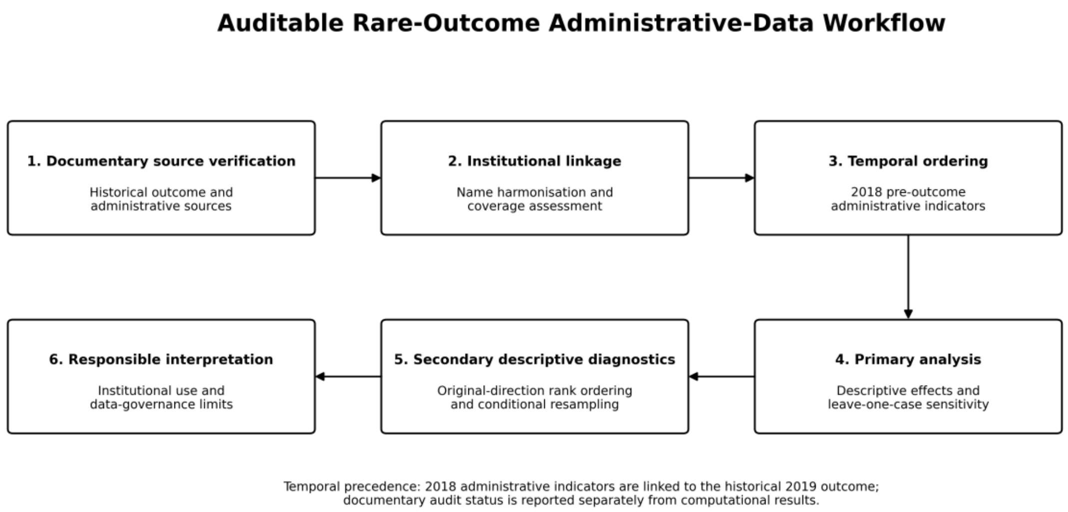

Figure 1. Auditable rare-outcome administrative-data workflow. Documentary source verification is reported separately from computational linkage and analysis.

Table 1. Data sources and traceability.

Source	Content Used	Source Period	Analytical Use	Traceability Information
CACES	Historical outcomes of the 2019 institutional evaluation process and evaluation framework	2019 process; outcomes resolved in 2020; contextual 2023 model	Definition and verification of the historical outcome; national quality-assurance context	Official 2020 institutional resolutions consolidated in Resolution No. 216-SO-29-CACES-2025 and maintained CACES pages
SENESCYT/SIIES - academic staff (docentes)	Total academic staff, PhD-qualified academic staff, and full-time academic staff	2015–2022	Institutional indicators for 2018	Original file, source URL, access date, temporal coverage, and checksum documented
SENESCYT/SIIES - enrolment	Aggregated institutional enrolment records	2015–2023	Enrolment and students per academic staff member in 2018	Original file, source URL, access date, temporal coverage, and checksum documented
SENESCYT/SIIES - indexed article records	Institutional publication records	2015–2023	Total and academic-staff-adjusted research output in 2018	Original file, source URL, access date, temporal coverage, and checksum documented
Integrated analytical workbook	Institution-year panel, name mapping, quality checks, and analytical cut	2015–2022 panel; 2018 indicators linked to the 2019 evaluation process resolved in 2020	Data linkage, quality control, and exploratory analysis	Author-constructed linkage; current registry used only as a reference

Note. Administrative indicators are partial institutional proxies and neither reproduce the complete CACES evaluation model nor directly measure educational quality. Source versions, original file names, access dates, temporal coverage, and SHA-256 checksums are documented in the accompanying audit package.

The 2018 indicators should not be interpreted as representing the complete body of evidence considered during accreditation. The official evaluation incorporated documentary, qualitative, and multi-year information, whereas the analytical cut provides a common pre-outcome administrative snapshot. The indicators are therefore treated as temporally ordered institutional descriptors rather than as complete predictors or reconstructions of the accreditation decision. Source coverage and the rationale for the analytical cut are documented in Table 1 and Supplementary Table S8.

2.3. Analytical Dataset and Outcome Coding

The final analytical dataset comprised 55 institutions with complete 2018 indicators and a document-verified historical outcome. Accreditation was coded as 1 and historical non-accreditation as 0 in the primary outcome variable. For secondary minority-oriented metrics, the coding was reversed so that historical non-accreditation represented the positive class. No missing values were imputed, and no over-sampling, under-sampling, synthetic-data generation, or class rebalancing was applied. Table 2 summarises cohort construction, while Table 3 defines the analytical hierarchy and inferential scope.

Table 2. Construction, documentary verification, and final composition of the linked analytical dataset.

Step	Description	Institutions	Interpretation
Current administrative reference panel	Institutions contained in the current administrative reference panel used to assess historical-cohort linkage and source coverage	63	Current reference panel; not the historical denominator
Comparable historical cohort with complete 2018 indicators	Institutions with document-verified outcomes from the 2019 evaluation process and complete 2018 core indicators	55	Final linked analytical dataset; no imputation
Document-verified outcome distribution	Accredited / historical non-accreditation	52 / 3	Extreme imbalance (approximately 17.3:1)
Outside the linked historical analytical cut	Four institutions evaluated in 2024; four without a comparable 2019-process outcome and complete 2018 core data	8	Exclusion rationale documented in the audit file

Note. The analytical dataset is a linked descriptive dataset rather than a reconstructed historical census. All 55 included outcomes and all eight exclusion decisions were documented through the audit protocol. The 63-institution reference panel was used only to assess linkage and coverage; the analytical denominator was the 55-institution historical cohort with a comparable 2019-process outcome and complete 2018 core indicators.

Table 3. Analytical hierarchy and inferential scope of the exploratory statistical analysis.

Element	Definition in this study
Unit and outcome	Institution-level observational analysis; document-verified outcome of the 2019 evaluation process, formally resolved in 2020
Positive class for secondary metrics	Historical non-accreditation
Primary analyses	Documentary and data-quality audit; observed prevalence; medians, interquartile ranges and full ranges; median differences; Cliff’s delta; leave-one-minority-case sensitivity

Element	Definition in this study
Secondary analyses	Majority-class reference; original-direction within-sample ROC AUC; conditional resampling ranges; supplementary precision–recall diagnostics
Inferential scope	Conditional description of the 55 linked institutions; no causal or population-level inference
Appropriate use	Data auditing, self-evaluation, and responsible institutional interpretation
Unsupported use	Stable prediction, ranking, sanctions, or automated regulatory classification

Note. Historical non-accreditation was defined as the positive class only for the secondary minority-oriented metrics. Administrative indicators were treated as partial institutional proxies and not as direct measures of educational quality.

2.4. Variables and Operational Definitions

Nine indicators were specified before the outcome-group results were examined. Selection was based on source availability, relevance to institutional capacity, comparability across institutions, complete coverage in the linked 2018 dataset, and the possibility of calculating transparent absolute or size-adjusted measures. The indicators were not intended to reconstruct the full CACES model or to represent every dimension of educational quality.

Variables were excluded when they lacked complete 2018 coverage, did not have sufficiently comparable definitions across institutions, could not be reconstructed transparently from the source files, or substantially duplicated information already represented by the selected measures. Indicator selection was therefore based on source quality, comparability, temporal consistency, and analytical transparency rather than on the magnitude or direction of the observed group differences. The complete source and variable audit is documented in Supplementary Tables S7, S8, and S14.

Table 4. Operational definitions and interpretive limits.

