International Journal of Literature and Arts
2026, Vol. 14, No. 5, pp. 120–134
https://doi.org/10.11648/j.ijla.20261405.12

Research Article

Assessing Information Literacy and Source-Based English Academic Writing Among Nigerian Undergraduates: Instruments and Analytical Framework

Basilia Igbokwe[1], **Nkechi Priscilla Nwanekezie**[2,*]

[1]Department of Humanities, Federal Polytechnic Nekede, Owerri, Nigeria
[2]University Library, Kingsley Ozumba Mbadiwe University, Ideato, Nigeria

Abstract

Undergraduates in Nigerian universities are expected to write essays and research papers that draw on scholarly sources, and their difficulties with finding, evaluating, integrating and acknowledging sources are often attributed to weak information literacy. Testing that explanation requires suitable measures of both information literacy and source-based writing, and a design that avoids artificial associations between them. Many studies measure both constructs by self-report, which inflates their association through shared method, and use writing criteria that overlap in content with information literacy items. This methodological paper proposes Version 1.0 instruments and an analytical framework for examining the relationship between self-reported information literacy and English academic writing among undergraduates in Nigerian federal universities. It reports no student data, and the statistical procedures it describes are proposed procedures, not results. The paper presents a 35-item Self-Reported Information Literacy Scale with five dimensions derived from the standards and frames of the Association of College and Research Libraries, five of them reverse-keyed; a six-dimension analytic rubric for source-based argumentative essays, adapted from Weigle's analytic scale and extended with source use and integration and critical thinking and analysis, with descriptors at all five levels; a rater training, adjudication and agreement procedure; and a preliminary audit of conceptual overlap between the scale and the rubric, with a protocol for independent reassessment. The framework specifies a supervised common writing task based on a closed source pack, stratified sampling, sample size requirements that allow for clustering, an ordinal confirmatory factor analysis, three hypotheses with prespecified decision rules, cluster-robust inference and sensitivity analyses that remove overlapping content. The scale and rubric are an initial item pool and scoring guide for subsequent validation; expert review, piloting and psychometric testing have not yet been carried out, and no reliability or validity is claimed for them.

Keywords

Information Literacy, Academic Writing, Analytic Rubric, Instrument Design, Source-based Writing, Nigeria

*Correspondence: Nkechi Priscilla Nwanekezie (nkechi.nwanekezie@komu.edu.ng)

Received: 21 October 2026; **Accepted:** 31 October 2026; **Published:** 16 November 2026

Copyright: © The Author(s), 2026. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Undergraduates in Nigerian universities write essays, term papers and research projects in English, which is the medium of instruction and, for most of them, a second language [1, 2]. Increasingly, these assignments ask students to use scholarly sources, and students are expected to locate, evaluate, combine and cite them. Evidence from Nigerian students, however, indicates that these tasks are challenging. Final-year undergraduates of library and information science rated themselves weakest in paraphrasing and in applying citation styles [3], and Nigerian postgraduates studying in the United Kingdom often had only a partial understanding of plagiarism [4]. So far, information literacy teaching in Nigerian universities has been offered mainly as library orientation and courses on the use of the library, and not within teaching in the disciplines [5].

Information literacy has long been defined as the ability to recognise when information is needed and to locate, evaluate and use it effectively [6]. The Framework for Information Literacy for Higher Education of the Association of College and Research Libraries (ACRL) now presents it as a set of interconnected frames that link information practices to inquiry, scholarly conversation and the ethical use of information [7]. Because many of these practices are exercised in writing from sources, information literacy is widely expected to be associated with the quality of students' source-based writing [8, 9]. Some recent conceptual studies have proposed incorporating information literacy into English language teaching and have advocated cooperation between Nigerian university libraries and English departments [10, 11].

But testing the association between information literacy and academic writing is not as easy as it sounds. Three measurement problems are common. First, when both constructs are measured by self-report questionnaires, part of their association reflects shared method, such as a general tendency to rate oneself favourably [12]. Second, writing criteria such as source integration and citation overlap in content with information literacy items about evaluating and acknowledging sources, so that the two measures partly assess the same thing. Third, where the writing evidence consists of assignments set by various lecturers in various courses, differences in tasks, topics and access to sources mask differences between the students. Further problems arise in design and analysis: students are clustered in departments and universities, analyses of many dimensions inflate the risk of chance findings, and sample sizes are often set by formulas that ignore the analyses to be performed.

This is a methodological paper. It proposes Version 1.0 of two instruments and an analytical framework, and it reports no data from students; the statistical procedures it describes are proposed, not completed. It has three aims: first, to present a proposed Self-Reported Information Literacy Scale (SRILS) and an analytic rubric for rating source-based argumentative essays; second, to specify a procedure for auditing and controlling the conceptual overlap between them; and third, to set out the design, sample size, validation procedure and analytical decision rules needed to test whether self-reported information literacy is associated with rated English academic writing. The instruments and the framework are provided for use by researchers, libraries and English departments, and their empirical validation is identified as the first step towards using them.

2. Literature Review

2.1. Information Literacy and Its Assessment

Conceptions of information literacy have shifted from a set of skills to a more holistic understanding, incorporating metacognition, critical evaluation and ethical participation in communities of learning [7, 13]. The earlier ACRL standards described five competencies: determining the information needed, accessing it, evaluating it, using it to accomplish a purpose, and using it ethically and legally [14]. The current ACRL Framework replaces these with six frames: Authority Is Constructed and Contextual; Information Creation as a Process; Information Has Value; Research as Inquiry; Scholarship as Conversation; and Searching as Strategic Exploration [7]. Knowledge practices and dispositions are integrated in each frame, which makes information literacy challenging to capture in a single test.

Assessment approaches differ in what they capture. Self-report scales record students' confidence and reported practices; they are efficient but measure perceived rather than demonstrated competence. Performance assessments examine what students actually do. Oakleaf [15] showed that rubrics can be used to assess information literacy in student work and that training improves agreement among raters. Salisbury and Karasmanis [16] surveyed first-year students and found that they could build basic searches but struggled to identify scholarly sources and to cite them correctly, and Pinto, Fernández-Pascual and Sales [17] developed a synthetic indicator of performance in the communication of information among social sciences students. Head and Eisenberg [18] found that the challenges students faced in research centred on finding, evaluating and selecting sources. The choice between self-report and performance measures therefore determines what kind of claim a study can make.

