Research Journal in Advanced Humanities

RJAH

https://doi.org/10.58256/rv53v440

Open Access

Check for updates

RESEARCH ARTICLE

Section: *Literature, Linguistics & Criticism*

Multicultural children's literature and story writing: A needs analysis in East Java

Hendrik Furqon[1,2]*, Sarwiji Suwandi[1] & Raheni Suhita[1]

[1]Faculty of Teacher Training and Education, Universitas Sebelas Maret, Surakarta, Indonesia

[2]Faculty of Teacher Training and Education, Universitas Islam Darul 'Ulum, Lamongan, Indonesia

*Correspondence: hendrikfurqon@student.uns.ac.id

ABSTRACT

Textbook development for prospective primary school teachers requires evidence about both learners' needs and the suitability of narrative models. This study quantitatively reanalyzes two four-category questionnaire distributions from 120 students in the reported East Java setting, integrates documentary lecturer and student needs, and audits a supplementary public corpus. Plot-development difficulty attracted 103 agreements (85.83%; conditional 95% Wilson interval 78.48–90.96%), and characterization difficulty attracted 105 (87.50%; 80.40–92.28%). Both items had a median response of agree; strongly agree responses were less frequent, at 12.50% and 15.00%. Enumeration of all 16 feasible joint tables established that 88–103 students must have reported both difficulties, assuming a common complete-response sample. Across those tables, exact McNemar p-values ranged from .500 to .860, providing no support for prioritizing one item from the small marginal difference. The separate corpus retained 163 books and 49,567 body tokens. Median within-book lexical diversity decreased by 0.0495 after body delimitation, and 40% of the raw-text highest-quartile selections changed. Documentary needs supported operational guidance, annotated models, child-directed language, and cultural interpretation. The resulting textbook specification connects character decisions with plot progression through planning, drafting, revision, and editing. These findings support an integrated development priority and transparent model selection; they do not demonstrate improved writing performance. Missing participant records, sampling information, and original ethics documentation constrain inference and require resolution before submission.

KEYWORDS: children's story writing, East Java, multicultural children's literature, needs analysis, pre-service primary school teachers, textbook development

Research Journal in Advanced Humanities

Volume 8, Issue 1, 2027

ISSN: 2708-5945 (Print)

ISSN: 2708-5953 (Online)

ARTICLE HISTORY

Submitted **27 August 2026**

Accepted: **14 October 2026**

Published: **1 January 2027**

HOW TO CITE

Furqon, H., Suwandi, S., & Suhita, R. (2027). Multicultural children's literature and story writing: A needs analysis in East Java. *Research Journal in Advanced Humanities*, *8*(1). https://doi.org/10.58256/rv53v440

Published in Nairobi, Kenya by Royallite Global, an imprint of Royallite Publishers Limited

© 2027 The Author(s). This is an open Access article distributed under the terms of the Creative Commons Attribution 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.

Introduction

Writing a children's story requires prospective teachers to connect a character's motives, intelligible events, and language appropriate for a child reader. Cultural representation further shapes whose experience matters and how conflicts are resolved. A textbook that treats these decisions separately may leave learners unable to construct a working narrative. The educational problem is therefore how to connect narrative craft and cultural interpretation through practical writing support.

East Java is the reported needs-study setting and intended location for textbook use. The target learners are Primary School Teacher Education (PGSD) students described as taking or having completed children's literature courses. The records do not identify all institutions, districts, or student backgrounds. Consequently, local relevance must be developed through situated examples and consultation; it cannot be inferred from provincial location alone.

A multicultural textbook must address more than stories from different places. Characters may appear diverse while having little opportunity to speak or influence events. Representation research therefore examines verbal and visual evidence together, while authenticity research questions whether apparently inclusive books recognize marginalized experiences (Adam, 2021; Adam & Urquhart, 2023; Adukia et al., 2023; Lewis et al., 2022). For writing instruction, this means evaluating voice, agency, perspective, and the consequences of characters' actions.

Indonesian work provides useful conceptual connections. Mandarani and Munir discuss criteria for selecting multicultural literature and ways to organize literature-based classrooms (Mandarani & Munir, 2021). Sudigdo and Pamungkas analyze multicultural values in elementary pupils' poetry, while Pamungkas and Sudigdo examine diversity in stories associated with MBKM students (Pamungkas & Sudigdo, 2022; Sudigdo & Pamungkas, 2022). A further study considers children's poetic expressions of national identity (Sudigdo et al., 2024). These studies justify attention to cultural meaning in literary production, but their participants and genres differ from the present PGSD writing task. They inform textbook questions without supplying missing observations about the East Java students.

Model selection presents another problem: public digital files combine stories with credits, licence notices, and institutional material. Whole-file language profiles may therefore misrepresent the passage students analyze. Child-directed print resources emphasize corpus construction and intended textual units (Green et al., 2024; Nag et al., 2024). Computational analysis can support model preparation, provided that the selected passage is inspected and its educational purpose remains explicit. It cannot substitute for local needs evidence or cultural judgment.

Literature review and conceptual framework

Learning to write for a child reader

Writing for children involves an audience relationship rather than a rule that every sentence should be short. A story may introduce unfamiliar language when actions, illustrations, or context make its meaning recoverable. Nation and colleagues distinguish the language experience offered by books from that of ordinary conversation (Nation et al., 2022). Lexical resources for picture books and books read independently provide further evidence that child-directed print is not a uniform register (Green et al., 2024; Korochkina et al., 2024). These findings support deliberate model selection when prospective teachers learn to write for a child audience. They do not provide a validated age threshold for the Indonesian stories examined here.

Studies of children's writing, freely told fantasy stories, and naturalistic picturebook reading document distinct production conditions (Hsiao et al., 2024; Stoops & Montag, 2024; van Dijk et al., 2023). PGSD learners are adults composing for children, so these resources cannot supply direct norms for their assessed performance. They instead support comparisons of audience, narrative choices, and spoken and written storytelling.

Morphological complexity varies with intended age and genre, while book morphology can contribute to language learning (Dawson et al., 2023; Korochkina & Rastle, 2025). Emotion vocabulary and lexical environments also differ across modalities (Dong & Nation, 2025; Li et al., 2025). Textbook exercises should ask why particular words help a child understand a scene, rather than reward removing every unfamiliar expression.

Multicultural literature in teacher education

Teacher education studies locate text selection within the development of professional judgment. Simoncini and colleagues examine prospective teachers' responses to diverse and multicultural books, and Iwai considers teacher self-identities through a multicultural literature project (Iwai, 2024; Simoncini et al., 2022). Sun addresses critical attention to hidden bias in literature within EFL teacher education (Sun, 2024). Together, these studies support tasks that require learners to explain their interpretations instead of simply identifying a cultural theme. Such tasks are relevant when PGSD students write their own stories because authors must make the same choices that critical readers evaluate. Their relevance remains conceptual unless tested in the East Java setting.

Text selection can reproduce cultural boundaries despite a curriculum's commitment to diversity. Research on racial representation, selective traditions, and literature courses makes book availability only part of the problem (Butković & Vidović, 2021; Liang et al., 2024; Quick, 2024). A textbook should explain selection criteria and visible perspectives, while allowing East Java students to compare local and unfamiliar experiences without treating either as homogeneous.

A practical multicultural orientation therefore concerns the narrative distribution of attention and power. A story can name several groups while allowing only one character to solve the problem. It can also portray cooperation without making the characters' differences relevant to their actions or understanding. Frameworks for inclusive literature and research on preservice teachers' interpretations of gender representation offer useful prompts for making such choices visible (Haghanikar et al., 2022; Pfundheller & Liesch, 2023). In the proposed textbook, students would identify who speaks, who makes a consequential decision, and whose account of a conflict is accepted. These are interpretive questions, not completed measurement categories in the present dataset.

Permatasari and Andriyanti (2021) examine cultural text-based teaching with engineering students and identify writing-practice constraints. That population cannot supply an effect estimate for PGSD story writing. It nevertheless encourages specifying how a cultural text becomes a learning activity. Each model should therefore lead to an explicit writing task and product whose connection to cultural interpretation can later be evaluated.