Domain	Indicator	Operational definition	Unit	Predefined conceptual direction	Main interpretive limitation
Academic staff	Total academic staff	Institutional academic-staff count in 2018	Count	No a priori direction	Strongly influenced by institutional size
Academic staff	PhD-qualified academic staff	Academic staff with a PhD in 2018	Count	Lower expected in historical non-accreditation	Influenced by size, mission, and disciplinary mix
Academic staff	Percentage of academic staff with a PhD	100 × PhD-qualified academic staff / total academic staff	Percentage	Lower expected in historical non-accreditation	Does not directly measure teaching quality
Academic staff	Full-time academic staff	Full-time academic-staff count in 2018	Count	Lower expected in historical non-accreditation	Influenced by size and contractual definitions
Academic staff	Percentage of full-time academic staff	100 × full-time academic staff / total academic staff	Percentage	Lower expected in historical non-accreditation	May vary by institutional mission and delivery mode
Institutional size	Total enrolment	Total institutional enrolment in 2018	Count	No a priori direction	Size descriptor; no directional accreditation hypothesis

Domain	Indicator	Operational definition	Unit	Predefined conceptual direction	Main interpretive limitation
Staffing ratio	Students per academic staff member	Total enrolment / total academic staff	Students per academic staff member	Higher expected in historical non-accreditation	Does not capture pedagogy, modality, or student support
Research	Indexed article records	Indexed article records attributed to the institution	Count	Lower expected in historical non-accreditation	Influenced by size and disciplinary field
Research	Article records per academic staff member	Article records / total academic staff	Article records per academic staff member	Lower expected in historical non-accreditation	Does not adjust for field, quality, or contribution

Note. Predefined conceptual directions were specified before group-specific results were interpreted. They reflect conceptual interpretation only, are not official CACES accreditation criteria, did not determine indicator inclusion, and were not used to invert original-direction AUCs or select thresholds. All indicators are partial administrative proxies.

The rationale for each predefined conceptual direction and its relationship to broad accreditation domains is documented in Supplementary Table S14. These directions were established for interpretive transparency and were not used as official CACES criteria, selection rules, or post hoc transformations of the statistical results.

To avoid treating scale and proportional measures as interchangeable, the indicators were divided into absolute and relative forms. Absolute measures included total academic staff, academic staff with a PhD, full-time academic staff, total enrolment, and indexed article records. Relative measures included the percentage of academic staff with a PhD, the percentage employed full time, students per academic staff member, and article records per academic staff member. The ratios reduce, but do not remove, the influence of institutional size. Neither group of indicators provides a direct measure of educational quality.

All relative indicators were recalculated from their constituent variables for the same institution and reporting year. The percentage with a PhD was calculated as 100 multiplied by PhD-qualified academic staff divided by total academic staff; the percentage full time used full-time academic staff as the numerator; students per academic staff member divided total enrolment by total academic staff; and article records per academic staff member used the same staff denominator. No institution had a zero denominator.

2.5. Data Linkage, Quality Control, and Article Deduplication

Institutional names were harmonised through a documented mapping that retained the source name, Unicode-normalised name, analytical identifier, and name used in the outcome source. Normalisation converted text to lowercase, removed accents, standardised punctuation, and regularised whitespace. The mapping yielded 53 exact matches, one match based on a normalised alias, and one record supported by both exact and alias evidence. Documentary verification of the outcome was kept separate from name matching. The workflow also checked for duplicate identifiers in the analytical worksheet, outcome file, and 2018 panel, and stopped if unresolved duplicates were found. Formula and range checks screened for invalid ratios, negative counts, components larger than total academic staff, and values requiring source review. Seven records were flagged and subsequently confirmed against the original staff, enrolment, or publication files; none required correction.

Publication records were reprocessed from the original SENESCYT workbook after locating the header in row 13 and mapping each record to an institution and year. The source did not provide a usable DOI field, and the records used in the sensitivity analysis contained no missing article titles or journal names. One deduplication key combined institution, publication year, and a Unicode-normalised title. A second combined the normalised title with journal or indexed-database information. Either rule could merge genuinely distinct records with identical titles or fail to join translated and variant titles, so both were treated as sensitivity checks on data quality rather than as a definitive bibliometric reconstruction. The recalculated institutional counts for 2018 agreed with the consolidated workbook.

Although exact and normalised-name matches were manually reviewed, residual linkage error cannot be ruled out completely. An undetected incorrect match could alter an institution's indicator profile or historical outcome assignment and could have a disproportionate effect because the minority group contains only three observations. Separating documentary outcome verification from computational name matching reduced this risk but did not eliminate it. The linkage results and verification categories are documented in Supplementary Table S1.

2.6. Exploratory Statistical Analysis

The analysis was deliberately exploratory and descriptive. Medians and rank-based effects were preferred because several indicators were skewed and closely tied to institutional size, while the historical non-accreditation group contained only three observations. Outcome prevalence was calculated for the linked dataset. For accredited institutions, continuous indicators were summarised with medians, interquartile ranges, and full ranges. For the three minority observations, the ordered values are provided in Supplementary Table S2 without preserving cross-variable institutional profiles. Names and direct identifiers were removed from public files, but the data are not described as fully anonymous because knowledgeable readers may still infer the institutions involved.

The primary analysis consisted of the data audit, observed outcome prevalence, descriptive distributions, median differences, Cliff's delta, and leave-one-minority-case sensitivity. The majority-class reference and original-direction within-sample ROC AUC were treated as secondary diagnostics.

Cliff's delta was defined as $\delta = P(XNA > XA) - P(XNA < XA)$, with ties contributing zero, where XNA represents an observation from the historical non-accreditation group and XA an accredited institution. Positive values indicate that the minority observations tended to be larger; negative values indicate the opposite. The statistic was used as an ordinal measure of group dominance (Cliff, 1993). Under a common group orientation and treatment of ties, Cliff's delta and ROC AUC are algebraically related expressions of the same pairwise ordering and should not be read as independent evidence. Cliff's delta was therefore the principal effect-size measure, whereas original-direction AUC served only as a secondary within-sample summary (Fawcett, 2006). No confirmatory null-hypothesis tests or multiplicity-adjusted claims of significance were made.

To illustrate the consequences of imbalance, a reference rule assigned every institution to the accredited class. Overall accuracy, recall of historical non-accreditation, specificity for accreditation, and balanced accuracy were then calculated. Clopper–Pearson exact binomial summaries were obtained with stats::binom.test in R. They are reported as mathematical descriptions of the linked data, not as estimates of generalisable predictive performance.

The analysis did not use over- or undersampling, synthetic minority generation, class weighting, or multivariable classifiers. With only three independent minority events, such procedures would repeatedly reuse the same outcome information and could create an unjustified impression of model stability.

For the secondary ordering analysis, historical non-accreditation was set as the positive class and larger raw indicator values were always treated as more positive. AUC estimates below 0.50 therefore show that minority observations tended to have lower raw values; they were not inverted after inspection. Total enrolment remained a size descriptor and had no directional accreditation hypothesis. Stratified percentile resampling used 3,000 replicates and deterministic indicator-specific seeds derived from the base seed 20260616, following the general framework of Efron and Tibshirani (1993). The 2.5th and 97.5th percentiles are reported as conditional resampling ranges rather than population confidence intervals. A 10,000-replicate check, repeated with the original and an alternative seed, was used to assess Monte Carlo variability. Because resampling cannot create new independent minority events, leave-one-minority-case analysis remained the main diagnostic of leverage. Average precision and precision–recall results are provided only as supplementary descriptive material and are not interpreted as evidence of predictive validity or as a basis for thresholds.

Spearman correlations among the nine indicators were used to examine redundancy and dependence between absolute counts and relative measures. The coefficients are described without p-values, multiple-testing claims, or fixed labels of magnitude.

All analyses were conducted in R 4.5.2 through the archived analytical workflow. The archive records the operating system, locale, time zone, package versions, principal functions, base and indicator-specific seeds, session information, and computational environment. A single pipeline generated the tables and figures. Exact binomial summaries were calculated with stats::binom.test, and Spearman correlations with stats::cor using method = "spearman". Custom functions in the archive implement the pairwise comparisons, rank-ordering summaries, resampling, and leave-one-minority-case diagnostics. The public reproducibility package is available in Zenodo at https://doi.org/10.5281/zenodo.21784668.