2.2. Assessing Source-based Academic Writing

Academic writing in a second language involves argument, organisation, accuracy and the conventions of a discipline, together with the use of sources [19]. Weigle [20] distinguished holistic scoring, which assigns a single overall score, from analytic scoring, which rates separate features of a text and gives more diagnostic information; analytic scales conventionally address content, organisation, vocabulary, language use and

mechanics. A review by Jonsson and Svingby [21] found that the reliability of rubric scoring improves with analytic criteria, clear descriptors, examples of performance at each level and rater training.

Conventional analytic scales, however, do not directly attend to the use of sources. Research on second-language writing shows that students often patchwrite, reproducing the wording of sources with small changes, and that this reflects difficulties in reading and in learning to write from sources as much as intent to deceive [22-24]. Howard, Serviss and Rodrigue [25] found that students worked mainly from isolated sentences rather than from complete texts. A rubric for source-based writing therefore needs criteria for how sources are represented, integrated and acknowledged, and for the critical analysis that distinguishes argument from summary.

2.3. Information Literacy and Writing

Elmborg [8] argued that information literacy and writing instruction share the goals of critical thinking and meaning making, and Grafstein [9] proposed situating information literacy within the disciplines in which students write. Students writing from sources must locate them, judge their credibility, combine their ideas and acknowledge them, activities that correspond closely to information literacy practices. In Nigeria, the formal register of academic English differs considerably from the English that students use in everyday life [26], which adds a further demand when reading sources, and explicit instruction in reading strategies has been proposed to give students greater control over difficult texts [27].

These arguments make an association between information literacy and writing plausible, and they also show how easily it can be exaggerated. If information literacy items ask students whether they paraphrase and cite correctly, and the writing rubric rates paraphrasing and citation, a correlation between the two is partly guaranteed by their content. A credible test therefore requires measures obtained by different methods, an explicit audit of overlapping content and analyses that show whether the association survives when overlapping elements are removed.

3. Methods of Development

The instruments and framework were developed in five steps.

Literature synthesis. The literature on information literacy and its assessment, second-language academic writing, rubric scoring and the relationship between information literacy and writing was reviewed narratively rather than systematically, using Google Scholar and publisher and journal websites and following the reference lists of key works.

Specification of threats. The measurement and design problems described in the Introduction were listed, and a feature of the instruments or the framework was specified for each: separation of methods, a common writing task, blind rating, an overlap audit, cluster-aware analysis and sample sizes based on the planned tests.

Item drafting. The five dimensions of the scale were taken from the five ACRL competency standards [14], because these describe practices that students can report on, whereas several of the current frames are threshold concepts and dispositions that are difficult to rate by self-report [7]. For each dimension, the authors drafted statements describing knowledge practices that correspond to the relevant standard and frames, and assigned each statement a primary frame (Appendix II). Seven items per dimension were retained, a number chosen to cover each dimension while keeping the scale short enough for class administration, and to allow weak items to be removed after piloting while leaving at least four items per dimension. No larger initial pool was retained as a separate document. In successive drafts the authors reworded items that described institutional access rather than personal ability, reworded one item in each dimension negatively, and removed duplication between dimensions.

Rubric drafting. The rubric was adapted from Weigle's analytic scale [20], as described in Section 5.1, and descriptors were written for all five levels (Table 2 and Appendix III).

Overlap audit. Each scale item was compared with each rubric dimension by the authors to identify conceptual overlap (Section 6). This audit is preliminary. No independent expert review has yet taken place, and no data were collected from students in preparing this paper. The questionnaire and rubric are presented as Version 1.0 (October 2026).

4. The Self-reported Information Literacy Scale

4.1. Structure

The SRILS is a 35-item self-report scale with five dimensions of seven items each, answered on a five-point scale from 1 (strongly disagree) to 5 (strongly agree). Table 1 shows the dimensions, the corresponding ACRL standard and frames and an example item; all items are given in Appendix I, and each item's primary frame in Appendix II. One item in each dimension is negatively worded and scored in reverse, so that a wording effect can be examined; reverse-keyed items do not by themselves guarantee that acquiescence can be identified or removed. Scores for each dimension are the means of their seven items, and the use of an overall score depends on the evidence described in Section 7.4. One instructed-response item asks students to select a particular answer and serves as an attention check.

Table 1. *Dimensions of the Self-Reported Information Literacy Scale.*

Dimension	Items	ACRL standard and main frames	Example item
Identifying information needs	1–7	Standard 1; Research as Inquiry	I can turn a broad assignment topic into specific questions.
Locating and accessing information	8–14	Standard 2; Searching as Strategic Exploration	I know how to search electronic databases for scholarly articles.
Evaluating information	15–21	Standard 3; Authority Is Constructed and Contextual; Information Creation as a Process	I consider the purpose or possible bias of a source.
Using information effectively	22–28	Standard 4; Scholarship as Conversation	I combine information from several sources to reach a conclusion.
Ethical use of information	29–35	Standard 5; Information Has Value	I acknowledge the sources of ideas as well as of direct quotations.

Note: Standards refer to the ACRL competency standards; the correspondence with the frames is approximate, and several frames bear on more than one dimension.

4.2. What the Scale Measures

The SRILS measures self-perceived information literacy: students' reported confidence and practices, not their demonstrated competence. This choice makes the scale efficient in large samples, and it has three consequences for interpretation. First, an association between SRILS scores and rated writing shows that students who report stronger information literacy tend to write better, not that their information literacy has been observed. Second, self-report responses are affected by social desirability and response styles. The reverse-keyed items allow a possible wording effect to be examined, and the framework recommends adding the 13-item short form of the Marlowe–Crowne Social Desirability Scale [28] so that its association with SRILS scores can be reported and adjusted for. Third, reported practices depend partly on the resources available; a student without database access cannot report using databases. Items about locating information are therefore worded as knowledge of how to perform a task (for example, "I know how to search electronic databases") rather than as frequency of use, and a separate four-item resource access checklist (Appendix I) records internet access, access to electronic databases on and off campus and library opening hours. It enters the models as a single covariate, the number of resources available (0 to 4), and the four indicators are entered separately in a sensitivity analysis. Studies that require evidence of demonstrated information literacy should add a performance measure, such as a rubric-scored annotated bibliography [15].