From reported needs to textbook requirements

Process-oriented instruction offers one way to organize those components. Reviews of writing instruction and feedback show that instructional focus and learner characteristics matter when evaluating outcomes (Kim et al., 2021; Scherer et al., 2024). Oral narrative instruction supplies a related example of connecting narrative work with writing in a different learner population (Kirby et al., 2021). These sources support separating planning, drafting, substantive revision, and editing in the proposed design. They do not justify importing published gains into an untested textbook. The relevant design question is which activity addresses each documented difficulty and what subsequent student work would demonstrate progress.

Computational support for choosing models

MATTR quantifies local lexical variation through a sliding window, with interpretation dependent on text length and measurement parameters (Bestgen, 2024; Covington & McFall, 2010; Zenker & Kyle, 2021). Research across readerships and language samples requires careful construct definition (Haverals et al., 2022; Yang et al., 2022); word-class measures add differentiation where annotation permits (Kim et al., 2026). Here, diversity describes extracted versions, not competence or multicultural quality. Readability indices also cannot settle accessibility alone (Matricciani, 2023).

Paratext mediates translated children's literature, and OCR errors affect downstream analysis (Lyu et al., 2021; Men et al., 2026; de Oliveira et al., 2023). Locating textual units and assigning categories are distinct annotation tasks (Dobrić, 2023). Dataset documentation and Indonesian-language corpus work reinforce the need to record collection contents (Cahyawijaya et al., 2023; Gebru et al., 2021). These considerations protect comparisons of candidate passages without validating their cultural relevance.

Research questions

Four questions guide the analysis: How prevalent are the two reported writing difficulties, and what overlap or difference is identifiable? What textbook support appears in the documentary accounts? How sensitive are

candidate model-text profiles to extraction boundaries? What instructional components follow from these findings? These questions address needs and development decisions, not learning gains.

Methods

Design and evidence hierarchy

We conducted a documentary reanalysis of an existing needs-study manuscript and a supplementary computational audit of public children's books. The original manuscript described a descriptive-exploratory mixed-methods study using lecturer and student questionnaires in East Java. The records available for the present analysis were the manuscript's aggregate tables, thematic descriptions, and selected reported quotations. The respondent-level questionnaire dataset and complete qualitative responses were unavailable. Accordingly, this article reports a reanalysis and design synthesis rather than claiming a newly completed mixed-methods field study. The public corpus supplies additional observations about candidate reading materials, not additional participants.

Separating these evidence sources protects the interpretation: learner needs guide instructional priorities, while book measurements inform how potential models are prepared.

Setting and available participant information

The source manuscript locates the study in higher education institutions offering PGSD programmes in East Java, Indonesia. Its results identify 120 student respondents, and both recoverable item distributions sum to that total. The methods section still contained an unresolved sample-size placeholder, which was reconciled against those results rather than treated as a separate cohort. Students were described as currently enrolled in or having completed children's literature courses. The institutional roster, number of institutions, recruitment procedures, dates, response rate, and demographic breakdown were not recoverable from the supplied material. The number of lecturers was also unavailable, so their accounts are not assigned a denominator.

Accordingly, the available sample supports a situated development rationale; it cannot estimate provincial prevalence or compare institutions and demographic subgroups.

Questionnaire recalculation and overlap bounds

The plot-development item contained 15 strongly agree, 88 agree, 16 disagree, and one strongly disagree response. The characterization item contained 18, 87, 14, and one response in the same order. We recalculated every percentage using 120 as the denominator and combined the first two categories as agreement. Conditional 95% Wilson intervals describe binomial uncertainty under independent responses (Wilson, 1927). They do not account for unknown institutional clustering, selection, or nonresponse. No reliability coefficient, factor model, subgroup test, or relationship with actual writing quality was estimated. With responses coded 1 for strongly disagree through 4 for strongly agree, we calculated marginal means and sample standard deviations from category counts, alongside ordinal medians and the full distributions. Equal category spacing is an explicit descriptive convention, not an established psychometric property. We therefore interpreted agreement prevalence and response intensity separately.

The two margins also permit exact bounds on overlap without reconstructing participant data. If P is the set agreeing with plot difficulty and C the set agreeing with characterization difficulty, the minimum possible intersection is max (0, 103 + 105 – 120), or 88. The maximum is min (103, 105), or 103. We enumerated all 16 integer-valued two-by-two tables consistent with the margins and checked their totals. These tables represent feasible arrangements, not observed responses or simulated participants. Bounds derived from them are logical identification limits, not confidence intervals. For every feasible table, we calculated the two-sided exact McNemar test from the discordant counts, using twice the smaller binomial tail under probability .5, capped at one. This is a sensitivity analysis over unidentified pairings, not a test on recovered individual records. We also calculated the attainable phi coefficients to determine whether the sign of association is identified. Both procedures require the same 120 respondents and common item completeness. They do not correct unknown recruitment bias or institutional dependence.

The analysis therefore prioritizes identifiable quantities and attainable limits, allowing instructional decisions while preserving uncertainty about individual relationships between responses.

Documentary synthesis of qualitative needs

We extracted the eight lecturer-needs dimensions from the source manuscript's first table and the five student-needs themes from its narrative results. For each dimension, we recorded its original location, respondent group, reported meaning, and possible textbook implication. Overlapping descriptions were grouped for presentation while retaining the original categories in the supplementary audit. The process reorganized documented findings; it did not repeat thematic coding on unavailable raw responses. Theme frequencies, saturation, and inter-coder reliability were therefore not calculated. The selected quotations in the source were paraphrased as reported accounts rather than presented as newly authenticated participant testimony. These accounts are most useful for specifying the form of instructional support, rather than estimating how commonly each preference occurs.

Supplementary corpus and model-text measurements

The public haznitrama/indonesian-children-books repository is documented in the supplied reproducibility package as retrieved on 3 November 2026 at revision e4d616bae9c887e2d017aaf626ce8773de0ae645. The controlling Parquet file contained 165,245 page records and 2,740 file-level title labels. A second file contained 1,595 whole documents; normalized filenames overlapped with 1,594 page-file titles, so the files were not pooled. File sizes and SHA-256 hashes were checked against repository objects. We retained books with an explicit early translated-edition CC BY-NC 4.0 notice and identifiable body boundaries. Four placeholder-only segments were excluded, leaving 163 file-level books.

The body began after the standardized early licence page and ended before a later institutional or contributor-credit marker. The rules and page decisions are supplied with the analysis. They identify a reproducible segment that may include activities and image-dependent language, rather than a human-validated pure story. We compared full raw extraction, layout-cleaned extraction, delimited body text, and a sensitivity version omitting the first and last retained body pages. Tokens were lowercase sequences of Unicode letters without stemming. We calculated body length, mean words per punctuation segment, matched quoted-token share, and moving-average type–token ratios with 50- and 100-token windows; texts shorter than a window had missing scores.

Book-level paired differences were summarized by medians and 95% percentile bootstrap intervals from 4,000 resamples with fixed seeds. Ranking stability compared eligible books across raw and body versions, with top-quartile size rounded upward. Sensitivity restrictions addressed body length, OCR-labelled pages, source-credit strings, and automatic language labels. A word-trigram TF–IDF similarity screen found no retained pairs above 0.90. The corpus is a selective resource pool whose extracted licence notices are not an independent legal review. Full books are not redistributed, and all model-text decisions require consultation of the original edition and its credits. The present revision independently recomputed all 15 supplied paired medians and the four top-quartile overlaps from book_features.csv. Bootstrap limits are retained from the documented package; the original source download and text extraction were not rerun for this revision. Thus, numerical consistency of derived results is distinguished from independent verification of every original page boundary.

Integration and ethical boundaries

The reanalysis used manuscript aggregates and public texts. Original ethics approval, consent, and institutional permissions were unavailable; no approval from another project is attributed to this study. Submission requires verification of the original fieldwork documentation. This evidential boundary permits analysis of recoverable findings while preventing unsupported claims about ethical clearance or instructional effectiveness.