2.7. Ethical Considerations

The analysis used only public, aggregated records at the institutional level and involved no individual participants or interventions. The main ethical concern is not personal-data exposure in the usual sense, but the risk that partial administrative proxies could be turned into definitive judgements about institutional quality. For that reason, the results are not presented as rankings, regulatory thresholds, sanctions, or automated decisions. Public files remove institutional names and direct identifiers, yet they are described as de-identified rather than anonymous because the three minority cases may remain inferable. Detailed audit and mapping files are therefore subject to controlled access that balances scientific verification with reputational and re-identification risks.

3. Results

3.1. Data Integration, Coverage, and Quality Checks

The contemporary administrative reference panel contained 63 institutional identifiers, but inclusion in that panel did not imply complete or comparable information across years and sources. The documentary and source audits identified 55 institutions that belonged to the comparable 2019 evaluation cohort, had a linked record for 2018, and were complete for all nine indicators. Four institutions were evaluated in a later process in 2024, and four lacked both a comparable 2019-process outcome and a complete core record for 2018. Within the 55 retained institutions, complete-case coverage was 100%, and no indicator or outcome was imputed.

Automated validation found no duplicated identifiers in the analytical, outcome, or 2018 panel files; no invalid derived formulas; no proportions outside their permissible range; no staff components larger than total academic staff; no negative counts; and no unmatched 2018 publication records. Seven values met robust or predefined source-review criteria.

Each was checked against the original staff, enrolment, or publication file and confirmed without correction. These results support internal consistency between sources and computations, but they do not establish that every indicator has complete construct validity.

The exclusion audit documented why each of the eight institutions outside the historical analytical cut could not be included. Because these institutions did not share both a comparable 2019-process outcome and complete 2018 indicators, comparing them with the retained group would not estimate a common target population. Supplementary Table S1 reports the audit in aggregate, while the institution-level records remain restricted
.

Table 5. Data-integration, coverage, and quality-control checks for the linked analytical dataset.

Linkage and coverage audit	Result
Analytical institutions	55
Indicator coverage	100.0% among retained institutions
Duplicated analytical/outcome/panel IDs	0 / 0 / 0
Flagged extreme-value records reviewed	7
Flagged values confirmed without correction	7
Flagged values corrected after source verification	0
Invalid formulas or out-of-range proportions	0
Unmatched 2018 article records	0

Note. Coverage refers to completeness among the 55 retained institutions. The checks assess structural, computational, and source consistency, not complete construct validity.

3.2. Historical Outcome Distribution and Descriptive Comparisons

Among the 55 linked institutions, 52 were accredited (94.55%) and three had a historical non-accreditation outcome (5.45%), corresponding to a ratio of approximately 17.3 to 1. These figures refer to the 2019 institutional evaluation process formally resolved in 2020. They describe the linked analytical cohort only and should not be interpreted as prevalence estimates for the Ecuadorian higher education system as a whole
.

Table 6. Document-verified outcomes of the 2019 institutional evaluation process in the linked analytical dataset.

Historical outcome	Primary outcome code	Positive-class code for secondary metrics	Institutions	Percentage
Accredited	1	0	52	94.55%
Historical non-accreditation	0	1	3	5.45%

Note. Historical non-accreditation was coded as the positive class only for the secondary minority-oriented metrics. All historical outcomes were verified through the documentary protocol described in Section 2.2.

Several indicators differed descriptively between the two historical outcome groups. Compared with accredited institutions, the historical non-accreditation group had lower medians for total academic staff (316 vs. 414), PhD-qualified staff (32 vs. 60.5), percentage of staff with a PhD (10.13% vs. 12.54%), full-time staff (251 vs. 285.5), indexed article records (78 vs. 129.5), and article records per staff member (0.247 vs. 0.264). Its medians were higher for the percentage of full-time staff (76.76% vs. 69.56%), enrolment (8,961 vs. 8,157), and students per academic staff member (28.358 vs. 16.458). For total enrolment, a positive median difference occurred alongside a negative Cliff's delta. This was not a contradiction: the three minority observations were heterogeneous, so their median was slightly higher even though pairwise comparisons more often placed them below accredited institutions. Given the size of the minority group, these contrasts are conditional descriptions rather than stable population differences or causal explanations.

Table 7. Descriptive distributions and rank-based effect sizes by historical accreditation outcome.

Indicator	Accredited institutions (n=52)	Historical non-accreditation (n=3)	Median difference	Cliff's delta
Total academic staff	414 [IQR 208.8–981; range 41–2,927]	316 [range 27–327]	−98	−0.436
PhD-qualified academic staff	60.5 [IQR 33.8–102.2; range 6–292]	32 [range 2–35]	−28.5	−0.673
Percentage of academic staff with a PhD	12.54% [IQR 8.64%–18.30%; range 2.00%–91.07%]	10.13% [range 7.41%–10.70%]	−2.42 percentage points	−0.500
Full-time academic staff	285.5 [IQR 134.2–636.5; range 27–1,874]	251 [range 18–297]	−34.5	−0.359
Percentage of full-time academic staff	69.56% [IQR 48.80%–80.49%; range 29.09%–99.43%]	76.76% [range 66.67%–93.99%]	+7.20 percentage points	+0.397
Total enrolment	8,157 [IQR 3,322.8–15,699.2; range 455–59,696]	8,961 [range 187–9,886]	+804	−0.269
Students per academic staff member	16.458 [IQR 13.477–20.477; range 4.775–44.112]	28.358 [range 6.926–30.232]	+11.900	+0.308
Indexed article records	129.5 [IQR 62–235.5; range 12–588]	78 [range 1–127]	−51.5	−0.474
Article records per academic staff member	0.264 [IQR 0.192–0.389; range 0.080–1.125]	0.247 [range 0.037–0.388]	−0.017	−0.256

Note. Accredited institutions are summarised as median [IQR; range]; the three minority observations are summarised as median [range]. For the historical non-accreditation group, n = 3; therefore, the reported median is the middle of the three ordered observations and should not be interpreted as a stable population location estimate. The ordered minimum, median, and maximum are reported in Supplementary Table S2. Median difference = historical non-accreditation median − accredited median. Positive Cliff's delta values indicate higher minority-group values; negative values indicate lower values. Total enrolment was defined as a size descriptor. No confirmatory null-hypothesis tests or multiplicity-adjusted significance claims were made.

