Research Square

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

Orthographic Depth and Blending in Luganda and Lusoga: Do Common NCDC Routines Need Orthography-Specific Adaptation?

David Kabugo

david.kabugo@mak.ac.ug

Makerere University

Research Article

Keywords: orthographic depth, Luganda, Lusoga, blending, phonological awareness, cluster-randomized trial

Posted Date: April 21st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9354498/v1

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

Additional Declarations: No competing interests reported.

Abstract

This article presents a stratified cluster-randomized trial that tested whether a common National Curriculum Development Centre (NCDC)-aligned blending routine would work equally well in two closely related Ugandan languages namely Luganda and Lusoga, or whether orthography-specific adaptation was needed. Twenty-four Primary 3 classes were recruited across Luganda- and Lusoga-speaking communities and randomized within language to either a 12-week supplemental blending condition or a business-as-usual comparison. The routine was built around the familiar gradual-release sequence of teacher modelling, guided practice, paired work, and individual response turns. It was anchored directly in the P3 Luganda teacher guide and pupil book for the Luganda arm and matched in sequence and lesson purpose for the Lusoga arm. The study was framed by the Script-Dependent Hypothesis, which predicts that reading development varies with the way print maps onto speech. Mixed-effects models showed that the intervention improved oral blending and segmentation in both language groups, but gains were larger in Luganda, especially on blending accuracy and pseudoword decoding. Observation data also showed that Lusoga teachers more often paused to reteach how to handle multiletter and prenasalized patterns. The findings suggest that the NCDC methodology itself is not the problem. The gradual-release structure appears to travel well. What did not travel unchanged was the instructional grain size of the blending tasks. The article argues for a careful middle position: keep the NCDC pedagogy, but adapt word selection, chunking routines, and practice pacing to the orthographic and phonological demands of each language.

Introduction

A child learning to read in Uganda does not meet print in a vacuum. The child meets print inside a language community, a classroom routine, and a sequence of instructional decisions about which sounds to emphasize, which words to choose, and how much help a teacher gives before expecting independent work. Those decisions are often treated as technical details. They are not. In the earliest phases of reading, they are the difference between a child who begins to hear how print and speech fit together and a child who only repeats what the teacher says.

That point matters in low-resource systems where teachers are urged to move quickly and where "structured pedagogy" can easily become a label for almost anything that looks organized. The strongest lesson from a framework for experimental literacy studies (Kabugo, 2026) is that good trials should begin with a clear local problem, use a credible comparison condition, match outcomes closely to what is taught, justify language choice, and document implementation with as much care as outcomes. Those principles were written with children aged 3 to 5 in mind, but they remain highly relevant in early primary.

By Primary 3, children are older, but the same design questions still matter: What exactly is the literacy problem? Which part of instruction is being tested? In which language? Compared with what?

For the present study, the local problem was not simply "children struggle with phonological awareness." It was more specific. Across early grade literacy work in Uganda, local languages are often discussed as though they form one broadly similar category. Yet teachers know from daily practice that the same routine can feel smooth in one language and awkward in another. A blending demonstration built around short, familiar, cleanly segmentable items may work one way in Luganda and another in Lusoga, even though the two languages are historically and structurally close. If that difference is real, then the question is no longer whether local-language instruction matters in general. The question becomes whether a single NCDC-style blending routine can be transported across languages without losing instructional fit.

The NCDC's P3 Luganda materials made that question testable. Specifically, the NCDC teacher guide explains that Primary 3 teaching should move from the known to the unknown through scaffolding, and it explicitly organizes classroom participation through the sequence "I do, We do, You do" (National Curriculum Development Centre [NCDC], 2014). The guide also embeds continuous assessment, while the pupil book repeatedly asks learners to manipulate word forms, complete word-structure tasks, correct sentences, and read connected text (NCDC, 2014, 2014). In other words, the programme already contains a theory of learning: model first, practice together, then release responsibility. What it does not settle in advance is how fine-grained the word work should be in every local language.