Results

Reported writing difficulties in the East Java sample

Agreement with plot-development difficulty was 103 of 120 students, or 85.83%, compared with 105 of 120, or 87.50%, for characterization. The corresponding conditional Wilson intervals were 78.48–90.96% and 80.40–92.28%. For plot, strongly agree and agree accounted for 12.50% and 73.33%; the remaining categories accounted for 13.33% and 0.83%. For characterization, the respective percentages were 15.00%, 72.50%, 11.67%, and 0.83%. These recalculations correct the earlier totals of 85.9% and 87.7% and several category percentages. They also replace the unsupported statement that all 120 students reported characterization

difficulty. Table 1 provides the complete distributions and supplementary descriptive summaries. The small number of strongest endorsements qualifies the meaning of the high combined-agreement proportions: many students acknowledged difficulty, but these items do not establish severe impairment.

Table 1. Questionnaire distributions and descriptive summaries for 120 students.

Response or statistic	Plot development	Characterization
Strongly disagree	1 (0.83%)	1 (0.83%)
Disagree	16 (13.33%)	14 (11.67%)
Agree	88 (73.33%)	87 (72.50%)
Strongly agree	15 (12.50%)	18 (15.00%)
Combined agreement	103 (85.83%)	105 (87.50%)
Conditional 95% Wilson interval	78.48–90.96%	80.40–92.28%
Mean (sample SD), coding 1–4	2.975 (0.542)	3.017 (0.550)
Median and mode	3; 3	3; 3

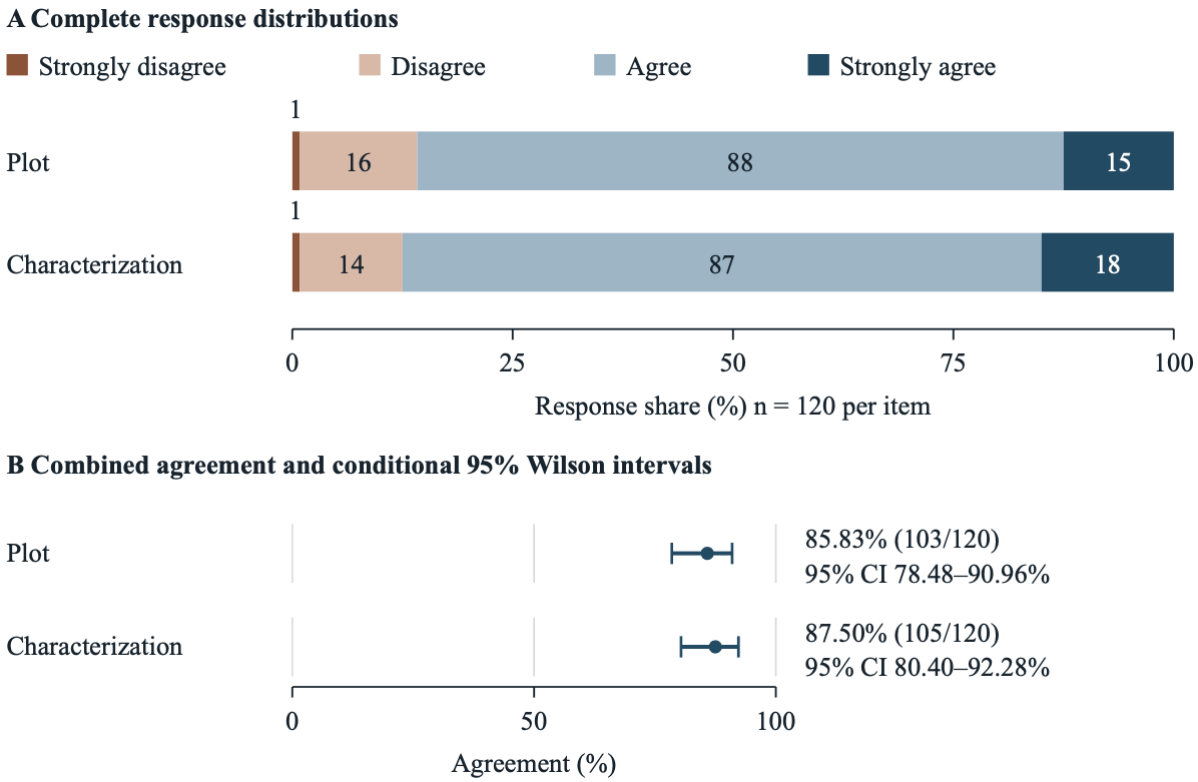


Figure 1. Reported plot and characterization difficulties among 120 PGSD students in the East Java source manuscript. The upper panel preserves all four response categories; the lower panel combines agreement categories and shows conditional 95% Wilson intervals. Counts are manuscript-reported aggregates. The intervals assume independent responses and do not establish provincial prevalence.

The combined-agreement estimates indicate broad demand for support, while the smaller strongly agree shares caution against equating prevalence with severity. Plot and characterization should both be available as core learning pathways, with initial writing used to adjust the intensity of support rather than infer individual deficits from aggregate responses.

Joint need that can be established from aggregate data

The margins imply that at least 88 students, or 73.33% of the reported sample, agreed with both difficulties. At most 103 students, or 85.83%, could have agreed with both. Between 105 and 120 students could have agreed with at least one item. Plot-only difficulty could involve zero to 15 students, characterization-only difficulty two to 17, and neither difficulty zero to 15. Figure 2 presents these attainable ranges. No particular arrangement within them is asserted to be the observed joint distribution. Marginal means were 2.975 (SD 0.542) for plot and 3.017 (SD 0.550) for characterization, with median and mode equal to 3 for both. These summaries support

the same broad pattern as the frequency distributions; they do not constitute a validated composite scale.

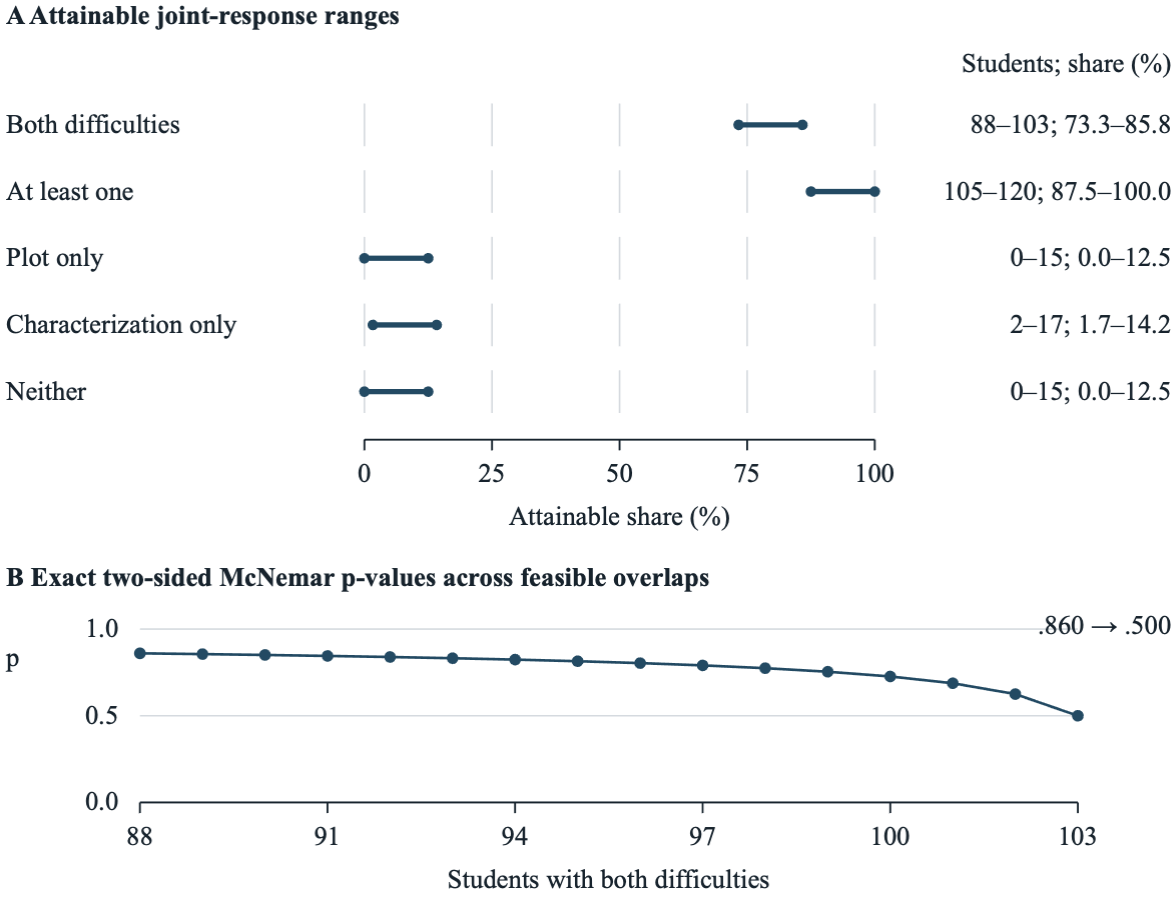


Figure 2. Partial identification from the two questionnaire margins. Panel A shows attainable count and percentage ranges. Panel B displays exact McNemar p-values across all 16 feasible overlaps. These are possible response arrangements, not observed individual records; the bounds are not confidence intervals. Both panels assume the same 120 respondents and no item-specific missingness.