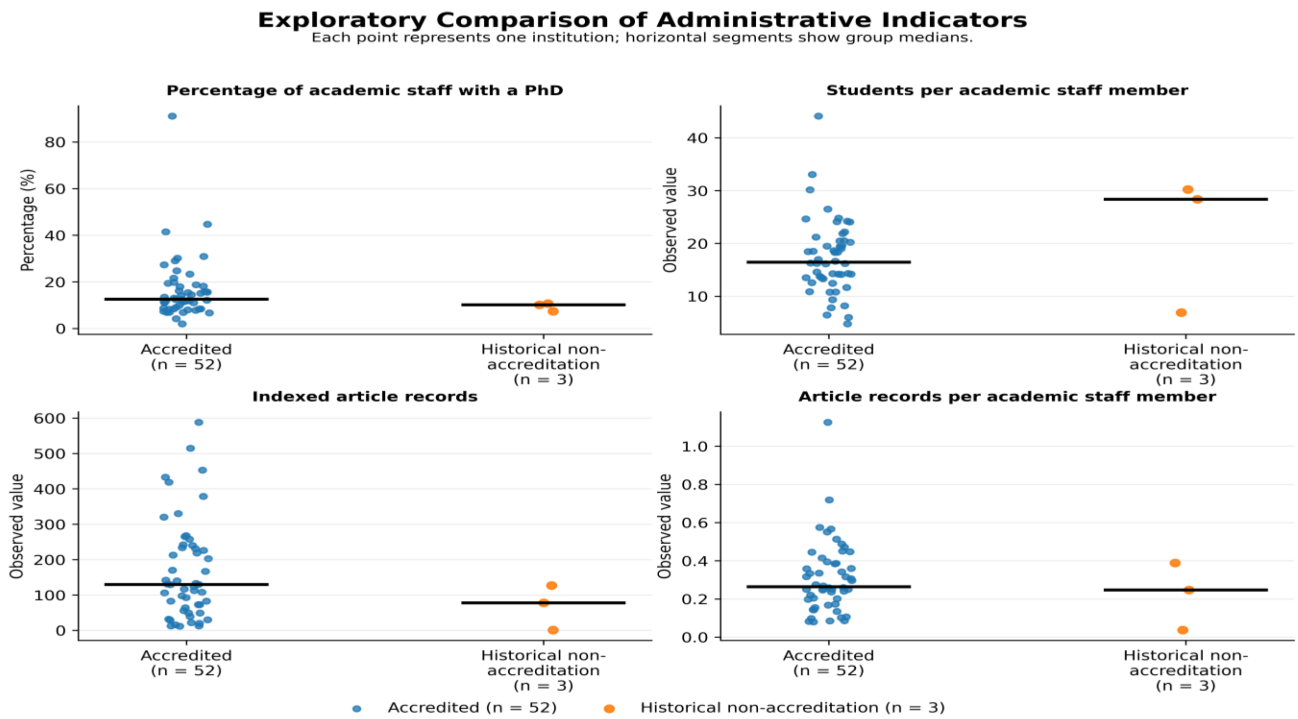

Figure 2. Individual-value distributions of four predefined illustrative indicators representing relative academic capacity, staffing ratio, research output, and size-adjusted research output. Each point represents one institution; horizontal segments indicate group medians, and descriptive results for all nine indicators are reported in Table 7.

3.3. Effect of Imbalance on the Majority-Class Reference

A rule that classified every institution as accredited produced an overall accuracy of 94.55%, specificity of 100% for accreditation, sensitivity of 0% for historical non-accreditation, and balanced accuracy of 50%. In other words, the apparently high accuracy arose entirely from the dominance of the majority class; none of the three minority outcomes was identified. Table 8 reports the point estimates, and Supplementary Table S3 provides the confusion matrix and exact binomial summaries.

Table 8. Majority-class reference metrics for the document-verified historical outcome in the linked analytical dataset.

Metric	Estimate	Interpretation
Overall accuracy	0.9455	Apparently high because the accredited class dominates the linked dataset
Sensitivity for historical non-accreditation	0.0000	No historical non-accreditation outcome is identified
Specificity for accreditation	1.0000	All accredited institutions are assigned to the majority class
Balanced accuracy	0.5000	Equivalent to a non-informative reference

Note. Sensitivity denotes recall for historical non-accreditation. Balanced accuracy is the arithmetic mean of minority-class sensitivity and accredited-class specificity. Point estimates describe the linked dataset only; Clopper–Pearson exact binomial summaries and the confusion matrix are reported in Supplementary Table S3. Matthews correlation coefficient was undefined because the reference rule predicted only one class.

3.4. Secondary Within-Sample Rank-Ordering Metrics

Original-direction within-sample ROC AUC was retained as a secondary description of pairwise rank ordering. The largest estimates were observed for percentage of full-time academic staff (AUC = 0.699) and students per academic staff member (AUC = 0.654), whereas the remaining estimates were below 0.50 in their original direction. Because the calculations relied on only three minority observations, they are not interpreted as evidence of out-of-sample discrimination. Complete estimates and conditional resampling ranges are reported in Supplementary Table S16 and Figure S8.

3.5. Leave-One-Minority-Case Sensitivity

Because resampling cannot add independent events, case deletion provided the clearest assessment of leverage. Cliff's delta changed substantially when each minority observation was omitted in turn. For students per academic staff member, the full-data estimate was 0.308, whereas the leave-one-case values ranged from 0.000 to 0.904. The corresponding ranges were 0.135 to 0.635 for percentage of full-time academic staff, −0.481 to 0.096 for total enrolment, −0.692 to −0.212 for indexed article records, and −0.635 to 0.115 for article records per academic staff member. Only two minority outcomes remained after each deletion. These values are diagnostics of observation-level influence, not validation results or alternative estimates of a population effect.

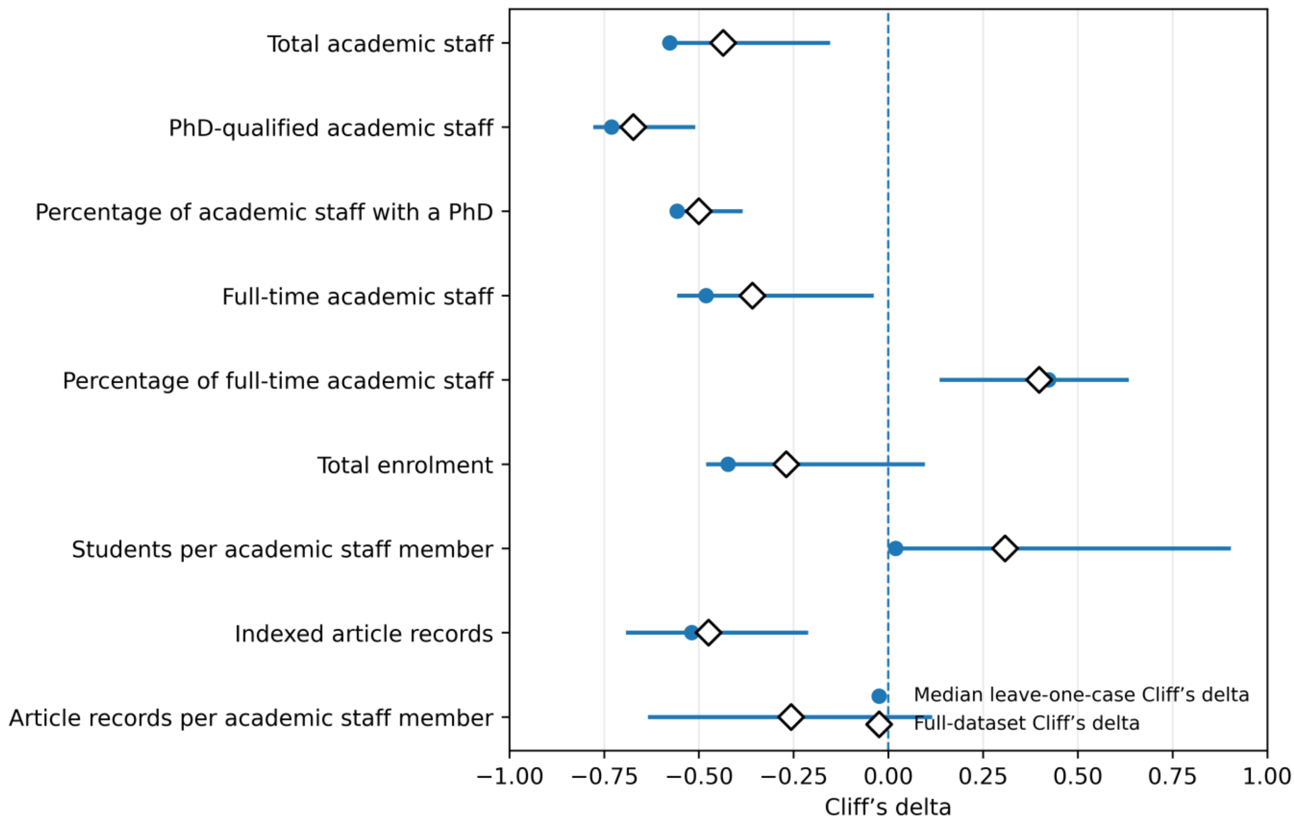

Figure 3. Leave-one-minority-case sensitivity of Cliff's delta estimates in predefined conceptual order. Horizontal segments show the minimum–maximum delta after removing each minority observation, points show median leave-one-case estimates, and open diamonds show full-dataset estimates. The dashed line indicates δ = 0.

3.6. Publication-Record Deduplication Sensitivity

Reprocessing the original publication workbook identified possible duplicate records under both the title-only and title-plus-source keys. Neither rule changed any institution's 2018 article-record count, and the descriptive effects and original-direction AUC values for the analytical year therefore remained unchanged. Detailed record-level and institution-year results are provided in Supplementary Table S4. The agreement across counting approaches confirms computational reconciliation for 2018, but it does not establish that every source record is bibliometrically unique or equivalent.