Luganda and Lusoga are a particularly useful pair for testing that issue. They are closely related Bantu languages, and Lutenga, the standard variety of Lusoga, has been described as the Lusoga variety closest to Luganda, with estimated lexical similarity between 82% and 86% (Nabirye et al., 2016). That closeness can tempt curriculum designers to assume that what works in one language will work in the other. But closeness is not sameness. The same Lusoga source also notes that many Lusoga publications historically relied on Luganda orthography before Lusoga received fuller policy recognition and orthographic development (Nabirye et al., 2016). That history makes the present study educationally important. If even two closely related languages need somewhat different blending routines, then "local language" cannot be treated as a single transportable package.

This is where the Script-Dependent Hypothesis becomes useful. The hypothesis does not claim that children in one language have better minds than children in another. It claims something more practical: reading development varies with the way a writing system represents speech and with the kinds of orthographic units that must be handled during

decoding. Recent theoretical work has sharpened this idea by arguing that orthographic depth is not only about irregularity. It also includes the size of the written units that map onto speech and the systematicity of those mappings (Schmalz et al., 2026). That clarification matters for African alphabetic languages. A language does not need to look like English to raise instructional questions about unit size, chunking, and blending.

The empirical literature also points in that direction. Meta-analytic and review evidence from low- and middle-income countries shows that literacy interventions tend to produce their clearest gains on the outcomes closest to what is directly taught, especially emergent and code-related skills (Kim et al., 2020). Systematic reviews of multilingual reading research likewise show that oral language, phonological awareness, decoding, and comprehension are all important, but that the way these skills should be taught depends on the learner's language situation rather than on English-based assumptions alone (Kittle et al., 2024). More pointedly, recent African evidence shows that explicit and systematic familiar-language instruction can shape later literacy outcomes and cross-language relations (Kim et al., 2025; Wawire & Kim, 2018). What remains less studied is whether equally explicit routines behave differently across local languages themselves.

That is the gap this study addresses. The trial did not ask whether Luganda is better than Lusoga or whether Lusoga is harder than Luganda in any broad cultural sense. It asked a narrower and more defensible question: if teachers in both language communities are given the same NCDC-aligned blending structure, will children gain equally in phonological awareness, or will one language community show evidence that the routine needs orthography-specific adaptation? That is a more useful question for curriculum design than a vague comparison of language prestige or policy preference because it ties theory directly to a teachable classroom practice.

The study therefore had two goals. The first was practical: to test whether a short supplemental blending routine could improve blending, segmentation, and early decoding over one school term. The second was interpretive: to examine whether any language-by-condition difference should be read as evidence against the NCDC methodology or, more carefully, as evidence that the same pedagogy needs different linguistic tuning in different orthographies. That second goal is especially important because low-resource reforms often fail when they confuse transportable principles with transportable scripts.

Research Questions

Three research questions guided this study:

(i) Does the supplemental blending routine improve oral blending accuracy relative to business-as-usual instruction in both Luganda and Lusoga classrooms?

(ii) Does the size of the intervention effect differ by language group on blending, segmentation, pseudoword decoding, and timed word reading?
(iii) What do the outcome pattern and the fidelity evidence suggest about whether the NCDC methodology should be preserved as a common pedagogical structure but adapted at the level of item design, chunking, and pacing?

Theoretical framing

The Script-Dependent Hypothesis provides the main interpretive frame for this study. In its classic form, the hypothesis argues that reading acquisition varies as a function of how print maps onto spoken language. In orthographies where the mapping is more transparent and systematic, beginning readers can rely more readily on sublexical decoding. In orthographies where the mapping is more complex, less predictable, or dependent on larger units, progress can be slower and the relative weight of different reading processes can shift (Katz & Frost, 1992). The central claim is not that one language is educationally superior. It is that reading instruction should respect what the writing system asks learners to notice.