The minimum overlap is sufficiently large to justify integrating character decisions with event progression even though individual pairings remain unknown. This is a curricular inference from shared prevalence, not evidence that the two difficulties are correlated or that one causes the other. The distinction permits a useful design decision without exceeding what the margins identify.

Sensitivity of the comparison to unknown pairing

The observed marginal difference is 2/120, or 1.67 percentage points. Every compatible table has two more characterization-only than plot-only endorsements, but the total number of discordant responses can range from 2 to 32. The exact McNemar p-value consequently ranges from .500 to .860 (Table 2). None of the feasible pairings supports a conventional .05 rejection of equal marginal probabilities under the test assumptions. This strengthens the recommendation to address both domains together, although non-rejection does not establish equivalence or equal educational importance.

Table 2. Selected feasible joint tables and exact McNemar sensitivity. Rows are possibilities, not observed pairings.

Both	Plot only	Character only	Neither	Exact p
88	15	17	0	.860
93	10	12	5	.832
98	5	7	10	.774
103	0	2	15	.500

The attainable phi coefficient ranges from –0.154 to 0.930. Consequently, the high lower bound on co-

occurrence does not identify a positive association: common difficulties can overlap extensively simply because each is prevalent. No correlation, regression coefficient, or latent writing-needs factor can be recovered from the two marginal distributions. This distinction matters for textbook design. Integrated activities are justified by the number of students necessarily encountering both problems, whereas an assertion that plot difficulty causes characterization difficulty would be unsupported. The strongest inferential contribution is therefore a bounded instructional decision, rather than a claim about mechanisms.

Lecturer and student accounts of textbook support

The lecturer table documented eight dimensions: content depth, practical writing procedures, multicultural integration, contextual relevance, child-directed language, narrative competence, model texts, and conceptual foundations. The narrative account emphasized incomplete practical examples and insufficient operational guidance. These are documentary findings from the source manuscript, not new frequency estimates. Because lecturer numbers were unavailable, terms such as majority or consensus are not used to describe their distribution. The most direct design implication is that explanation should be accompanied by a visible writing procedure. A chapter must show what students are expected to produce after encountering its concepts.

The student narrative documented five additional themes: vocabulary and language adaptation, creativity and plot consistency, implicit communication of values, annotated model stories, and scaffolded adaptation exercises. These themes expand the two questionnaire items into more specific instructional requests. For example, characterization support involves language and viewpoint as well as a list of a character's traits. Plot support involves managing repetition and moving beyond an initial idea, not simply identifying the beginning, middle, and end of a published story. The source also describes concern about moral messages becoming overly didactic. That concern connects the writing task to multicultural orientation because cultural values must influence events and relationships rather than appear only as concluding advice.

Table 3. Documented needs and proposed instructional responses. Qualitative themes have no recoverable frequency denominators.

Documented need	Source evidence	Development implication
Plot and characterization	Student Tables 2–3: 103/120 and 105/120 agreement.	Connect a child character's goal, decisions, obstacles, and consequences.
Operational guidance	Lecturer dimensions 2 and 6; student theme 5.	Provide staged tasks and worked revisions rather than definitions alone.
Models and interpretation	Lecturer dimensions 1, 7, and 8; student theme 4.	Annotate selected models and link each observation to a writing task.
Child-directed language	Lecturer dimension 5; student themes 1 and 5.	Practice adapting diction, dialogue, and explanations for a specified audience.
Context and cultural meaning	Lecturer dimensions 3 and 4; student theme 3.	Examine voice, agency, cultural perspective, and implicit messages.
Sustaining narrative development	Student theme 2 and plot item.	Use causal outlines and revision prompts to address stalled or repetitive plots.

These accounts explain the kind of support the numerical prevalence alone cannot specify: learners need to see how a narrative decision is made, attempted, and revised. The textbook should consequently connect each model to an explicit writing task. Because themes have no frequency denominators, their instructional relevance should not be confused with a measured ranking of priorities.

Supplementary evidence for preparing model texts

The retained corpus contained 49,567 body tokens across 163 books and 1,908 retained page records. Median body length was 245 tokens, compared with 891 in the raw documents. Median body punctuation-segment length was 7.18 words, and median matched quoted-token share was 8.77%. These distributions provide an inspectable range of textual features for choosing short passages and comparing how language is used. They do not identify an optimal vocabulary level, ideal dialogue proportion, or validated reading age. Figure 3 therefore

presents distributions rather than thresholds for acceptable student writing.

Table 4. Within-book changes from raw extraction to body text. Percentile bootstrap limits are from the supplied package, with 4,000 book resamples.

Measure	Paired n	Median change	95% bootstrap interval
Token count	163	−639	−647 to −632
MATTR50	160	−0.0495	−0.0623 to −0.0434
MATTR100	139	−0.0664	−0.0786 to −0.0605
Words per punctuation segment	163	−1.842	−2.253 to −1.645
Quoted-token share, percentage points	163	+5.426	+3.751 to +8.547