5. The Analytic Writing Rubric and Writing Task

5.1. Rubric

The rubric rates six dimensions of a source-based argumentative essay on a scale from 1 to 5. Table 2 gives the descriptors for levels 1, 3 and 5, and Appendix III gives the complete rubric with descriptors for all five levels in one table. Three dimensions are retained from Weigle's analytic scale [20]: content, here content and argument development; organisation, here organisation and coherence; and language use, here language use and accuracy, which also covers vocabulary. Mechanics is included within academic conventions. Two dimensions are added for source-based writing: source use and integration, informed by research on patchwriting and textual borrowing [22, 24, 25], and critical thinking and analysis.

The writing composite is the unweighted mean of the six dimension scores. Equal weighting is used because no dimension has a stronger theoretical claim than the others to represent source-based writing, and because equal weights are transparent and easy to replicate. As a sensitivity analysis, the composite is recomputed with weights from a one-factor confirmatory model of the six dimensions. Because such weights are estimated in the same sample and may not generalise, the unweighted composite remains the primary outcome, and dimension-level results are reported alongside it.

Table 2. Analytic rubric for source-based argumentative essays: descriptors for levels 1, 3 and 5.

Dimension	Level 1	Level 3	Level 5
Content and argument development	No clear claim; ideas listed without development	Claim stated and partly developed; some points unsupported	Clear, arguable claim developed throughout with well-chosen evidence
Organisation and coherence	No discernible structure; paragraphs lack focus	Recognisable structure with lapses in paragraphing or transitions	Logical progression; unified paragraphs; effective transitions
Source use and integration	Sources absent, copied or inserted without connection; extensive patchwriting	Required sources used and mostly paraphrased; limited synthesis	Sources accurately summarised or paraphrased, synthesised and used to advance the argument
Language use and accuracy	Frequent errors that obscure meaning	Errors noticeable but meaning clear; limited range	Accurate, varied language in an appropriate academic register
Academic conventions	Citations and references missing or unusable	Citations and references present with inconsistencies	Accurate, consistent citation and referencing in a recognised style; formatting conventions observed
Critical thinking and analysis	Description or summary only	Some evaluation of sources or ideas	Evaluates evidence, compares positions and addresses counterarguments

Note: The complete rubric, including levels 2 and 4, is given in Appendix III.

5.2. Writing Task

To make essays comparable, the framework specifies a common writing task based on a closed source pack. All participants write an argumentative essay of 800 to 1,000 words on a prompt set by the research team, using only a supplied pack of five academic sources that present different positions on a question of general interest. External sources are not permitted, so that every student works with the same material; the ability to locate sources is measured by the SRILS rather than by the essay. The task is completed in two supervised sessions of two hours each, one for reading and planning and one for writing, so that external assistance, including generative artificial intelligence, can be excluded. Students may use a printed dictionary. The source pack is assembled by the research team from open access sources, abridged to excerpts of similar length (about 800 to 1,200 words each, or 4,000 to 6,000 words in total) and comparable difficulty, so that the reading can be completed within the first session; the feasibility of this length for the intended students is checked in the pilot. If more than 20% of pilot students cannot finish the reading within the first session, the excerpts are shortened to about 600 to 800 words each or the pack is reduced to four sources, and the revised pack is piloted again. The pack is reviewed by the content validation panel for suitability, balance and reading demand. Because the pack is printed, the task does not depend on a university's database subscriptions.

5.3. Rater Training, Adjudication and Agreement

A rater manual, containing the rubric, decision rules for borderline cases and benchmark essays at each level of each dimension, is prepared from essays written during piloting. Two raters with experience of teaching academic writing are trained in two sessions with the manual and practise until their scores fall within one point of each other on at least 80% of dimension ratings. Essays are anonymised before rating, and raters are blind to students' SRILS scores, names and universities. Each essay is rated independently by both raters. If the two raters differ by no more than one point on a dimension, the operational score is the mean of their ratings. If they differ by more than one point, the dimension is referred to a consensus review: the two raters and an experienced adjudicator examine the essay against the written rubric and the rater manual, and the adjudicator records the final score, with a brief reason. Because the adjudicator decides, ties do not arise. All original ratings are retained, and the proportion of dimension scores referred to consensus review is reported. Differences in severity between the raters are examined by comparing their mean scores on each dimension. To detect drift, a random 10% of essays are rated again by both raters at the end of the rating period.

Inter-rater reliability is reported from the original ratings before consensus review, and the final operational scores are used only in the substantive analyses. Agreement is reported

in two ways. For the writing composite, an intraclass correlation coefficient from a two-way random-effects model with absolute agreement is reported with its 95% confidence interval, for single raters and for the mean of two raters [29, 30]; values above .75 are regarded as good and above .90 as excellent [30]. For each dimension separately, quadratic-weighted kappa is reported [31]. Pearson correlations between raters are not used as indices of agreement, because they ignore systematic differences in severity.

6. Controlling Construct Overlap

Table 3. *Preliminary classification of overlap between scale items and rubric dimensions.*

Items	Practice described	Rubric dimension	Category
22, 24, 25, 28	Paraphrasing, combining and summarising sources; explaining how a source supports a point	Source use and integration	Substantial
30, 31, 33, 34	Acknowledging ideas, citing in a referencing style, avoiding unacknowledged copying, knowing what needs citation	Academic conventions; source use and integration	Substantial
18, 21	Comparing sources; checking whether evidence supports claims	Critical thinking and analysis	Partial
23	Organising information from sources around main ideas	Organisation and coherence; content and argument development	Partial
27	Selecting information relevant to a question	Content and argument development	Partial
All other items	Identifying needs, searching, judging credibility, recording sources, understanding rules	None	None

Note: No item duplicates the wording of a rubric descriptor. The classification is to be reassessed independently as described in the text.

Some scale items and rubric dimensions refer to related practices. Overlap is classified in three categories: *substantial*, when an item describes the same practice that a rubric dimension rates in the essay, such as citing sources; *partial*, when an item describes a related practice that contributes to, but is not the same as, what the dimension rates; and *none*. Table 3 records the authors' preliminary classification of all 35 items against the six rubric dimensions; it is labelled preliminary because it has not been independently replicated.

Independent reassessment. During content validation, two judges who are not authors classify each of the 210 item-by-dimension pairs (35 items by 6 dimensions) independently, using the category definitions above. Agreement is reported with Cohen's weighted kappa [31], disagreements are resolved by a third judge, and the final classification replaces Table 3.