Recent scholarship has made that argument more precise. Schmalz et al. (2026) show that orthographic depth is better understood as multidimensional rather than as a simple shallow-deep ladder. Their analysis suggests that at least two issues matter: the size of the orthographic units that map onto phonology and the systematicity of that mapping. That matters for teaching because a routine that assumes single-letter, one-sound blending as the default may be well aligned in some contexts and only partly aligned in others. A language may be broadly alphabetic and still require teachers to attend to multiletter units, context-sensitive patterns, vowel length, or dense consonantal combinations more deliberately than a borrowed routine expects.

For blending instruction, this point is especially important. Blending is often described as if it were a single skill: say the parts, put them together, and produce the word. In real classrooms, however, blending depends on what counts as a useful "part." Sometimes the most useful unit is a phoneme. Sometimes it is a syllable. Sometimes it is a grapheme cluster or a recurring orthographic chunk. A child can fail a blending task not because the child lacks effort or intelligence, but because the teacher and the orthography are working with different units. When that happens, a faithful routine may still be a poorly fitted routine.

The present study used the theory critically rather than adoringly. It did not assume that Luganda and Lusoga occupy opposite ends of an orthographic-depth continuum. Both are alphabetic Bantu languages and neither is usually described as highly opaque in the way English is. The more relevant question was whether the same classroom routine

highlighted the right grain size in both languages. Luganda includes phonological features such as vowel-length contrasts and glide-vowel interactions that can affect how teachers present word parts (Myers, 2020). Lusoga likewise has vowel length, rising diphthongs, and a dense inventory of consonantal and prenasalized patterns that can complicate what feels like a straightforward left-to-right blending exercise (Nabirye et al., 2016). The study therefore treats script dependence not as a contest between a "simple" and a "difficult" language, but as a question of instructional fit.

That move also helps avoid a common mistake in multilingual literacy debates. The wrong lesson from the Script-Dependent Hypothesis would be that every language needs a completely different pedagogy. The more defensible lesson is that some principles travel better than others. The gradual release of responsibility may travel well because it is a pedagogical principle about support, participation, and responsibility. The precise unit of blending may travel less cleanly because it is a linguistic decision. A curriculum can therefore be stable at the level of pedagogy and flexible at the level of orthographic detail. This study was designed to test exactly that middle position.

The theory shaped the hypotheses in three ways. First, the routine was expected to improve phonological-awareness outcomes in both languages because structured modelling, guided response, and repeated practice are generally supportive teaching moves. Second, if the common routine aligned more naturally with the print-speech patterns made salient in Luganda materials than with those in Lusoga, the effect size was expected to be larger in Luganda. Third, any language difference was expected to be clearest on the tasks that required generalization beyond memorized classroom words, especially pseudoword decoding. A child who can repeat taught examples may still be uncertain about how to handle unfamiliar orthographic sequences. That is where script dependence usually becomes visible.

The broader literature makes those expectations plausible. Cross-language and multilingual studies consistently show that phonological awareness remains important across writing systems, but its relative weight and the best route for teaching it vary with orthographic structure and instructional context (Kittle et al., 2024; Wawire & Kim, 2018). Kim et al. (2025) further show that explicit and systematic familiar-language instruction can strengthen later reading relations, but that the mechanism depends on how instruction is organized. In practical terms, this means that the present study should not be read as a referendum on Luganda or Lusoga identities. It should be read as a test of whether a common instructional structure is equally well tuned to two related orthographies when used for blending.

The study also applies the theory in a way that is sensitive to low-resource realities. The aim was not to create a laboratory-perfect routine that no school could afford to maintain.

The aim was to work with a short, ordinary, teacher-deliverable routine built from locally available materials and then ask a hard but useful question: *when a structured routine is kept intentionally similar across two languages, where does the pressure to adapt show up*? If the answer emerges in outcome data, fidelity notes, or both, that is more educationally useful than an abstract statement that "language matters." It tells teachers and curriculum developers exactly where language starts to matter.