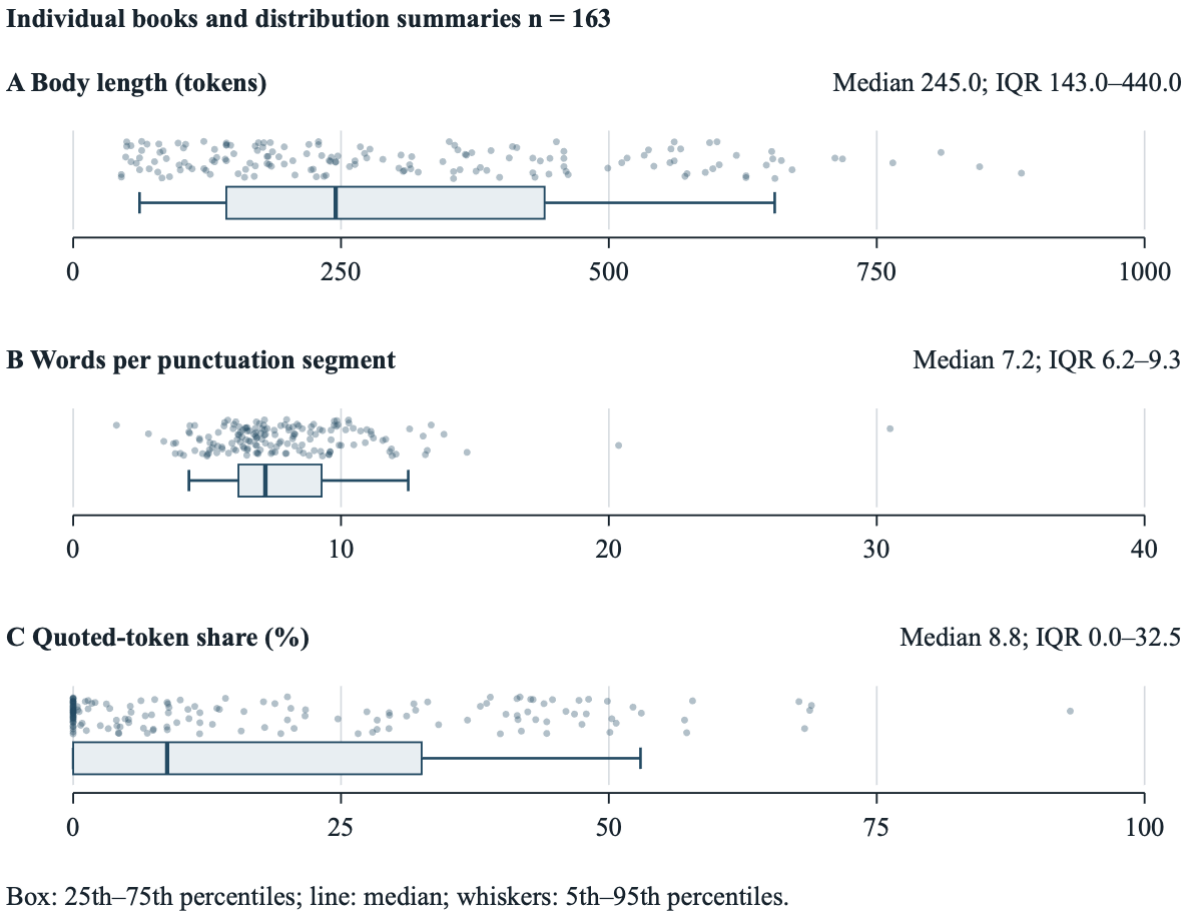


Figure 3. Book-level distributions of body length, punctuation-segment length, and quoted-token share (n = 163). Points represent books; boxes show the interquartile range and median, and whiskers mark the 5th and 95th percentiles. Each panel has its own measurement scale. Dispersion describes candidate materials, not participant performance or validated reading levels.

The usable subset represented 5.95% of the repository's file-level title labels. Detected credit strings grouped 77 books with Pratham, 18 with African Storybook, 12 with Book Dash, and 56 as other or unresolved. An automatic language screen labelled 150 bodies Indonesian, ten Malay, and three Javanese, without human adjudication. Only three retained body pages carried an OCR extraction label. These facts describe the resource's limitations and provenance; they do not classify authors' or characters' cultural identities. In particular, the three automatically labelled Javanese texts cannot be treated as evidence that the corpus represents East Javanese culture.

Boundary selection materially affected computational comparisons. For 160 eligible books, the median within-book MATTR50 change from raw extraction to body text was −0.0495, with a bootstrap interval of −0.0623 to −0.0434. Among the 40 books in the highest raw-text quartile, only 24 remained in the highest body-text quartile. The resulting replacement rate was 40%; the corresponding MATTR100 rate was 37.14%. Segment-length and quoted-share selections each replaced 24.39% of their raw top-quartile members. The

robustness and selection-loss comparison in Figure 4 shows why a textbook developer should inspect the actual passage before using a feature-based shortlist.

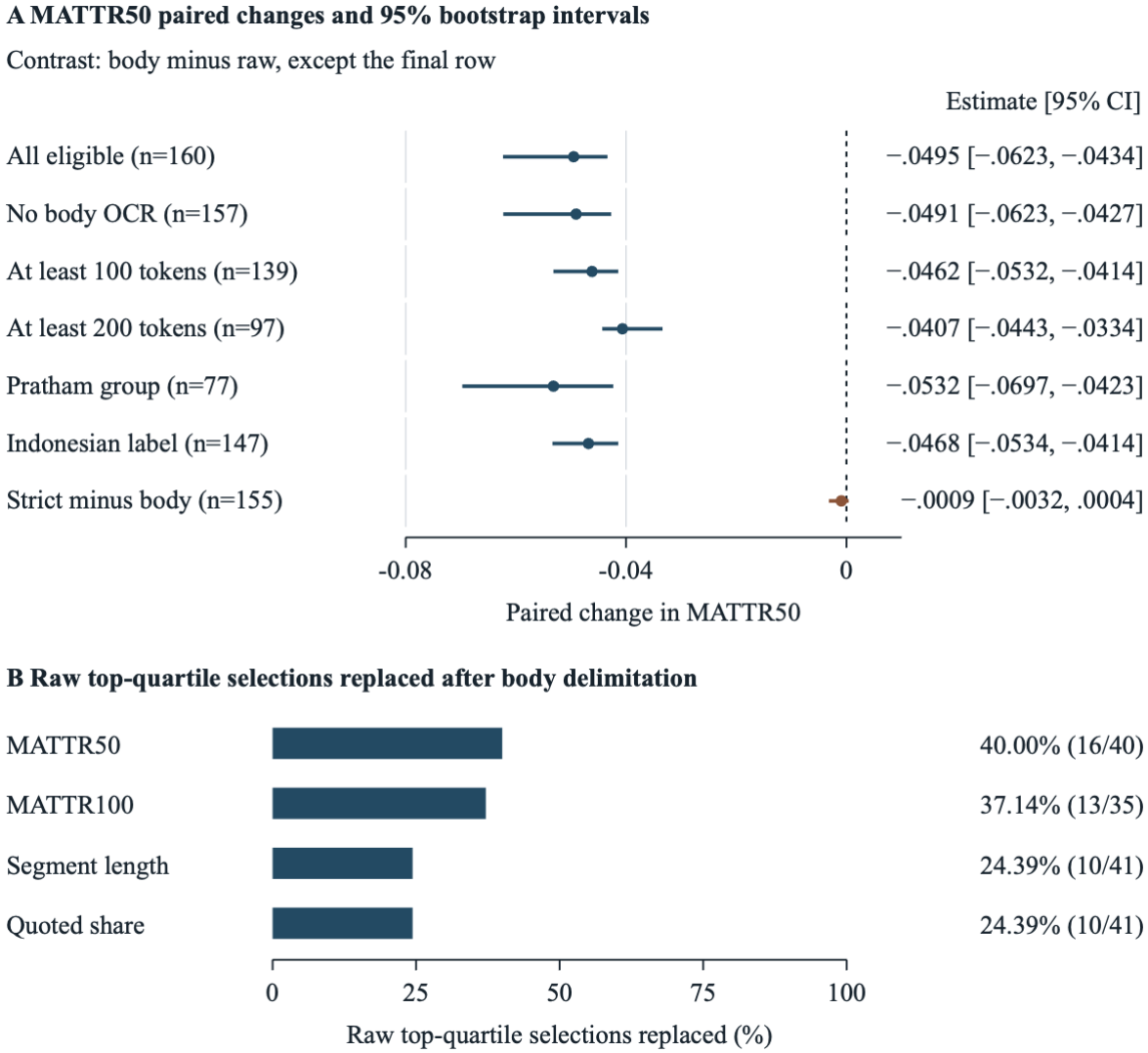


Figure 4. Robustness of model-text comparisons. Panel A reports median paired MATTR50 changes with supplied 95% percentile bootstrap intervals. The strict-body comparison uses a different contrast, identified in its label. Panel B reports the proportion and count of raw top-quartile selections absent from the body-text quartile. These results quantify processing sensitivity, not literary quality.

Table 5 shows that the main directions persisted under minimum-length, no-body-OCR, source-restricted, and automatic Indonesian-language restrictions. Omitting the first and last retained body pages produced a much smaller MATTR50 change, with a median of –0.0009 and an interval spanning zero. The supplementary package retains the full paired-effect and sensitivity displays. For the textbook, the consequence is procedural: document the edition, select the intended passage, inspect its boundaries, and then examine its language profile. A numerical shortlist should be followed by human assessment of narrative quality, cultural representation, illustration, and classroom purpose. Corpus processing supports that sequence without becoming the central educational outcome. The raw files contained 155,414 tokens, so 105,847 tokens (68.11%) lay outside the retained body segments. This corpus-wide removal fraction differs from the median book-level removal share of 72.51%, because the former weights long books more heavily. Paired feature changes must likewise be distinguished from differences between separate raw and body medians. For example, the median within-book token change was –639, whereas subtracting the two marginal medians yields –646. Table 4 reports the paired estimates appropriate to the processing comparison.