Control of overlap. Overlap is controlled in three ways. First, the methods differ: the SRILS is a self-report questionnaire, whereas writing is rated from essays by raters blind to SRILS scores. This separation reduces shared-method bias, although it does not eliminate all correlated error, confounding or effects of the way students respond. Second, overlapping content is removed from both measures together. An overlap-free writing composite, the mean of content and argument development, organisation and coherence, and language use and accuracy, is combined first with a reduced SRILS score that excludes items classified as substantial, and then with a further reduced score that also excludes items classified as partial. These two analyses form the test of H3 (Section 7.6); analyses that remove overlap from only one of the two measures are reported descriptively, to show where any reduction in the association arises. Third, the interpretation of the primary hypothesis depends on whether the association survives the removal of overlap.

7. Analytical Framework

7.1. Hypotheses

The framework yields three hypotheses.

H1. Self-reported information literacy is positively associated with the rubric-rated writing composite.

H2. The five dimensions of self-reported information literacy together explain variance in the rubric-rated writing composite beyond that explained by year level, discipline, prior English attainment, prior writing instruction, resource access

and university.

H3. The association in H1 remains positive when overlapping content is removed.

H1 and H3 refer to an overall score of self-reported information literacy. If the evidence described in Section 7.4 does not support an overall score, H1 and H3 cannot be tested as written; they are then reported as not testable, and any analyses of individual dimensions are labelled exploratory rather than substituted for them. H2 does not depend on an overall score. The relative strength of individual dimensions, and associations at the level of individual rubric dimensions, are treated as exploratory. In this paper the hypotheses are theoretical propositions; they become confirmatory only when a study applying the framework registers them, with its analysis plan, before data collection.

7.2. Design and Sampling

The framework specifies a cross-sectional correlational design, from which associations but not causal effects can be estimated, and is designed for federal universities in Nigeria; application to state or private universities may require adaptation. At least four federal universities from at least two geopolitical zones are recommended. Within each university, students at 200, 300 and 400 level are sampled by stratified random selection from enrolment lists, with year level as the stratum and allocation proportional to enrolment, so that students who rarely use the library are as likely to be selected as frequent users. Discipline is recorded in four categories: arts and humanities, social sciences, sciences and education. First-year students are excluded because most have not yet completed the general studies course in academic writing and have little experience of source-based assignments. The population figures used for allocation should be the documented enrolment of the eligible levels in each university. The numbers selected, contacted, consenting, completing the questionnaire, completing the essay and retained for analysis should be reported for each university and level.

7.3. Sample Size

The sample size is set by the planned analyses. All calculations use the noncentral F distribution for the G*Power test "Linear multiple regression: fixed model, R^2 increase" [32], with $f^2 = \Delta R^2/(1 - R^2\text{full})$, $\alpha = .05$ and power = .80, and 16 predictors in the full model (11 control terms, comprising two for year level, three for discipline, three for university and one each for prior English attainment, prior writing instruction and resource access, and five scale dimensions). Detecting a small unique contribution of one dimension ($f^2 = .02$; one tested predictor, numerator df = 1) requires 395 participants (denominator df = 378), and detecting a small to moderate joint contribution of the five dimensions ($f^2 = .035$; numerator df = 5) requires 373 (denominator df = 356). Because f^2 is defined relative to the unexplained variance, these effect sizes correspond to different increments in R^2 depending on the variance explained by the full model; if the full model explains 30% of the variance, $f^2 = .02$ corresponds to an increment of 0.014 and $f^2 = .035$ to an increment of 0.025. For H1, about 194 participants give 80% power to detect a correlation of .20 (two tailed, Fisher z approximation).

Students sampled individually are clustered in departments and universities. If about 24 students come from each department and the department-level intraclass correlation of writing scores is .02, the design effect is $1 + 23 \times .02 = 1.46$. These figures approximate the effect of clustering by dividing the sample by the design effect and recomputing power from the noncentral F distribution for the resulting effective sample; correlated clusters were not simulated. With 480 valid cases the effective sample would be about 329, giving power of about .72 for a single dimension and .74 for the joint test, and about .95 for H1. Retaining power of .80 for the joint test under this design effect requires about 545 valid cases; with 600 valid cases the effective sample is about 411, giving power of about .82 for a single dimension and .84 for the joint test. The framework therefore recommends a target of 600 valid cases, distinguishing this from the number recruited: allowing for 20% loss through incomplete questionnaires, absence from the writing sessions or unusable essays, about 750 students should be recruited. The calculations are approximate and provisional, because the true intraclass correlation is unknown; it should be estimated and reported, and the calculations repeated with the observed value. The settings used are summarised in Appendix IV so that the figures can be reproduced.

7.4. Validation of the Instruments

The SRILS and the rubric presented in Sections 4 and 5 are an initial item pool and scoring guide: they have not yet undergone expert review, piloting or psychometric testing, and the following steps must be completed before either is used to test the hypotheses.

Content validation. A panel of at least five experts, including specialists in library and information science and in English language education, rates the relevance of each item and rubric descriptor on a four-point scale. Items with an item-level content validity index below .78 are revised or removed [33]; with five experts, this means that an item is retained only if at least four of the five rate it as relevant (an index of .80). The scale-level index, averaged across items, should reach at least .90 [33]. The panel also reviews the source pack and the item-to-frame mapping.

Piloting. The questionnaire and writing task are piloted with about 60 students outside the main sample. Cognitive interviews with a small number of students check how items are understood, the essays are used to prepare the rater manual and benchmark essays, and a subset of essays is rated by two raters to check the rubric before the main study.

Psychometric analysis. In the main sample, a confirmatory

factor analysis tests the five-dimension structure of the SRILS. Because the items are ordinal with five categories, the primary estimator is WLSMV, which treats responses as ordered categories. Departmental clustering is accommodated in the primary analysis by estimating the model in Mplus with the complex survey option (TYPE = COMPLEX), with department as the cluster variable and university as the stratum, which adjusts standard errors and the chi-square test of fit for clustering [34]. With four universities as strata, each contributing several departments, this specification is implementable provided that every stratum contains at least two departments. Robust maximum likelihood, which performs well with five categories when distributions are not strongly skewed [35], is used as a sensitivity analysis in R with the lavaan package [36], again with standard errors clustered by department. Fit is judged by the comparative fit index (CFI ≥ .95 good, ≥ .90 acceptable), the root mean square error of approximation (≤ .06 good, ≤ .08 acceptable) and the standardised root mean square residual (≤ .08) [37, 38]. These values are used as guidelines rather than as decisive evidence of validity, and the decision about the structure also considers the size of loadings, residual correlations and the interpretability of the factors. Because negatively worded items can form a method factor of their own, the five-dimension model is compared with a model that adds a method factor for the five reverse-keyed items, and the effect of this factor on loadings and fit is reported. Reliability is estimated with Cronbach's alpha and McDonald's omega for each dimension, and measurement invariance between female and male students is tested by comparing configural, metric and scalar models [39].