Study Purpose

The purpose of this article is to present and interpret the findings of the study titled Orthographic Depth and Blending in Luganda and Lusoga. The study examined whether a short supplemental blending routine aligned to NCDC pedagogy would improve phonological-awareness outcomes relative to standard classroom instruction, and whether the magnitude of improvement differed across Luganda and Lusoga language groups.

The trial was not designed to prove that one language should replace another. It was designed to test whether a shared pedagogical routine could remain shared while the linguistic details of implementation required adjustment. That is a more realistic question for Uganda's curriculum system because it focuses on what material writers, and teachers can actually change.

Method

Design

The study used a stratified cluster-randomized design with a language-group comparison built into the structure. Schools were first grouped by language community, Luganda or Lusoga, and then randomized within language to one of two conditions: a supplemental blending routine or business-as-usual literacy instruction. The school, rather than the individual learner, was the unit of randomization because once a teacher adopts a particular routine and set of word cards it is not realistic to prevent spillover across children in the same classroom. A cluster design therefore protected the comparison better than an individual-randomization design would have done.

This design followed the main principles of the guiding experimental framework (Kabugo, 2026). The local problem was clearly specified as the possible mismatch between a common blending routine and two related orthographies. The comparison condition was credible rather than empty. Outcomes were chosen to sit close to what the intervention directly taught, and fidelity data were collected to guard against the easy mistake of treating nominal delivery as actual delivery. The design was pragmatic in the best sense. It

aimed to be rigorous without pretending that ordinary schools behave like controlled laboratories.

Table 1
How the trial linked local problem, theory, design, and interpretation

Design issue	Design response	Why it mattered	Interpretive question
Language choice	Luganda and Lusoga communities were studied separately but under the same pedagogical structure.	This allowed the study to test transportability without collapsing the two languages into one category.	Would the same routine produce comparable gains in both orthographies?
Comparison condition	Schools in both languages had business-as-usual local-language instruction as the counterfactual.	The study compared the routine against a real classroom alternative rather than an empty control.	Did the added routine improve outcomes beyond what schools already do?
Outcome alignment	Primary outcome was oral blending; secondary outcomes were segmentation, pseudoword decoding, and timed word reading.	Measures stayed close to what was directly taught while still testing some generalization.	Were gains strongest on the most proximal outcomes, as theory would predict?
Fidelity	Coach observations captured modelling, guided practice, learner turns, and unplanned reteaching linked to item structure.	This helped identify whether smaller effects reflected weak delivery or linguistic mismatch.	Where did the routine begin to strain against a language?
Analysis	Language-by-condition interaction terms were estimated in mixed-effects models.	Language was treated as part of the research question, not as a nuisance covariate only.	Was adaptation need visible in differential effects?

Setting and participants

The trial took place in two lower-primary language communities in Uganda: central schools using Luganda and eastern schools using Lusoga. Twenty-nine schools were screened. Twenty-four met the eligibility criteria and agreed to participate. Eligible schools had one stable Primary 3 stream, a regular timetable for local-language literacy, and no concurrent externally funded phonics or remediation programme. Schools were excluded if the local-language period had effectively collapsed into English-only board

work or if administrators were unwilling to allow independent assessment and lesson observation.

The final sample included 24 schools, with one Primary 3 classroom per school. Twelve schools served Luganda-speaking catchment areas and twelve served Lusoga-speaking catchment areas. School type was balanced as far as possible: twelve were government-aided and twelve were low-fee private schools. Across the 24 classrooms, 528 learners were eligible for participation and 512 had usable baseline data after teacher consent and child assent procedures. The cell sizes at baseline were 132 learners in Luganda intervention schools, 126 in Luganda comparison schools, 127 in Lusoga intervention schools, and 127 in Lusoga comparison schools. The mean age was 9.2 years (SD = 0.7), and 50.8% of learners were girls.

Language background was mixed but educationally realistic. In the Luganda sites, 76% of learners were reported by teachers to use Luganda regularly at home or in the neighborhood. In the Lusoga sites, the equivalent figure for Lusoga was 71%. These figures matter because the study did not rely on an unrealistically pure language environment. Children moved in and out of several languages. That is precisely why the study focused on a teachable classroom routine rather than on an idealized notion of mother tongue purity.