3.7. Correlation Structure

The absolute indicators were strongly redundant with institutional size. Total academic staff correlated closely with full-time academic staff (Spearman ϱ = 0.950), total enrolment (ϱ = 0.949), PhD-qualified academic staff (ϱ = 0.824), and indexed article records (ϱ = 0.818). Full-time staff also correlated with enrolment (ϱ = 0.950) and indexed article records (ϱ = 0.830). Among the relative measures, article records per academic staff member correlated with the percentage of staff holding a PhD (ϱ = 0.419) and the percentage employed full time (ϱ = 0.412). These coefficients are reported descriptively and are not assigned fixed qualitative categories.

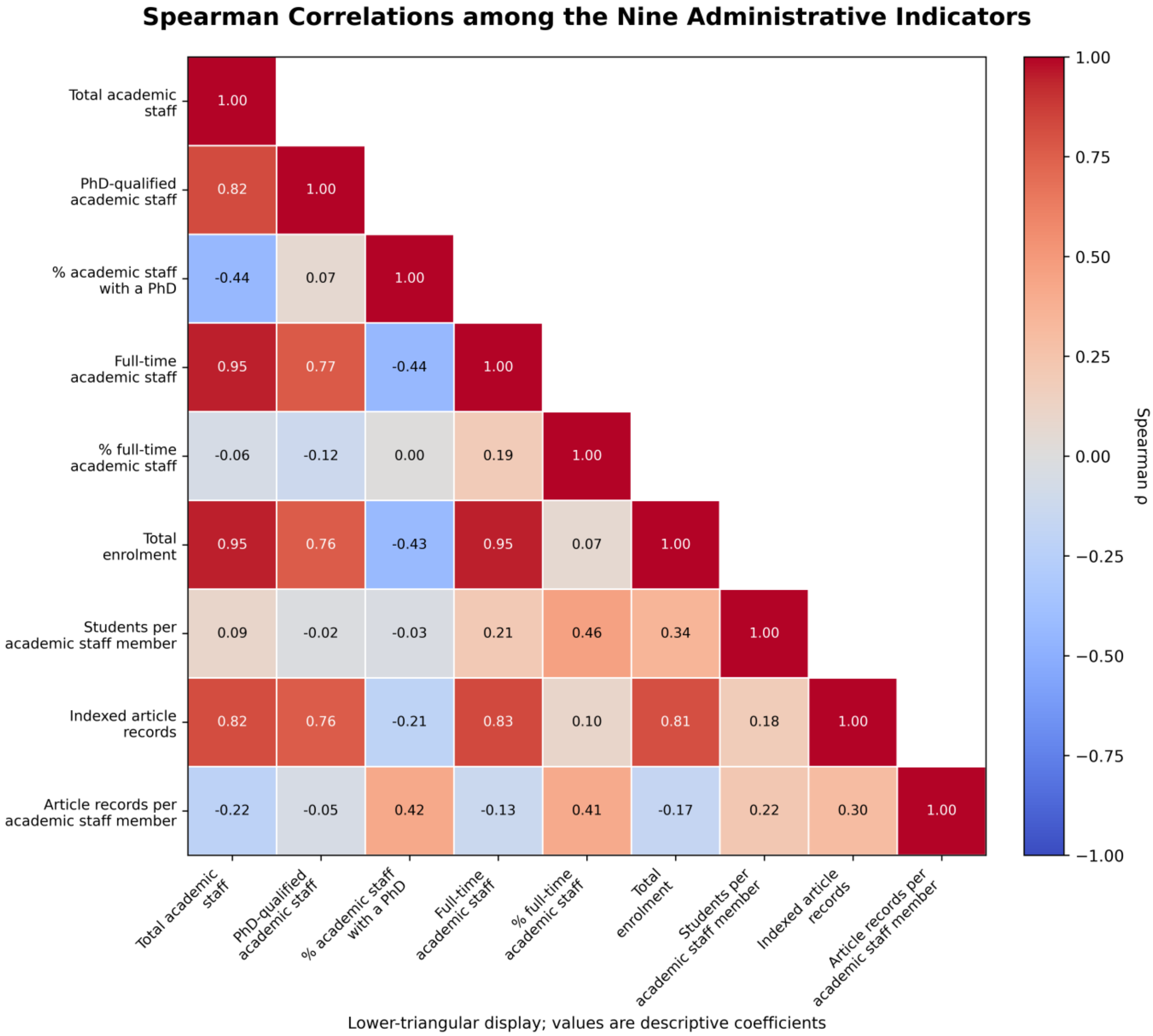

Figure 4. Lower-triangular Spearman correlation matrix for the nine 2018 administrative indicators across the 55 linked institutions. The lower-triangular display avoids symmetric duplication and

illustrates the redundancy between several absolute indicators and institutional size. Coefficients are presented descriptively without significance testing.

4. Discussion

4.1. Principal Findings

When an institutional outcome is represented by only three independent minority events, the interpretation of statistical summaries becomes highly dependent on the individual cases. The majority-class reference illustrates this clearly: it achieved high overall accuracy while missing every historical non-accreditation outcome. The rank-based effects and leave-one-case results likewise showed that the apparent separation between outcome groups was specific to the linked dataset and sensitive to particular minority observations. Alternative performance metrics can describe different aspects of the same data, but they cannot replace independent outcome information (Chicco & Jurman, 2020; He & Garcia, 2009; Saito & Rehmsmeier, 2015).

Independent numerical recalculation reproduced the archived point results within the predefined tolerance, supporting numerical consistency. Full end-to-end execution of the archived R workflow remains an additional reproducibility check. Statistical stability is a different question. Here, the dominant source of uncertainty was the small number of independent minority outcomes, as the case-deletion analysis made explicit. The value of the study lies less in classifier performance than in bringing documentary provenance, data-quality control, descriptive effect sizes, and observation-level sensitivity into a single transparent analysis.

4.2. Interpretation of Administrative Indicators

The historical non-accreditation group had lower medians for doctoral qualifications, absolute staff counts, and publication output, but higher medians for the student-to-staff ratio and the percentage of staff employed full time. These patterns may reflect differences in institutional size and mission, variation in reporting practices, or characteristics of the historically linked cohort. However, these possible explanations were not tested directly and should be treated as contextual hypotheses rather than empirical findings of the present study. None of the observed contrasts should be read as a causal mechanism or as evidence that a single indicator determines accreditation.

The higher median percentage of full-time academic staff in the historical non-accreditation group should not be interpreted as either a protective or an adverse factor. It may reflect institutional mission, contractual structure, delivery mode, reporting practice, or the particular composition of the three minority cases. Similarly, the higher students-per-academic-staff median may reflect differences in institutional scale or staffing organisation, but the present data cannot determine the underlying mechanism. The coexistence of these patterns illustrates why isolated administrative indicators require contextual and qualitative interpretation (Beerkens, 2018; Sarrico, 2022; Sharvashidze et al., 2023).

These measures are administrative descriptors, not direct observations of educational quality. The number of staff with a PhD may reflect one aspect of academic capacity, but it says little by itself about teaching effectiveness, curriculum relevance, student learning, or fulfilment of institutional mission. Counts of staff and publications also rise with institutional scale. With only three minority outcomes and strong correlations among several indicators, the data cannot support stable multivariable estimation.

4.3. Imbalance, Recall, and Methodological Implications

The results reinforce the well-established caution against using overall accuracy as the main performance measure in imbalanced data (Chicco & Jurman, 2020; He & Garcia, 2009; Saito & Rehmsmeier, 2015). Neither resampling nor greater algorithmic complexity can create information from additional independent events when only three minority outcomes exist. Conditional resampling repeatedly reuses the same cases, whereas leave-one-minority-case analysis shows directly how much each case matters. Under these conditions, the data cannot support stable thresholds, calibration, multivariable effects, or generalisable prediction.