Table 5. MATTR50 sensitivity to corpus restrictions. Negative changes denote lower body-text diversity than raw-text diversity.

Restriction	Paired n	Median change	95% bootstrap interval
No body OCR pages	157	–0.0491	–0.0623 to –0.0427
At least 100 body tokens	139	–0.0462	–0.0532 to –0.0414
At least 200 body tokens	97	–0.0407	–0.0443 to –0.0334
Pratham credit group	77	–0.0532	–0.0697 to –0.0423
Automatically labelled Indonesian	147	–0.0468	–0.0534 to –0.0414

Across the processing comparisons, the consistent issue is the identity of the text being measured. Boundary checks change both estimated linguistic characteristics and membership of a numerical shortlist, while neither result certifies pedagogical quality. The defensible sequence is therefore to delimit the passage, inspect its profile, and justify its narrative and cultural suitability through human reading.

Proposed textbook architecture for the East Java setting

The integrated findings support a proposed architecture that connects narrative craft and multicultural interpretation throughout the writing process. The architecture begins with identifying a child audience and a locally situated story idea, then links character decisions with plot development. Annotated models provide examples of how those decisions are realized in text. Drafting is followed by separate opportunities to revise narrative and cultural meaning before editing language. Each stage produces a student artifact that can support feedback and later evaluation. Table 6 specifies these components as development outputs rather than tested instructional effects.

Table 6. Proposed textbook sequence and assessable products. Activities have not been classroom-tested.

Stage and purpose	Proposed activity	Student product and evaluation focus
1 Context and audience	Investigate a responsibly described East Java setting or experience; specify the intended child reader.	Context note and story premise; review source use, perspective, and audience rationale.
2 Reading models	Inspect the original edition and delimit a passage; annotate plot, character decisions, language, and cultural meaning.	Annotated passage and selection rationale; assess evidence for interpretations.
3 Planning	Connect the protagonist's goal, obstacle, decision, and consequence; consider whose voice is missing.	Character–plot plan; examine causal progression and child agency.
4 Drafting	Write a complete story using the plan while allowing justified changes; test dialogue and viewpoint.	First draft; assess coherence, characterization, and intended audience.
5 Substantive revision	Use focused peer feedback on plot, characterization, and representation before correcting surface form.	Revised draft and change log; examine whether identified problems were addressed.
6 Editing and reflection	Review diction, sentence clarity, conventions, and implicit messages; explain major choices.	Final story and reflection; use a piloted human rubric in future evaluation.

The proposed sequence makes development accountable by connecting each activity to an observable product and an identified need. Its coherence lies in the progression from interpretation to production and revision, rather than the number of topics included. Classroom testing must establish whether that progression improves students' stories; the architecture itself is a design outcome, not an effectiveness result.

Discussion

The central result is the scale of shared instructional need, rather than a defensible hierarchy between plot and characterization. Agreement reached 85.83% and 87.50%, respectively, and at least 73.33% of the common sample must have endorsed both difficulties. Yet the response distributions concentrate in agree rather than strongly agree, and exact McNemar results remain non-significant across every feasible pairing. These findings justify coordinated support without portraying all participants as severely deficient or treating non-rejection as equivalence. Writing-instruction and feedback syntheses show why instructional focus and learner characteristics matter, rather than supporting a universal procedure detached from the task (Kim et al., 2021; Scherer et al., 2024). Narrative instruction also offers a rationale for connecting oral planning to written production, although evidence from kindergarten cannot establish effectiveness among adult PGSD students (Kirby et al., 2021). The implication is to organize textbook work around a character's goal, obstacle, decision, and consequence, so that characterization generates plot progression. An initial story-writing task should then distinguish perceived difficulty from demonstrated performance and determine how much support each learner needs. The quantitative contribution is a transparent basis for this integrated priority, not a retrospective diagnosis of individual competence.

That integrated priority becomes multicultural when narrative decisions redistribute voice and agency, rather than merely introduce culturally marked names or settings. Research on authenticity and representation shows why the presence of diverse characters cannot alone establish inclusion, especially when verbal and visual portrayals distribute visibility and power differently (Adam, 2021; Adam & Urquhart, 2023; Adukia et al., 2023; Lewis et al., 2022). Teacher-education studies likewise locate cultural interpretation within professional judgment and the examination of hidden bias (Iwai, 2024; Simoncini et al., 2022; Sun, 2024). Indonesian studies of diversity in student stories and poetry provide relevant conceptual connections, while retaining different participants and genres from this study (Pamungkas & Sudigdo, 2022; Sudigdo & Pamungkas, 2022; Sudigdo et al., 2024). The textbook should therefore ask whose problem organizes the story, who can act, and whose explanation becomes authoritative. Annotated drafts and revision notes can make those choices inspectable. This connects the reported need for operational guidance with cultural interpretation through a concrete writing practice, while leaving the proposed relationship between instruction and inclusive storytelling open to empirical evaluation.

Contextual relevance in East Java should similarly be established through accountable interpretation rather than an assumed common cultural voice. Rifai and Pamungkas (2025) examine a Banyumasan folk story through literary anthropology, illustrating a culturally situated object of study rather than supplying evidence about the present students. Fadila and Pamungkas (2025) connect emotion, character development, and conflict in adult fiction; their approach suggests questions about motivation but cannot prescribe child-directed content. Research on bilingual picturebooks and a conceptual framework for inclusive multilingual learning further support explaining choices of dialogue, glossing, and participation (Ayuningtyas & Rakhmawati, 2026; Potter et al., 2026). In practice, students should justify how an expression or cultural detail helps a specified child reader understand an action. Familiar local material still needs scrutiny for stereotypes, and translated material needs a clear rationale for its instructional use. Cross-curricular literary work offers another possible application, provided that a story and its accompanying exercise have distinct purposes (Throop Robinson, 2023). Locality thus becomes an explicit condition of model selection and writing, rather than a label presumed to guarantee authenticity.

The corpus results reveal a separate threat to instructional judgment: a numerical profile may describe the downloaded document more strongly than the passage learners actually encounter. Delimitation removed 68.11% of pooled tokens, reduced median within-book MATTR50 by 0.0495, and replaced 40% of the highest-quartile selections. Sensitivity restrictions retained the negative direction, whereas the smaller strict-body contrast indicates that initial boundary identification matters more than the tested edge-page adjustment. These processing comparisons strengthen the case for documenting textual units, consistent with corpus-construction and dataset-documentation principles (Gebru et al., 2021; Nag et al., 2024). Lexical-diversity research further explains why window size and text length constrain interpretation (Bestgen, 2024; Covington & McFall, 2010; Zenker & Kyle, 2021). Neither rank stability nor a high lexical-diversity score establishes literary quality, reading age, or cultural adequacy. Computational screening should therefore precede human review of narrative

structure, illustrations, and classroom purpose. If students use generative tools during drafting, attribution and retained revisions can also clarify their own decisions, an issue raised conceptually in discussions of AI-assisted authorship (Amil, 2025). Automated support remains useful when its contribution is traceable and its limits remain visible.

Together, the findings support a development specification whose effectiveness must be tested with actual student writing. A subsequent evaluation should use comparable prompts, documented instructional exposure, independently scored drafts, and a piloted rubric covering plot coherence, characterization, child-directed language, and cultural representation. Feedback research supports evaluating the substance of revision, while annotation research cautions that agreement depends on how units and categories are defined (Dobrić, 2023; Scherer et al., 2024). The current evidence cannot supply that evaluation: participant-level responses, institution identifiers, lecturer denominators, complete qualitative records, and ethics documentation were unavailable. Conditional Wilson intervals do not correct selection or clustering, and the attainable overlap bounds do not identify the correlation between needs. The 163-book subset is also selective and text-focused; automatic language and source labels cannot establish cultural identity. Consequently, the strongest contribution is the connection between quantified needs, documented requests, and assessable textbook activities. Future work should test whether that connection produces better stories and more defensible representational choices, rather than treating textbook completion or favorable perceptions as proof of improved writing.