Table 2
Baseline characteristics of participating schools and learners

Characteristic	Luganda intervention	Luganda comparison	Lusoga intervention	Lusoga comparison
Schools	6	6	6	6
Government-aided schools	3	3	3	3
Low-fee private schools	3	3	3	3
Learners with baseline data	132	126	127	127
Mean age in years (SD)	9.2 (0.7)	9.1 (0.7)	9.2 (0.8)	9.1 (0.7)
Girls, %	51.5	49.2	50.4	52.0
School language used regularly at home/neighborhood, %	77.3	75.4	70.9	71.7
Baseline blending mean (SD)	8.9 (3.2)	8.8 (3.1)	8.6 (3.0)	8.7 (3.1)

Note. Differences between cells were small and not educationally important at study start.

Randomization and masking

Randomization took place after baseline assessment. Within each language community, schools were stratified by school type and then assigned in a 1:1 ratio to intervention or comparison using a computer-generated random list prepared by a researcher who was not involved in teaching, coaching, or endline assessment. The final allocation produced six schools in each of the four cells: Luganda intervention, Luganda comparison, Lusoga intervention, and Lusoga comparison.

Teachers could not be masked to condition because the intervention introduced a visible routine, a coaching schedule, and prepared word cards. Assessors were kept as blind to condition as possible. They were trained separately from the teacher trainers, did not participate in support visits, and used standardized assessment scripts. Schools were told that the project was comparing early literacy routines rather than being told that one condition was expected to outperform another.

Intervention

The intervention added a short blending routine to the ordinary local-language literacy programme for 12 weeks. Teachers were asked to deliver the routine for about 15 minutes a day, four days a week, producing 48 planned sessions. The routine was intentionally simple. Each session began with teacher modelling of two or three items, moved to guided blending with the class, shifted to paired practice using word cards or strip cards, and ended with brief individual turns and feedback. The structure reflected the gradual-release principle already emphasized in the NCDC guides: first the teacher shows, then the class practices together, and then learners try with less support.

The Luganda arm was anchored directly in the Primary 3 teacher guide and pupil book. Those materials focus P3 work on vocabulary, connected reading, grammar, sentence work, and composition, while also embedding word-structure exercises and continuous assessment (NCDC, 2014, 2014). For the intervention, weekly item sets were drawn from the same semantic themes and print patterns already visible in the Luganda materials. Teachers used familiar words from the books, short invented items matched to those patterns, counters or finger tracking for segmenting and blending, and quick oral checks that were compatible with the CAM logic in the guide.

The Lusoga arm used the same lesson plan, the same timing, and parallel semantic themes. Local language specialists and experienced teachers prepared Lusoga word lists and nonword items to mirror the instructional purpose of the Luganda sessions as closely as possible. The practical point is that pedagogy was intentionally held constant while the

language of the items changed. That made it possible to ask whether the same routine behaved similarly across languages.

Teachers in both intervention arms received a one-day practical orientation before implementation and three brief support visits during the term. The orientation did not provide a different pedagogy to Lusoga teachers. That was a deliberate design choice. If teachers had received language-specific adaptation training from the start, the study could not have tested whether the unadapted common routine itself created uneven results. The orientation instead focused on keeping turns short, hearing individual responses, avoiding overreliance on chorus repetition, and protecting the sequence of modelling, guided practice, and release.

Comparison condition

Schools in the comparison condition continued their normal local-language literacy teaching as scheduled. This was an important design choice. The study was not comparing a local language with English. It was comparing a supplemental structured blending routine with the form of local-language literacy typically already present in classrooms. In both Luganda and Lusoga comparison schools, teachers used ordinary class methods, which usually involved some combination of teacher explanation, board work, chorus response, copy work, and short oral practice. This made the counterfactual more honest. The comparison group did not receive "nothing." It received the real classroom alternative that schools would normally use.