Statistical approaches for sparse binary outcomes include bias-reduced likelihood, penalised logistic regression, and rare-event logistic regression. Firth's method was developed to reduce maximum-likelihood bias, while later work demonstrated its usefulness under logistic separation; King and Zeng proposed corrections for probability estimation in rare-event data (Firth, 1993; Heinze & Schemper, 2002; King & Zeng, 2001). However, these methods do not create additional independent institutional outcomes. With only three minority events and strongly correlated indicators, their application here would still depend heavily on untestable modelling assumptions and individual observations.

The original direction of AUC was preserved so that the sign of the observed ordering remained visible; estimates were not converted after inspection to max(AUC, 1 − AUC). Median differences and Cliff's delta provide the main descriptions of between-group ordering, while the leave-one-case ranges show more clearly how unstable those descriptions become when a single minority institution is removed.

4.4. Contribution from Applied Statistics to Educational Research

The disciplinary contribution lies in applied statistics rather than in a new accreditation model or a novel estimator. Its value comes from combining established procedures—source verification, linkage, quality control, rank-based effects, conditional resampling, and case-level sensitivity—within a single auditable workflow for an administrative outcome with exceptionally few independent events. This integration makes the limits of the data, the leverage of individual observations, and the range of defensible interpretations visible to the reader.

This matters because universities and quality-assurance agencies increasingly depend on administrative indicators, while research continues to emphasise the need for data quality, transparent methods, responsible governance, and safeguards against unfair or overextended use (Kalita et al., 2025; Komljenovic et al., 2025; Pham et al., 2025; Sarrico, 2022; Sharvashidze et al., 2023). Applying the workflow elsewhere would require a comparable definition of the outcome, a justified temporal window, renewed verification of the source data, enough independent events for the intended statistical analysis, and review by experts familiar with the local context.

Transferability requires a clearly defined and document-verifiable outcome, a defensible common temporal window, comparable administrative definitions, traceable institutional linkage, transparent exclusion rules, and enough independent minority observations for the intended analytical purpose. No universal minimum number of minority events is proposed because adequacy depends on the planned analysis. When unfavourable outcomes remain extremely scarce, the workflow should be limited to auditing, descriptive effects, and case-level sensitivity rather than prediction, calibration, multivariable modelling, or threshold development (He & Garcia, 2009; Chen et al., 2024).

4.5. Institutional and Educational Implications

Universities and polytechnic schools could use the workflow to organise data inventories, verify source records, check temporal consistency, frame questions for self-evaluation, and identify evidence that needs documentary or qualitative follow-up. As a contribution from applied statistics in the Exact Sciences, its role is to strengthen data governance, reproducibility, uncertainty assessment, and quality control of institutional evidence. It is not a direct evaluation of pedagogy. Any practical interpretation should remain connected to curriculum, teaching practice, student support, institutional mission, and the broader evaluation context.

In institutional self-evaluation, the workflow can be used to verify whether reported indicators share consistent definitions, reporting periods, institutional identifiers, calculation rules, and documentary sources. Comparable guidance in Colombia and Brazil treats self-evaluation as a structured process for identifying strengths, weaknesses, and evidence for improvement rather than as the mechanical completion of an accreditation requirement. Under this approach, an unusual numerical value should trigger source verification and contextual review before it is interpreted as institutional performance (Consejo Nacional de Acreditación, 2022; Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira [INEP], 2004).

For external reviewers, the audit trail and the leave-one-minority-case analysis can indicate whether an apparent difference is consistently supported by the linked data or depends strongly on a particular observation. This information should guide requests for documentary clarification and contextual explanation rather than produce an automatic judgement. Accreditation research has shown that its institutional value depends not only on formal certification, but also on how evidence is interpreted, discussed, and incorporated into internal quality-assurance practices (Acevedo-De-los-Ríos & Rondinel-Oviedo, 2022; Salto, 2018).

Administrative indicators are most useful when their definitions, coverage, and limitations are documented. Under those conditions they can support evidence organisation, contextual review, and internal monitoring, although their meaning still depends on institutional mission, evaluation rules, and reporting practice (Beerkens, 2018; Harvey & Williams, 2010; Sarrico, 2022; Sharvashidze et al., 2023; Stensaker, 2008). Work on quality assurance in Latin America similarly links external accountability with institutional self-regulation and continuous improvement (Aguilera, 2019). The present findings add a specific statistical caution: with only three unfavourable outcomes, even technically appropriate metrics remain highly sensitive to individual institutions and cannot justify automated accreditation decisions.

At the system level, quality-assurance agencies could incorporate comparable procedures into data-submission protocols, version control, institutional-name harmonisation, evidence verification, correction mechanisms, and pre-evaluation checks. Official accreditation frameworks in Ecuador, Colombia, Chile, and Brazil already recognise that evaluation requires multiple sources of evidence, institutional context, self-evaluation, and external review. The proposed workflow contributes specifically to the statistical auditability of that evidence and should remain subordinate to the broader evaluative process (CACES, 2023; Consejo Nacional de Educación Superior, 2020; Comisión Nacional de Acreditación, 2021; INEP, 2004).

For quality-assurance bodies, a comparable workflow could improve interoperability, source auditing, and preparation for external review. It should not be converted into an automatic accreditation rule. Any consequential use would require expert judgement, qualitative evidence, a documented audit trail, opportunities for contextual explanation, and a clear mechanism for correction or appeal. Table 9 summarises the interpretations supported by the analysis and the uses that fall outside its scope.

Table 9. Supported and unsupported interpretations and uses within the scope of this study.

Dimension	Supported interpretation or use	Unsupported interpretation or use

Dimension	Supported interpretation or use	Unsupported interpretation or use
Statistical	Explore imbalance, conditional uncertainty, effect sizes, and descriptive ordering	Claim prediction, stable classification, population inference, or generalisable performance
Institutional	Support self-evaluation, internal evidence review, and data-quality assessment	Rank universities or compare institutional reputation using indicators alone
Quality assurance	Prepare evidence, improve traceability, and inform contextual external review	Replace peer review, qualitative evidence, or expert judgement
Public policy	Strengthen information systems and identify data-governance needs	Apply sanctions or allocate resources based only on unvalidated indicator separation
Data governance	Document source versions, access controls, human review, correction procedures, and an audit trail	Release identifiable minority profiles or use unverified data for consequential decisions

Note. The workflow supports educational review and quality-assurance preparation; it does not replace peer judgement or contextual institutional evaluation.

4.6. Responsible Data Governance

Responsible institutional analytics depends on purpose limitation, transparency, auditability, proportionality, controlled access, and human oversight (Cerratto Pargman & McGrath, 2021; Guan et al., 2023; Khalil et al., 2023; Komljenovic et al., 2025; Slade & Prinsloo, 2013; Viberg et al., 2022). Misclassification can affect institutions, academic communities, and reputations, especially when administrative indicators are presented as rankings (Hazelkorn, 2015). Public reporting should therefore distinguish de-identified profiles from genuinely anonymous data, avoid releasing identifiable minority profiles, specify permitted uses, and retain an auditable route for verification and correction.

5. Limitations

The most important limitation is the scarcity of events. Three historical non-accreditation outcomes are insufficient for confirmatory inference, stable prediction, calibration, multivariable modelling, or the selection of operational thresholds. Figure 3 shows the resulting case-level leverage. Conditional resampling ranges describe repeated draws from the observed data, but they cannot substitute for additional independent outcomes.

The analysis also uses a single year before the outcome and an institution-level unit of observation. Selecting 2018 improved temporal clarity, but it does not capture institutional trajectories or the multi-year evidence considered during formal accreditation. The findings cannot be transferred to individual students or members of staff.

The outcome represents a specific historical accreditation process and should not be interpreted as the institutions' current accreditation status or present educational quality. Changes in institutional conditions, reporting practices, or evaluation frameworks after the analysed period fall outside the scope of the study.

Residual measurement error remains possible even though the audit verified the historical outcomes, exclusions, and flagged values. The administrative and documentary records were originally produced for different institutional purposes and may not be fully equivalent. Moreover, the four institutions assessed in 2024 and the four without both a comparable historical outcome and complete 2018 data do not form a common comparison group with the retained cohort. The results are therefore conditional on the final linked dataset.