Conclusion

This study establishes a quantitative rationale for developing a multicultural children's literature textbook that integrates plot construction and child characterization for pre-service primary school teachers in the reported East Java setting. Of 120 students, 103 endorsed plot difficulty and 105 endorsed characterization difficulty; at least 88 must have endorsed both under the common-sample assumption. The small marginal difference does not support separating the two domains into unequal instructional priorities. Documentary accounts further connect these needs with annotated models, operational writing guidance, child-directed language, and substantive cultural interpretation. The supplementary corpus analysis demonstrates that model selection is sensitive to document boundaries, reinforcing the need to define and inspect the actual text before interpreting its numerical profile. The study's contribution is therefore an evidence-based development specification that connects narrative craft, cultural representation, and transparent preparation of teaching materials.

The proposed textbook should translate this specification into a sequence of contextual inquiry, model reading, planning, drafting, substantive revision, and editing, with an assessable student product at each stage. Its practical value will depend on whether learners can produce more coherent stories, give child characters meaningful agency, and justify cultural and linguistic choices for their intended readers. These outcomes have not yet been demonstrated. The next phase should recover the original sampling and ethics records, pilot the assessment rubric, and evaluate actual writing under comparable conditions with independent scoring. Institutional dependence and differences in learners' starting competence should be incorporated where the data permit. Such evaluation would extend the present needs analysis into evidence of instructional effectiveness while preserving the distinction between a justified textbook design and a tested improvement in writing.

Data availability and declarations

The public corpus is available at https://huggingface.co/datasets/haznitrama/indonesian-children-books. The supplied reproducibility package contains derived data, source hashes, eligibility decisions, feature measurements, and analysis code. The author-supplied manuscript provides local aggregate and documentary evidence. Participant-level responses, complete qualitative records, and student writing samples were unavailable. The present revision's verification scripts and joint-table sensitivity calculations are retained with the working analysis. Full public books and personal participant information are not redistributed.

Funding

This research received financial support from BPI (Indonesian Education Scholarship), PPAPT (Center for Higher Education Funding and Assessment), and LPDP (Indonesian Endowment Fund for Education), as reported in the original manuscript. No grant or award number was supplied.

Acknowledgements

The authors thank BPI, PPAPT, and LPDP for the support acknowledged in the original study. The authors also acknowledge the creators, translators, publishers, and repository contributors whose credited children's books enabled the supplementary analysis of model-text preparation.

Author contributions

The original manuscript attributes the following contributions: **Hendrik Furqon**: conceptualization, data curation, formal analysis, investigation, methodology, visualization, writing the original draft, and review and editing. **Sarwiji Suwandi** and **Raheni Suhita**: supervision, validation, methodology, and review and editing.

Conflicts of interest

The original manuscript declares no conflict of interest.

Research integrity statement

Original ethics approval and consent records were unavailable for this reanalysis; the responsible authors must supply the applicable fieldwork documentation before submission. Funding is carried forward from the original study without an unsupported claim about funders' involvement. Final approval of this revised analysis remains with the named authors. AI assistance supported analysis code, source checking, and manuscript editing; accountability for the data, interpretation, and final submission remains with the authors.

References

Adam, H. (2021). When authenticity goes missing: How monocultural children's literature is silencing the voices and contributing to invisibility of children from minority backgrounds. Education Sciences, 11(1), 32. https://doi.org/10.3390/educsci11010032

Adam, H., & Urquhart, Y. (2023). A cause for hope or an unwitting complicity? The representation of cultural diversity in award-listed children's picturebooks in Australia. Bookbird: A Journal of International Children's Literature, 61(2), 48–58. https://doi.org/10.1353/bkb.2023.0023

Adukia, A., Eble, A., Harrison, E., Runesha, H. B., & Szasz, T. (2023). What we teach about race and gender: Representation in images and text of children's books. The Quarterly Journal of Economics, 138(4), 2225–2285. https://doi.org/10.1093/qje/qjad028

Amil, A. J. (2025). Negotiating authorship and creativity in AI-assisted poetry: A critical analysis of contemporary literary debates. Journal of Language, Literature, and Cultural Dynamics, 2(3), 23–36. https://doi.org/10.65840/jllcd.v2i3.46

Ayuningtyas, A. P., & Rakhmawati, A. (2026). Reframing inclusive language learning in Indonesian multilingual classrooms: A TPACK-based model for meaningful and transformative pedagogy. Journal of Language, Literature, and Cultural Dynamics, 3(1), 22–42. https://doi.org/10.65840/jllcd.v3i1.36

Bestgen, Y. (2024). Measuring lexical diversity in texts: The twofold length problem. Language Learning, 74(3), 638–671. https://doi.org/10.1111/lang.12630

Butković, M., & Vidović, E. (2021). Representation of racial diversity in picturebooks in teacher education programs in the Republic of Croatia. Journal of Literary Education(4), 50–68. https://doi.org/10.7203/jle.4.21012

Cahyawijaya, S., Lovenia, H., Koto, F., Adhista, D., Dave, E., Oktavianti, S., Akbar, S., Lee, J., Shadieq, N., Cenggoro, T. W., Linuwih, H., Wilie, B., Muridan, G., Winata, G., Moeljadi, D., Aji, A. F., Purwarianti, A., & Fung, P. (2023). NusaWrites: Constructing high-quality corpora for underrepresented and extremely low-resource languages. Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (Volume 1: Long Papers), 921–945. https://doi.org/10.18653/v1/2023.ijcnlp-main.60

Covington, M. A., & McFall, J. D. (2010). Cutting the Gordian knot: The moving-average type–token ratio (MATTR). Journal of Quantitative Linguistics, 17(2), 94–100. https://doi.org/10.1080/09296171003643098

Dawson, N., Hsiao, Y., Tan, A. W. M., Banerji, N., & Nation, K. (2023). Effects of target age and genre on morphological complexity in children's reading material. Scientific Studies of Reading, 27(6), 529–556. https://doi.org/10.1080/10888438.2023.2206574

de Oliveira, L. L., Vargas, D. S., Alexandre, A. M. A., Cordeiro, F. C., Gomes, D. d. S. M., Rodrigues, M. d. C., Romeu, R. K., & Moreira, V. P. (2023). Evaluating and mitigating the impact of OCR errors on information retrieval. International Journal on Digital Libraries, 24(1), 45–62. https://doi.org/10.1007/s00799-023-00345-6

Dobrić, N. (2023). Identifying errors in a learner corpus – the two stages of error location vs. error description and consequences for measuring and reporting inter-annotator agreement. Applied Corpus Linguistics, 3(1), Article 100039. https://doi.org/10.1016/j.acorp.2022.100039

Dong, Y., & Nation, K. (2025). Charting the frequency and diversity of emotion words in children's language: Written language matters. First Language, 45(4), 457–475. https://doi.org/10.1177/01427237251339788

Fadila, A. W., & Pamungkas, O. Y. (2025). Emotional representation of characters in the novel After Tunangan by Agnes Jessica: A literary psychology perspective. Journal of Language, Literature, and Cultural Dynamics, 2(2), 1–9. https://doi.org/10.65840/jllcd.v2i02.24

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86–92. https://doi.org/10.1145/3458723

Green, C., Keogh, K., Sun, H., & O'Brien, B. (2024). The children's picture books lexicon (CPB-Lex): A large-scale lexical database from children's picture books. Behavior Research Methods, 56(5), 4504–4521. https://doi.org/10.3758/s13428-023-02198-y

Haghanikar, T. M., Jiang, S., Tomek, S., & Hooper, L. M. (2022). Exploring pre-service teachers' perspectives: Gender and gender representation across human protagonists in picture books. International Journal of Educational Methodology, 8(1), 91–105. https://doi.org/10.12973/ijem.8.1.91

Haverals, W., Geybels, L., & Joosen, V. (2022). A style for every age: A stylometric inquiry into crosswriters for children, adolescents and adults. Language and Literature: International Journal of Stylistics, 31(1), 62–84. https://doi.org/10.1177/09639470211072163

Hsiao, Y., Dawson, N. J., Banerji, N., & Nation, K. (2024). A corpus-based developmental investigation of linguistic complexity in children's writing. Applied Corpus Linguistics, 4(1), Article 100084. https://doi.org/10.1016/j.acorp.2024.100084