Overall score. Failure of the five-dimension model does not justify a single total score. An overall score is used only if a bifactor or higher-order model fits acceptably and the general factor accounts for most of the reliable variance, judged by hierarchical omega and explained common variance [40], with values of at least .70 and .60 respectively adopted here as working criteria. If these conditions are not met, no overall score is reported, H1 and H3 are reported as not testable as written, and associations of the individual dimensions with writing are reported only as an exploratory family with Holm correction.

7.5. Missing Data

The primary analysis uses students with both a questionnaire and an essay. Students with only one of these are compared with complete cases on sex, age, year level, discipline and university, using tests with the correct degrees of freedom for the groups compared. In the primary analysis, a scale score is computed from the items answered if at least five of the seven items in a dimension are answered, and is otherwise treated as missing; this keeps the primary inference, including the CR2 confidence intervals and the HTZ joint test, within a single dataset. The confirmatory factor analysis uses the observed ordinal responses. Under WLSMV, missing responses are handled by pairwise deletion, which gives consistent estimates when responses are missing completely at random but not in general when they are missing at random. Multiple imputation is therefore used as a sensitivity analysis, under the assumption that data are missing at random given the observed variables [41]. Twenty imputed datasets are created in R with the mice package [42], using proportional odds models for the ordinal scale items and predictive mean matching for continuous variables; the imputation model includes all scale items, the writing dimension scores, the controls, resource access and university and department indicators, the last as fixed indicators to reflect the cluster structure. The confirmatory factor analysis is repeated in each imputed dataset and the estimates pooled. In the regressions, each imputed dataset is analysed with CR2 standard errors, and coefficients and their CR2 variances are pooled with Rubin's rules. Because there is no established method for pooling the HTZ joint test across imputations, the joint test is reported for each imputed dataset, with the median and range of its p values, rather than as a single pooled test. In a further sensitivity analysis, missing essays are imputed for students who completed the questionnaire, using the same variables and the prior English grade as auxiliary information. Little's test is reported [43], with the caution that it can detect departures from data missing completely at random but cannot establish that data are missing at random.

7.6. Analysis and Decision Rules

Clustering. University is entered as a fixed effect, because there are too few universities for multilevel modelling [44], and standard errors are clustered by department with the CR2 small-sample correction [45]. Ordinary standard errors are not used for inference.

H1. The association is estimated as the standardised coefficient from a regression of the writing composite on the SRILS score with university fixed effects, which in this form approximates a partial correlation within universities, with a 95% confidence interval from CR2 standard errors. H1 is supported if the coefficient is positive and the interval excludes zero. The unadjusted Pearson correlation is reported for description.

H2. A two-step regression is estimated. Step 1 includes the controls; Step 2 adds the five scale dimensions. The inferential test of H2 is a cluster-robust Wald test of the five added coefficients, using the CR2 variance estimator with the HTZ small-sample correction [45]; H2 is supported if this test is significant at $\alpha = .05$. The ordinary change in R^2 is reported only as a descriptive effect size, together with each coefficient and its CR2 confidence interval. Variance inflation factors, residual plots and influential cases are examined and reported.

H3. The H1 model is reestimated with overlapping content removed from both measures at the same time: first with the overlap-free writing composite and the SRILS score excluding items classified as substantial, and second with the overlap-free writing composite and the score that also excludes items classified as partial. H3 is supported if both coefficients

are positive with intervals excluding zero. Analyses that remove overlapping content from only one of the two measures are reported descriptively, to show where any reduction in the association arises. If the association is substantially weaker when overlap is removed, the main association is interpreted as partly reflecting overlapping content.

Exploratory analyses and effect sizes. Associations between each scale dimension and each rubric dimension form an exploratory family of 30 comparisons, with p values adjusted by the Holm procedure. Correlations of about .10, .30 and .50 are described as small, medium and large [46], as a convention rather than a judgement of practical importance.

Table 4. *Hypotheses, analyses and decision criteria.*

Hyp.	Association tested	Analysis	Criterion for support
H1	Self-reported information literacy and writing composite	Standardised coefficient with university fixed effects; CR2 confidence interval	Positive coefficient; CI excludes zero
H2	Five dimensions of self-reported information literacy and writing composite, beyond controls	Two-step regression; cluster-robust Wald test (CR2, HTZ) of the five added coefficients	Wald test significant at $\alpha = .05$
H3	H1 with overlapping content removed from both measures	H1 model with the overlap-free composite combined with each of the two reduced scale scores	Both coefficients positive; CIs exclude zero

Note: H1 and H3 require an overall score. If the evidence for an overall score is insufficient (Section 7.4), H1 and H3 are reported as not testable, and analyses of individual dimensions are exploratory only. All associations are cross-sectional. CI: confidence interval.

Reporting. Studies applying the framework should follow the STROBE guidelines for cross-sectional studies [47], report the flow of participants, the rater agreement statistics and the proportion of adjudicated scores, and emphasise effect sizes and confidence intervals over significance tests. Reverse causality and confounding by general academic ability, reading proficiency and motivation are plausible and should be acknowledged.

7.7. Ethical Safeguards

Ethical approval should be obtained from the research ethics committee of each participating university before data collection. Students give informed consent to complete the questionnaire and to the research use of their essays. Because the writing sessions take place in scheduled course time, participation must be clearly voluntary: the essay is not graded and does not count towards any course, students who decline are offered an alternative activity, and lecturers are not told who took part. Essays are anonymised and linked to questionnaires by a study code, and the key linking codes to names is stored separately and destroyed after linkage.

7.8. Feasibility

The framework is designed to be feasible in Nigerian university settings where resources are limited. The printed source pack removes dependence on database subscriptions, and the questionnaire takes about 15 minutes. The main demand is rating. Because each of the two raters independently scores every essay, at about 15 minutes per essay each rater needs about 150 hours for 600 essays. Training, consensus review of discrepant scores and the 10% repeat ratings add further hours. The load can be shared by recruiting raters from English departments in all the participating universities.