The indicators themselves are incomplete and often scale-dependent. They omit important academic processes, student experience, governance, and the qualitative or peer-reviewed evidence used in accreditation. Definitions and reporting practices may vary, and computational agreement does not establish construct validity. Title-based deduplication may merge distinct records that share a normalised title or fail to join translated and variant titles. Although the supplementary files report ordered values without direct identifiers, combinations of indicators may still allow indirect identification; detailed mappings therefore remain restricted. The study does not estimate pedagogical quality, learning outcomes, teaching effectiveness, or the effect of an educational intervention.

The workflow was not formally evaluated with institutional self-evaluation teams, external reviewers, CACES personnel, or SENESCYT data administrators. The present study therefore demonstrates analytical and computational feasibility rather than confirmed operational usability. Future research should examine whether the workflow is understandable, practical, and compatible with existing institutional and external-review procedures.

Finally, the study has no external or temporal validation. Future work should seek additional independent outcomes, repeat source verification for later accreditation cycles, assess temporal replication, involve relevant stakeholders, and strengthen safeguards against re-identification or regulatory misuse. Questions of implementation cost and usability should be examined only if the workflow is later developed for operational use.

6. Conclusions

Using document-verified accreditation outcomes from Ecuadorian higher education, this study examined how data quality, rank-based effects, and case-level instability behave under extreme class imbalance. The majority-class reference achieved high accuracy only because accredited institutions dominated the dataset; it failed to identify any historical non-accreditation outcome. Accuracy alone was therefore an inadequate description of performance.

The three minority outcomes placed a strict limit on what could be learned from the data. Effect sizes and rank-ordering summaries changed when individual cases were removed, and conditional resampling could not add new independent information. The evidence does not support stable prediction, calibration, operational thresholds, multivariable modelling, institutional ranking, or automated classification.

The principal contribution is an auditable applied-statistics workflow that integrates documentary provenance, institutional linkage, data-quality control, rank-based description, and case-level sensitivity. Developed within the applied-statistics research line of the Department of Exact Sciences and applied to higher education quality assurance, the workflow can strengthen institutional self-evaluation, external-review preparation, reproducibility, and responsible data governance. Its application must preserve the role of contextual evidence, qualitative assessment, and expert judgement. Future research should prioritise additional independent outcomes and replication across accreditation cycles.

Supplementary Materials: The supporting files include Table S1, aggregated documentary-verification and exclusion audit; Table S2, ordered minority-group values without cross-variable institutional linkage; Tables S3a–S3b, exact summaries and the confusion matrix for the majority-class reference; Table S4, sensitivity of publication-output measures to record deduplication; Table S5, average precision by administrative indicator; Table S6, computational stability of the conditional resampling limits; Table S7, source review of flagged values; Table S8, source metadata and reference audit; Table S9, detailed leave-one-minority-case results; Table S10, the Spearman correlation matrix; Tables S11–S12, exploratory precision–recall results; Table S13, complete descriptive effects, AUC estimates, and leave-one-case ranges; Table S14, the conceptual accreditation-domain crosswalk; Table S15, comparison of recalculated and workbook publication counts; Table S16, original-direction within-sample ROC AUC and conditional resampling ranges; and Figures S1–S8, supplementary computational, sensitivity, rank-ordering, and precision–recall displays.

Author Contributions: Conceptualization, D.S.; Methodology, D.S.; Formal Analysis, D.S.; Data Curation, D.S.; Visualization, D.S.; Writing—Original Draft Preparation, D.S.; Writing—Review and Editing, D.S. The author has read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical review was not required because the study exclusively analysed publicly available, aggregated, institution-level administrative records and involved no human participants, interventions, or individual-level personal data.

Informed Consent Statement: Not applicable.

Data Availability Statement: The public reproducibility package containing processed non-identifying outputs, R code, variable definitions, data-quality summaries, computational-environment records, supplementary tables and figures, and scripts used to generate the analyses is available in Zenodo at https://doi.org/10.5281/zenodo.21784668. The package allows reproduction of the univariate descriptive, rank-based, conditional resampling, and leave-one-minority-case analyses and verification of the aggregate correlation matrix. Recalculating institution-level cross-indicator correlations requires the restricted linked dataset. Source mappings, identifiable audit records, the linked institutional workbook, and minority-case profiles remain under controlled access because complete anonymisation cannot be guaranteed. Authorised editors or reviewers may request access under documented confidentiality and data-governance conditions. Public and restricted materials may not be used to identify, rank, sanction, or automatically classify institutions.

Conflicts of Interest: The author declares no conflict of interest.

Use of Generative Artificial Intelligence: During manuscript preparation, the author used OpenAI ChatGPT (GPT-5.6 Thinking) exclusively to support English-language editing, structural review, and the rephrasing of selected passages in the Introduction, Materials and Methods, Discussion, Limitations, and Conclusions. The tool was not used for data collection, documentary verification, institutional linkage, statistical analysis, result calculation, figure generation, or independent scientific interpretation. All AI-assisted changes were reviewed and verified by the author, who accepts full responsibility for the final manuscript.

Acknowledgments: The author acknowledges the institutional and public data sources that supported the documentary and administrative-data audit.

References

Acevedo-De-los-Ríos, A., & Rondinel-Oviedo, D. R. (2022). Impact, added value and relevance of an accreditation process on quality assurance in architectural higher education. Quality in Higher Education, 28(2), 186–204. https://doi.org/10.1080/13538322.2021.1977482

Aguilera, R. (2019). Aseguramiento de la calidad de la educación superior. Revista Educación Superior y Sociedad, 22(22), 131–154. https://ess.iesalc.unesco.org/index.php/ess3/article/view/36

Beerkens, M. (2018). Evidence-based policy and higher education quality assurance: Progress, pitfalls and promise. European Journal of Higher Education, 8(3), 272–287. https://doi.org/10.1080/21568235.2018.1475248

Cerratto Pargman, T., & McGrath, C. (2021). Mapping the ethics of learning analytics in higher education: A systematic literature review of empirical research. Journal of Learning

Analytics, 8(1), 123–139. https://doi.org/10.18608/jla.2021.1

Chen, W., Yang, K., Yu, Z., Shi, Y., & Chen, C. L. P. (2024). A survey on imbalanced learning: Latest research, applications and future directions. Artificial Intelligence Review, 57, Article 137. https://doi.org/10.1007/s1139-04-1139-04

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient over F1 score and accuracy in binary classification evaluation. BMC Genomics, 21, Article 6. https://doi.org/10.1186/s1139-04-139-04

Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. Psychological Bulletin, 114(3), 494–509. https://doi.org/10.1037/139-04.114.3.494

Comisión Nacional de Acreditación. (2021). Criterios y estándares de calidad para la acreditación institucional del subsistema universitario. CNA Chile. Retrieved August 3, 2026, from https://www.cnachile.cl/noticias/SiteAssets/Paginas/Forms/AllItems/CyE%20INSTITUCIONAL%20SUBSISTEMA%20UNIVERSITARIO.pdf

Consejo Nacional de Acreditación. (2022). Actualización de los aspectos por evaluar para la autoevaluación con fines de acreditación en alta calidad de las instituciones de educación superior. CNA Colombia. Retrieved August 3, 2026, from https://www.cna.gov.co/1779/articles-412513_norma.pdf

Consejo Nacional de Educación Superior. (2020). Acuerdo 02 del 1 de julio de 2020: Por el cual se actualiza el modelo de acreditación en alta calidad. Ministerio de Educación Nacional de Colombia. Retrieved August 3, 2026, from https://www.mineducacion.gov.co/portal/Educacion-superior/CESU/399567:Acuerdo-02-del-1-de-julio-de-2020

Council for Quality Assurance in Higher Education. (2019). Modelo de evaluación externa de universidades y escuelas politécnicas 2019. CACES. Retrieved July 24, 2026, from https://www.caces.gob.ec/universidades-y-escuelas-politecnicas-2/

Council for Quality Assurance in Higher Education. (2023). Modelo de evaluación externa con fines de acreditación para el aseguramiento de la calidad de las universidades y escuelas politécnicas 2023. CACES. Retrieved July 24, 2026, from https://www.caces.gob.ec/universidades-y-escuelas-politecnicas-3/

Council for Quality Assurance in Higher Education. (2025, October 22). Resolución No. 216-SO-29-CACES-2025: Ampliación de la vigencia de acreditación de las universidades y escuelas politécnicas. CACES.