Iwai, Y. (2024). Preservice teachers' reflections on teacher self-identities through a multicultural children's literature project. Multicultural Learning and Teaching, 19(2), 269–292. https://doi.org/10.1515/mlt-2022-0021

Kim, T., Larsson, T., Kaatari, H., Wang, Y., & Sundqvist, P. (2026). Introducing and evaluating a measure of lexical diversity across word classes. TESOL Quarterly, Article tesq.70154. https://doi.org/10.1002/tesq.70154

Kim, Y. S. G., Yang, D., Reyes, M., & Connor, C. (2021). Writing instruction improves students' writing skills differentially depending on focal instruction and children: A meta-analysis for primary grade students. Educational Research Review, 34, Article 100408. https://doi.org/10.1016/j.edurev.2021.100408

Kirby, M. S., Spencer, T. D., & Chen, Y. J. I. (2021). Oral narrative instruction improves kindergarten writing. Reading & Writing Quarterly, 37(6), 574–591. https://doi.org/10.1080/10573569.2021.1879696

Korochkina, M., & Rastle, K. (2025). Morphology in children's books, and what it means for learning. npj Science of Learning, 10(1), Article 22. https://doi.org/10.1038/s41539-025-00313-6

Korochkina, M., Marelli, M., Brysbaert, M., & Rastle, K. (2024). The children and young people's books lexicon (CYP-Lex): A large-scale lexical database of books read by children and young people in the United Kingdom. Quarterly Journal of Experimental Psychology, 77(12), 2418–2438. https://doi.org/10.1177/17470218241229694

Lewis, M., Cooper Borkenhagen, M., Converse, E., Lupyan, G., & Seidenberg, M. S. (2022). What might books be teaching young children about gender? Psychological Science, 33(1), 33–47. https://doi.org/10.1177/09567976211024643

Li, L., Song, M., & Cai, Q. (2025). The words children hear and see: Lexical diversity across-modalities and its impact on lexical development. Developmental Science, 28(2), Article e13601. https://doi.org/10.1111/desc.13601

Liang, L. A., Cromwell, R., & Hacker, D. J. (2024). The impact of children's and young adult literature courses on teachers' selection of global and culturally diverse texts for the classroom. Journal of Adolescent & Adult Literacy, 68(2), 118–128. https://doi.org/10.1002/jaal.1341

Lyu, L., Koutraki, M., Krickl, M., & Fetahu, B. (2021). Neural OCR post-hoc correction of historical corpora. Transactions of the Association for Computational Linguistics, 9, 479–493. https://doi.org/10.1162/tacl_a_00379

Mandarani, V., & Munir, A. (2021). Incorporating multicultural literature in EFL classroom. IJELTAL (Indonesian Journal of English Language Teaching and Applied Linguistics), 6(1), 1–12. https://doi.org/10.21093/ijeltal.v6i1.832

Matricciani, E. (2023). Readability indices do not say it all on a text readability. Analytics, 2(2), 296–314. https://doi.org/10.3390/analytics2020016

Men, D., Wang, Y., Ganapathy, M., & Baagbah, S. Y. S. (2026). Paratextual adaptation in English-Chinese translation of children's literature: A grounded theory study. Sage Open, 16(4), Article 21582440261497482. https://doi.org/10.1177/21582440261497482

Nag, S., John, S., & Agrawal, A. (2024). NSP-SCD: A corpus construction protocol for child-directed print in understudied languages. Behavior Research Methods, 56(4), 2751–2764. https://doi.org/10.3758/s13428-024-02339-x

Nation, K., Dawson, N. J., & Hsiao, Y. (2022). Book language and its implications for children's language, literacy, and development. Current Directions in Psychological Science, 31(4), 375–380. https://doi.org/10.1177/09637214221103264

Pamungkas, O. Y., & Sudigdo, A. (2022). Profile of Pancasila students: Implementation of diversity in MBKM student's stories in UST Yogyakarta. Daengku: Journal of Humanities and Social Sciences Innovation, 2(2), 156–164. https://doi.org/10.35877/454ri.daengku870

Permatasari, I., & Andriyanti, E. (2021). Developing students' intercultural communicative competence through cultural text-based teaching. Indonesian Journal of Applied Linguistics, 11(1), 72–82. https://doi.org/10.17509/ijal.v11i1.34611

Pfundheller, M., & Liesch, J. (2023). Framework for inclusive literature in teacher education. Journal of Higher Education Theory and Practice, 23(15). https://doi.org/10.33423/jhetp.v23i15.6406

Potter, C. E., Castellana, M. A., Guerra, M. D., & Benitez, V. L. (2026). The balance of Spanish and English child-directed text in bilingual picture books. Journal of Early Childhood Literacy, 26(1), 138–171. https://doi.org/10.1177/14687984241257400

Quick, J. (2024). Subverting the selective tradition? A self-exploration of text selection in pre-service teacher education. The Australian Journal of Language and Literacy, 47(1), 53–72. https://doi.org/10.1007/s44020-023-00047-1

Rifai, I., & Pamungkas, O. Y. (2025). Javanese culture in the folk story of Banyumasan Panembahan Situmpur Kapiten Kwe Situmpoa: A study of literary anthropology. Journal of Language, Literature, and Cultural Dynamics, 2(1), 1–11. https://doi.org/10.65840/jllcd.v2i01.20

Scherer, S., Graham, S., & Busse, V. (2024). How effective is feedback for L1, L2, and FL learners' writing? A meta-analysis. Learning and Instruction, 93, Article 101961. https://doi.org/10.1016/j.learninstruc.2024.101961

Simoncini, K. M., Smith, H., Cain-Gray, L., & Sebalj, D. (2022). Books that tell my story: Transforming the attitudes of Australian preservice teachers towards children's diverse and multicultural literature. Australian Journal of Teacher Education, 47(9), 100–114. https://doi.org/10.14221/ajte.2022v47n9.6

Stoops, A., & Montag, J. L. (2024). A novel corpus of naturalistic picture book reading with 2-to-3 year old children. Language Development Research, 4(1). https://doi.org/10.34842/3kz6-4s17

Sudigdo, A., & Pamungkas, O. Y. (2022). Multiculturalism in children's literature: A study of a collection of poems by elementary school students in Yogyakarta. Daengku: Journal of Humanities and Social Sciences Innovation, 2(3), 266–278. https://doi.org/10.35877/454ri.daengku902

Sudigdo, A., Pamungkas, O. Y., Hastangka, Suprapto, Sumarni, Abdurrahman, O., & Mastiyah, I. (2024). Children's voices in poetry: The illustration and re-imagination of national identity. Journal of Language Teaching and Research, 15(5), 1581–1592. https://doi.org/10.17507/jltr.1505.19

Sun, L. (2024). Developing and sustaining critical intercultural competence through confronting hidden bias in children's literature in EFL preservice teacher education. Language and Intercultural Communication, 24(6), 710–725. https://doi.org/10.1080/14708477.2023.2291476

Throop Robinson, E. (2023). Count me in: Exploring equity, diversity, and inclusion through mathematics and children's literature. The Reading Teacher, 77(2), 189–198. https://doi.org/10.1002/trtr.2233

van Dijk, B., van Duijn, M., Verberne, S., & Spruit, M. (2023). ChiSCor: A corpus of freely-told fantasy stories by Dutch children for computational linguistics and cognitive science. Proceedings of the 27th Conference on Computational Natural Language Learning (CoNLL), 352–363. https://doi.org/10.18653/v1/2023.conll-1.23

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. Journal of the American Statistical Association, 22(158), 209–212. https://doi.org/10.1080/01621459.1927.10502953

Yang, J. S., Rosvold, C., & Bernstein Ratner, N. (2022). Measurement of lexical diversity in children's spoken language: Computational and conceptual considerations. Frontiers in Psychology, 13, Article 905789. https://doi.org/10.3389/fpsyg.2022.905789

Zenker, F., & Kyle, K. (2021). Investigating minimum text lengths for lexical diversity indices. Assessing Writing, 47, Article 100505. https://doi.org/10.1016/j.asw.2020.100505