8. Discussion

8.1. Contributions

The paper has three contributions. First, it provides an initial item pool for measuring self-reported information literacy in five dimensions derived from the ACRL standards and mapped to the current frames, together with an analytic rubric that extends a conventional analytic scale with criteria for source use and critical analysis and gives descriptors at every level. Second, it makes the conceptual overlap between information literacy and writing measures explicit, through a classification with defined categories and a protocol for independent reassessment, and it specifies how the influence of overlap can be tested. Third, it sets out a design in which the two constructs are measured by different methods, with a supervised common writing task, blind rating, sample sizes that allow for clustering and decision rules fixed in advance.

8.2. Implications

For researchers, the framework shows that the strength of an association between information literacy and writing depends heavily on how the two are measured; an association between two self-report scales with overlapping content says

little about either construct. For libraries and English departments, the rubric offers a shared vocabulary for describing students' source-based writing, and the dimensions of the scale identify areas, such as evaluating sources and acknowledging them, in which collaborative teaching may be planned [11].

8.3. Limitations

The instruments are an initial item pool and scoring guide. They have been reviewed only by the authors and have not been reviewed by independent experts, piloted or administered to students; their structure, reliability and validity remain to be established. Until the content validation, piloting and psychometric analysis in Section 7.4 have been completed, they should not be treated as validated instruments, and no reliability or validity is claimed for them. Recent validation studies of student self-assessment scales for related literacies have followed the same sequence of factor analysis and reliability estimation [48], and the five-dimension structure of the SRILS should likewise be regarded as a hypothesis until it is tested. The SRILS measures self-perceived rather than demonstrated information literacy and may be affected by social desirability and response styles. A supervised task with a closed source pack improves comparability but differs from the extended, open assignments students normally write, and it does not assess the ability to find sources. The rubric criteria for source use and critical thinking require careful rater training. Recent studies have found that raters score source use less consistently in argumentative reading-to-write essays than in summaries [49], and that the amount of source material and how well it supports the writer's claim affect raters' judgements of argumentation [50], so agreement on these dimensions may be lower than on the others. The overlap classification was made by the authors and requires independent confirmation. Sample size calculations depend on an assumed intraclass correlation, and although bias-reduced cluster-robust standard errors generally perform acceptably with few clusters [51], inference with a small number of departments remains approximate. The framework was designed for undergraduates in Nigerian federal universities and may need adaptation for other institutions and countries.

9. Conclusions

This paper has proposed a Self-Reported Information Literacy Scale, an analytic rubric for source-based argumentative essays and an analytical framework for examining the relationship between self-reported information literacy and English academic writing among undergraduates in Nigerian federal universities. The framework separates the methods by which the two constructs are measured, classifies and tests their conceptual overlap, specifies a supervised common writing task and blind rating, and sets sample sizes and decision rules in advance. These tools are offered for validation and use, with the aim of making research on information literacy and academic writing in Nigerian universities more rigorous and more useful to the libraries and English departments that teach them.

Abbreviations

ACRL	Association of College and Research Libraries
CFI	Comparative Fit Index
CI	Confidence Interval
CR2	Bias-Reduced Cluster-Robust Variance Estimator
HTZ	Hotelling's T^2 Approximation with Small-Sample Correction
SRILS	Self-Reported Information Literacy Scale
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
WLSMV	Weighted Least Squares, Mean and Variance Adjusted

Author Contributions

Basilia Igbokwe: Conceptualization, Methodology, Writing – original draft, Writing – review & editing

Nkechi Priscilla Nwanekezie: Conceptualization, Methodology, Writing – original draft, Writing – review & editing

Conflicts of Interest

The authors declare no conflicts of interest.

Appendix

Appendix I: Self-Reported Information Literacy Scale and Resource Access Checklist (Version 1.0)

Please indicate how far you agree with each statement. (1 = strongly disagree; 2 = disagree; 3 = neither agree nor disagree; 4 = agree; 5 = strongly agree). Items marked (R) are scored in reverse.

Identifying information needs. 1. Before I start an assignment, I work out what kind of information I need. 2. I can turn a broad assignment topic into specific questions. 3. I identify the key concepts and terms in a research topic. 4. I usually start searching before deciding what information I need. (R) 5. I recognise when I need more information to support a point. 6. I can distinguish between background information and evidence for an argument. 7. I revise my research question when I learn more about a topic.

Locating and accessing information. 8. I know how to use the library catalogue to find books and other materials. 9. I know how to search electronic databases for scholarly articles.

10. I use synonyms and related terms when a search gives few results. 11. I use search operators such as AND, OR and NOT to focus a search. 12. I know how to obtain the full text of an article through a library or open access sources. 13. I often lose track of sources I have found. (R) 14. I know when and how to ask a librarian for help.

Evaluating information. 15. I check the qualifications or expertise of an author before using a source. 16. I can tell the difference between a peer-reviewed article and other publications. 17. I check whether the information in a source is current enough for my purpose. 18. I compare information from different sources before accepting it. 19. I consider the purpose or possible bias of a source. 20. I accept information from websites without checking who produced it. (R) 21. I check whether the evidence in a source supports its claims.

Using information effectively. 22. I take notes from sources in my own words. 23. I organise information from different sources around my main ideas. 24. I combine information from several sources to reach a conclusion. 25. I can summarise the main argument of an article. 26. I relate new information to what I already know about a topic. 27. I often include information that is not relevant to my question. (R) 28. I can explain how a source supports a point I am making.

Ethical use of information. 29. I understand what counts as plagiarism. 30. I acknowledge the sources of ideas as well as of direct quotations. 31. I know how to cite sources in at least one referencing style. 32. I understand why intellectual property should be respected. 33. I avoid copying text from the internet without acknowledgement. 34. I am often unsure whether information needs a citation. (R) 35. I follow my university's rules on acknowledging the sources and tools I use in assignments.

Attention check. For this statement, please select "disagree". (Not scored.).

Resource access checklist (yes or no). R1. I have regular access to the internet for my studies. R2. I can use my university's electronic databases on campus. R3. I can use my university's electronic databases off campus. R4. The library's opening hours allow me to use it when I need to.

Appendix II: Primary ACRL Frame of Each Scale Item

Table A1. *Items of the Self-Reported Information Literacy Scale by primary ACRL frame.*

Primary frame	Items
Research as Inquiry	1, 2, 3, 4, 5, 6, 7
Searching as Strategic Exploration	8, 9, 10, 11, 12, 13, 14
Authority Is Constructed and Contextual	15, 19, 20, 21
Information Creation as a Process	16, 17
Scholarship as Conversation	18, 22, 23, 24, 25, 26, 27, 28
Information Has Value	29, 30, 31, 32, 33, 34, 35

Note: Each item is assigned to one primary frame; several items also bear on other frames.