Council for Quality Assurance in Higher Education. (n.d.-a). Evaluación externa con fines de acreditación de las UEP no acreditadas en el proceso de evaluación institucional 2019. Retrieved July 23, 2026, from https://www.caces.gob.ec/evaluacion-externa-confines-de-acreditacion-de-las-uep-no-acreditadas-2019/

Council for Quality Assurance in Higher Education. (n.d.-b). Listado de universidades y escuelas politécnicas acreditadas. Retrieved July 23, 2026, from https://www.caces.gob.ec/listado-de-universidades-y-escuelas-politecnica/

Efron, B., & Tibshirani, R. J. (1993). An introduction to the bootstrap. Chapman & Hall/CRC.

European Association for Quality Assurance in Higher Education, European Students' Union, European University Association, & European Association of Institutions in Higher Education. (2015). Standards and guidelines for quality assurance in the European Higher Education Area.

Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Fernández Lamarra, N. (2003). Higher education, quality evaluation and accreditation in Latin America and MERCOSUR. European Journal of Education, 38(3), 253–269. https://doi.org/10.1111/139-04.00145

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. International Journal of Technology Enhanced Learning, 4(5/6), 304–317. https://doi.org/10.1504/IJTEL.2012.051816

Firth, D. (1993). Bias reduction of maximum likelihood estimates. Biometrika, 80(1), 27–38. https://doi.org/10.1093/biomet/80.1.27

Guan, X., Feng, X., & Islam, A. Y. M. A. (2023). The dilemma and countermeasures of educational data ethics in the age of intelligence. Humanities and Social Sciences Communications, 10, Article 138. https://doi.org/10.1057/s4139-04-01633-x

Harvey, L., & Green, D. (1993). Defining quality. Assessment & Evaluation in Higher Education, 18(1), 9–34. https://doi.org/10.1080/0260293930180102

Harvey, L., & Williams, J. (2010). Fifteen years of quality in higher education. Quality in Higher Education, 16(1), 3–36. https://doi.org/10.1080/13538321003679457

Hazelkorn, E. (2015). Rankings and the reshaping of higher education: The battle for world-class excellence (2nd ed.). Palgrave Macmillan. https://doi.org/10.1057/9781137446671

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 21(9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239

Heinze, G., & Schemper, M. (2002). A solution to the problem of separation in logistic regression. Statistics in Medicine, 21(16), 2409–2419. https://doi.org/10.1002/sim.1047

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2004). Roteiro de auto-avaliação institucional: Orientações gerais. INEP/CONAES/SINAES. Retrieved August 3, 2026, from https://download.inep.gov.br/publicacoes/institucionais/avaliacoes_e_exames_da_educacao_superior/roteiro_de_auto_avaliacao_institucional_2004.pdf

International Network for Quality Assurance Agencies in Higher Education. (2016). Guidelines of good practice.

Kalita, E., Oyelere, S. S., Gaftandzhieva, S., Rajesh, K. N. V. P. S., Jagatheesaperumal, S. K., Mohamed, A., Elbarawy, Y. M., Desuky, A. S., Hussain, S., Cifci, M. A., Theodorou, P., Hilčenko, S., Hazarika, J., & Ali, T. (2025). Educational data mining: A 10-year review. Discover Computing, 28, Article 81. https://doi.org/10.1007/s1139-04-09589-z

Khalil, M., Prinsloo, P., & Slade, S. (2023). Fairness, trust, transparency, equity, and responsibility in learning analytics. Journal of Learning Analytics, 10(1), 1–7. https://doi.org/10.18608/jla.2023.7983

King, G., & Zeng, L. (2001). Logistic regression in rare events data. Political Analysis, 9(2), 137–163. https://doi.org/10.1093/oxfordjournals.pan.a004868

Komljenovic, J., Sellar, S., & Birch, K. (2025). Turning universities into data-driven organisations: Seven dimensions of change. Higher Education, 89, 1369–1386. https://doi.org/10.1007/s1139-04-01277-z

Lemaitre, M. J. (2017). Quality assurance in Latin America: Current situation and future challenges. Tuning Journal for Higher Education, 5(1), 21–40. https://doi.org/10.18543/tjhe-5(1)-2017pp139-04

Martin, M., & Stella, A. (2007). External quality assurance in higher education: Making choices. UNESCO International Institute for Educational Planning.

Pham, N., Pham Ngoc, H., & Nguyen-Duc, A. (2025). Fairness for machine learning software in education: A systematic mapping study. Journal of Systems and Software, 219, Article 112244. https://doi.org/10.1016/j.jss.2024.112244

Rifo, M., & Durán, A. (2023). Serving two masters: First and second-generation quality assurance reforms in Latin America. Journal of Higher Education Policy and Leadership Studies, 4(2), 74–90. https://doi.org/10.61186/johepal.4.2.74

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLOS ONE, 10(3), e0118432. https://doi.org/10.1371/journal.pone.0118432

Salto, D. J. (2018). Quality assurance through accreditation: When resistance meets over-compliance. Higher Education Quarterly, 72(2), 78–89. https://doi.org/10.1111/hequ.12151

Sarrico, C. S. (2022). Quality management, performance measurement and indicators in higher education institutions: Between burden, inspiration and innovation. Quality in Higher Education, 28(1), 11–28. https://doi.org/10.1080/13538322.2021.1951445

Secretariat of Higher Education, Science, Technology and Innovation. (n.d.-a). Base de datos de artículos publicados por universidades y escuelas politécnicas en revistas indexadas, 2015–2023 [Data set; source file: Base_estadistica_articulos_UEP_15_23.xlsx]. SENESCYT/SIIES. Retrieved July 23, 2026, from https://www.datosabiertos.gob.ec/dataset/base-de-datos-de-articulos-publicados-uep-en-revistas-indexadas

Secretariat of Higher Education, Science, Technology and Innovation. (n.d.-b). Base estadística de docentes de universidades y escuelas politécnicas, 2015–2022 [Data set; source file: Reporte_estadistico_docentes_UEP_139-04.xlsx]. SENESCYT/SIIES. Retrieved July 23, 2026, from https://www.datosabiertos.gob.ec/dataset/base-de-datos-docentes-uep-ano-2022

Secretariat of Higher Education, Science, Technology and Innovation. (n.d.-c). Base estadística de matrícula de universidades y escuelas politécnicas, 2015–2023 [Data set; source file: Base_estadistica_matricula_UEP_15_23.xlsx]. SENESCYT/SIIES. Retrieved July 23, 2026, from https://www.datosabiertos.gob.ec/dataset/base-de-datos-de-matricula-de-uep-ano-2021

Sharvashidze, G., Grdzelidze, I., Sikharulidze, D., & Gabrichidze, T. (2023). Data-driven analysis for evidence-based decision making at universities: Benchmarking, software tools and weighted indicators in quality assurance—Analytical framework of Tbilisi State University. Quality in Higher Education, 29(1), 42–59. https://doi.org/10.1080/13538322.2022.2122107

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. EDUCAUSE Review, 46(5), 30–32.

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. American Behavioral Scientist, 57(10), 1510–1529. https://doi.org/10.1177/0002764213479366

Stensaker, B. (2008). Outcomes of quality assurance: A discussion of knowledge, methodology and validity. Quality in Higher Education, 14(1), 3–13. https://doi.org/10.1080/13538320802011532

Viberg, O., Mutimukwe, C., & Grönlund, Å. (2022). Privacy in learning-analytics research: Understanding the field to improve the practice. Journal of Learning Analytics, 9(1), 169–182. https://doi.org/10.18608/jla.2022.7751