Appendix III: Complete Analytic Rubric (Version 1.0)

Table A2. *Descriptors for all five levels of the analytic rubric.*

Dimension	Level	Descriptor
Content and argument development	1	No clear claim; ideas listed without development
	2	A position can be inferred but is not stated clearly; most points are asserted rather than supported
	3	Claim stated and partly developed; some points unsupported
	4	Clear claim developed with relevant evidence, with occasional gaps or weakly supported points
	5	Clear, arguable claim developed throughout with well-chosen evidence
	1	No discernible structure; paragraphs lack focus

Dimension	Level	Descriptor
Organisation and coherence	2	Some grouping of ideas, but paragraphs often lack a controlling idea and links are missing
	3	Recognisable structure with lapses in paragraphing or transitions
	4	Clear structure and focused paragraphs; transitions mostly effective, with occasional abrupt shifts
	5	Logical progression; unified paragraphs; effective transitions
Source use and integration	1	Sources absent, copied or inserted without connection; extensive patchwriting
	2	Sources mentioned but largely quoted or closely reproduced; little connection to the writer's points
	3	Required sources used and mostly paraphrased; limited synthesis
	4	Sources accurately represented and mostly integrated into the argument; synthesis present but uneven
	5	Sources accurately summarised or paraphrased, synthesised and used to advance the argument
Language use and accuracy	1	Frequent errors that obscure meaning
	2	Frequent errors that sometimes obscure meaning; limited vocabulary
	3	Errors noticeable but meaning clear; limited range
	4	Generally accurate language with minor errors; mostly appropriate academic register
	5	Accurate, varied language in an appropriate academic register
Academic conventions	1	Citations and references missing or unusable
	2	Some citations present, but many sources unacknowledged or references incomplete
	3	Citations and references present with inconsistencies
	4	Citations and references accurate and consistent, with minor lapses in style or formatting
	5	Accurate, consistent citation and referencing in a recognised style; formatting conventions observed
Critical thinking and analysis	1	Description or summary only
	2	Mostly summary, with occasional unsupported evaluative comments
	3	Some evaluation of sources or ideas
	4	Evaluates sources and compares positions; counterarguments acknowledged but not fully addressed
	5	Evaluates evidence, compares positions and addresses counterarguments

Note: Each dimension is scored from 1 to 5. The writing composite is the unweighted mean of the six dimension scores.

Appendix IV: Settings for the Sample Size Calculations

Table A3. *Settings used in the sample size calculations.*

Setting	Value
Test family and test	F tests; linear multiple regression, fixed model, R^2 increase (noncentral F distribution, as in G*Power 3.1)
Effect size definition	$f^2 = \Delta R^2/(1 - R^2\text{full})$
Effect sizes	$f^2 = .02$ (one dimension); $f^2 = .035$ (five dimensions jointly)
Significance level and power	$\alpha = .05$; power = .80
Predictors in full model	16 (11 control terms and 5 scale dimensions)
Tested predictors (numerator df)	1 for a single dimension; 5 for the joint test

Required sample (denominator df)	395 (378) for a single dimension; 373 (356) for the joint test
Correlation for H1	ρ = .20, two tailed, Fisher z approximation: about 194
Clustering assumption	About 24 students per department; intraclass correlation .02; design effect 1.46
Conversion for clustering	Effective sample = valid cases ÷ design effect; power recomputed from the noncentral F distribution for the effective sample
Resulting targets	600 valid cases (effective sample about 411); about 750 recruited, allowing 20% loss

References

[1] Federal Republic of Nigeria. *National Policy on Education*. 6th ed. Lagos: Nigerian Educational Research and Development Council; 2013.

[2] Bamgbose, A. *Language and Exclusion: The Consequences of Language Policies in Africa*. Münster: LIT Verlag; 2000.

[3] Okeji, C. C., Ilika, O. M., Baro, E. E. Assessment of information literacy skills: A survey of final year undergraduates of library and information science in Nigerian universities. *Global Knowledge, Memory and Communication*. 2020, 69(6/7), 521–535. https://doi.org/10.1108/GKMC-10-2019-0130

[4] Orim, S.-M., Davies, J. W., Borg, E., Glendinning, I. Exploring Nigerian postgraduate students' experience of plagiarism: A phenomenographic case study. *International Journal for Educational Integrity*. 2013, 9(1). https://doi.org/10.21913/ijei.v9i1.845

[5] Baro, E. E., Zuokemefa, T. Information literacy programmes in Nigeria: A survey of 36 university libraries. *New Library World*. 2011, 112(11/12), 549–565. https://doi.org/10.1108/03074801111190428

[6] American Library Association. *Presidential Committee on Information Literacy: Final Report*. Chicago, IL: American Library Association; 1989. Available from: https://www.ala.org/acrl/publications/whitepapers/presidential

[7] Association of College and Research Libraries. *Framework for Information Literacy for Higher Education*. Chicago, IL: American Library Association; 2016. Available from: https://www.ala.org/acrl/standards/ilframework

[8] Elmborg, J. Critical information literacy: Implications for instructional practice. *The Journal of Academic Librarianship*. 2006, 32(2), 192–199. https://doi.org/10.1016/j.acalib.2005.12.004

[9] Grafstein, A. A discipline-based approach to information literacy. *The Journal of Academic Librarianship*. 2002, 28(4), 197–204. https://doi.org/10.1016/S0099-1333(02)00283-5

[10] Igbokwe, B., Nwanekezie, N. P. A conceptual model for integrating information literacy and task-based language teaching: Rethinking English language instruction in Nigerian secondary schools. *International Journal of English Language, Literature and Translation Studies*. 2026, 13(3), 121–134. http://ijelr.in

[11] Igbokwe, B., Nwanekezie, N. P. Hedging, signposting and stance: A conceptual model of library–English department collaboration for developing metadiscourse competence in Nigerian undergraduate research proposals. *Research Journal of English Language and Literature*. 2026, 14(3), 508–522. rjelal.com

[12] Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., Podsakoff, N. P. Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*. 2003, 88(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

[13] Bruce, C. S. *Informed Learning*. Chicago, IL: Association of College and Research Libraries; 2008.

[14] Association of College and Research Libraries. *Information Literacy Competency Standards for Higher Education*. Chicago, IL: American Library Association; 2000.

[15] Oakleaf, M. Using rubrics to assess information literacy: An examination of methodology and interrater reliability. *Journal of the American Society for Information Science and Technology*. 2009, 60(5), 969–983. https://doi.org/10.1002/asi.21030

[16] Salisbury, F., Karasmanis, S. Are they ready? Exploring student information literacy skills in the transition from secondary to tertiary education. *Australian Academic & Research Libraries*. 2011, 42(1), 43–58.

[17] Pinto, M., Fernández-Pascual, R., Sales, D. Communication of information in the digital age among social sciences students: Uncovering a synthetic indicator of performance. *Aslib Journal of Information Management*. 2018, 70(4), 326–343. https://doi.org/10.1108/AJIM-02-2018-0035

[18] Head, A. J., Eisenberg, M. B. *Truth Be Told: How College Students Evaluate and Use Information in the Digital Age*. Project Information Literacy Progress Report. Seattle, WA: University of Washington Information School; 2010.

[19] Hyland, K. *Second Language Writing*. 2nd ed. Cambridge: Cambridge University Press; 2019.

[20] Weigle, S. C. *Assessing Writing*. Cambridge: Cambridge University Press; 2002.

[21] Jonsson, A., Svingby, G. The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*. 2007, 2(2), 130–144. https://doi.org/10.1016/j.edurev.2007.05.002

[22] Pecorari, D. Good and original: Plagiarism and patchwriting in academic second-language writing. *Journal of Second Language Writing*. 2003, 12(4), 317–345. https://doi.org/10.1016/j.jslw.2003.08.004

[23] Shi, L. Textual borrowing in second-language writing. *Written Communication*. 2004, 21(2), 171–200. https://doi.org/10.1177/0741088303262846

[24] Pecorari, D., Petrić, B. Plagiarism in second-language writing. *Language Teaching*. 2014, 47(3), 269–302. https://doi.org/10.1017/S0261444814000056

[25] Howard, R. M., Serviss, T., Rodrigue, T. K. Writing from sources, writing from sentences. *Writing and Pedagogy*. 2010, 2(2), 177–192. https://doi.org/10.1558/wap.v2i2.177

[26] Jowitt, D. *Nigerian English Usage: An Introduction*. Ikeja: Longman Nigeria; 1991.

[27] Igbokwe, B., Nwanekezie, N. P. Metacognitive reading strategies and English academic comprehension: A conceptual framework for Nigerian LIS education. *Research Journal of English Language and Literature*. 2025, 13(4), 515–526. http://rjelal.com

[28] Reynolds, W. M. Development of reliable and valid short forms of the Marlowe–Crowne Social Desirability Scale. *Journal of Clinical Psychology*. 1982, 38(1), 119–125. https://doi.org/10.1002/1097-4679(198201)38:1<119::AID-JCLP2270380118>3.0.CO;2-I

[29] Shrout, P. E., Fleiss, J. L. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*. 1979, 86(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420

[30] Koo, T. K., Li, M. Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*. 2016, 15(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

[31] Cohen, J. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*. 1968, 70(4), 213–220. https://doi.org/10.1037/h0026256

[32] Faul, F., Erdfelder, E., Buchner, A., Lang, A.-G. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*. 2009, 41(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

[33] Polit, D. F., Beck, C. T., Owen, S. V. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health*. 2007, 30(4), 459–467. https://doi.org/10.1002/nur.20199

[34] Muthén, L. K., Muthén, B. O. *Mplus User's Guide*. 8th ed. Los Angeles, CA: Muthén & Muthén; 1998–2017.

[35] Rhemtulla, M., Brosseau-Liard, P. É., Savalei, V. When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*. 2012, 17(3), 354–373. https://doi.org/10.1037/a0029315

[36] Rosseel, Y. lavaan: An R package for structural equation modeling. *Journal of Statistical Software*. 2012, 48(2), 1–36. https://doi.org/10.18637/jss.v048.i02

[37] Hu, L., Bentler, P. M. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*. 1999, 6(1), 1–55. https://doi.org/10.1080/10705519909540118

[38] Kline, R. B. *Principles and Practice of Structural Equation Modeling*. 4th ed. New York, NY: Guilford Press; 2016.

[39] Cheung, G. W., Rensvold, R. B. Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*. 2002, 9(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

[40] Rodriguez, A., Reise, S. P., Haviland, M. G. Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*. 2016, 21(2), 137–150. https://doi.org/10.1037/met0000045

[41] Enders, C. K. *Applied Missing Data Analysis*. New York, NY: Guilford Press; 2010.

[42] van Buuren, S., Groothuis-Oudshoorn, K. mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*. 2011, 45(3), 1–67. https://doi.org/10.18637/jss.v045.i03

[43] Little, R. J. A. A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*. 1988, 83(404), 1198–1202. https://doi.org/10.1080/01621459.1988.10478722

[44] McNeish, D., Stapleton, L. M. Modeling clustered data with very few clusters. *Multivariate Behavioral Research*. 2016, 51(4), 495–518. https://doi.org/10.1080/00273171.2016.1167008

[45] Pustejovsky, J. E., Tipton, E. Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. *Journal of Business & Economic Statistics*. 2018, 36(4), 672–683. https://doi.org/10.1080/07350015.2016.1247004

[46] Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum; 1988.

[47] von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., Vandenbroucke, J. P. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *The Lancet*. 2007, 370(9596), 1453–1457. https://doi.org/10.1016/S0140-6736(07)61602-X

[48] Kim, J., Hong, L., Yoon, A. University students' self-assessment of data literacy: A validation study. *PLOS ONE*. 2025, 20(4), e0322104. https://doi.org/10.1371/journal.pone.0322104

[49] Xie, Q. Assessing source use: Summary vs. reading-to-write argumentative essay. *Assessing Writing*. 2023, 57, 100755. https://doi.org/10.1016/j.asw.2023.100755

[50] Chuang, P.-L. Do source use features impact raters' judgment of argumentation? An experimental study. *Language Testing*. 2025, 42(1), 48–72. https://doi.org/10.1177/02655322241263629

[51] Zhang, Y., Lai, M. H. C. Evaluating two small-sample corrections for fixed-effects standard errors and inferences in multilevel models with heteroscedastic, unbalanced, clustered data. *Behavior Research Methods*. 2024, 56(6), 5930–5946. https://doi.org/10.3758/s13428-023-02325-9