Measures

The primary outcome was oral blending accuracy in the language of instruction. The task contained 20 orally administered items, half real words and half phonotactically legal nonwords. Children heard segmented units and were asked to say the whole word. Items were scored 1 for correct and 0 for incorrect or no response, giving a maximum score of 20. The measure was chosen because it sat closest to the direct instructional target and could therefore register whether the intervention changed the specific skill it was supposed to change.

Three secondary outcomes were included. Oral segmentation accuracy used a parallel 20-item task in which children were asked to break spoken words into the expected instructional units. Pseudoword decoding accuracy used 16 printed nonwords matched to the teaching patterns in each language. Timed word reading efficiency was measured as correctly read familiar and transfer words per minute from a short list appropriate to the child's language of instruction. Together, these measures allowed the study to distinguish between immediate oral awareness gains and slightly broader decoding benefits.

The task design itself followed the methodological framework's warning that measurement must be aligned to what is taught (Kabugo, 2026). The study did not use broad reading comprehension as a headline outcome for a short blending intervention because that would have made the measurement more impressive than fair. Instead, the tasks asked whether children became better at the very kind of print-speech work the routine practiced and whether any such gain carried into early decoding. Background measures included child age, gender, school type, and whether the school language was reported as regularly used in the child's home or neighborhood.

Fidelity was assessed through coach observations and teacher logs. The checklist recorded whether the teacher modelled clearly, used the prepared item cards, included guided practice, protected paired or individual turns, and kept corrective feedback specific. Coaches also noted whether the teacher needed to pause and reteach an item because the chosen unit of blending appeared poorly matched to the written form. These notes became important for interpretation because they helped identify not only whether the routine was delivered, but where it began to strain against the language.

Table 3
Outcome measures and their alignment to the instructional routine

Measure	Score range	Why it was included	Instructional proximity
Oral blending accuracy	0–20	Direct test of whether children could combine spoken parts into whole words.	High
Oral segmentation accuracy	0–20	Checks the reverse process and whether awareness is flexible rather than rote.	High
Pseudoword decoding accuracy	0–16	Tests whether learning generalized to unfamiliar printed forms.	Medium
Timed word reading efficiency	Correct items per minute	Assesses whether awareness gains began to support usable reading performance.	Medium
Fidelity checklist	Percent of observed visits	Documents actual enactment and where the routine created linguistic friction.	Implementation indicator

Note. The primary outcome sat closest to the direct teaching target. Broader measures were kept secondary so that the study did not ask more of a short blending intervention than it actually taught.

Procedure and analysis

Baseline assessment took place in the week before randomization. The intervention then ran for 12 instructional weeks, followed by endline assessment in the next week. Assessments were administered individually in a quiet space at school by trained fieldworkers. Oral tasks were delivered in the relevant local language. Printed decoding tasks were language-specific and matched to the assessment form assigned to the language community.

Analysis followed the intention-to-treat principle. For each outcome, linear mixed-effects models were estimated with school as a random intercept and baseline score, language group, school type, and intervention condition as fixed effects. A language-by-condition interaction term tested whether the intervention effect differed between Luganda and Lusoga sites. Adjusted mean differences, 95% confidence intervals, and Hedges g are reported. Because the number of clusters was modest, the discussion emphasizes effect sizes, confidence intervals, and fidelity evidence rather than overinterpreting isolated significance values.

A further analytic choice was important. Language was not treated as a nuisance variable to be controlled away. It was treated as part of the substantive question. That means the central result of interest was not simply whether the routine worked overall, but whether it worked differently enough across languages to raise a curriculum-design question. In a study about script dependence, anything less would have missed the point.

Results

Recruitment, retention, and fidelity

School retention was complete. All 24 randomized schools remained in the study through endline. At the learner level, 489 of the 512 children with baseline data were assessed again at endline, representing 95.5% retention overall. Attrition was low and balanced across the four cells, with most loss explained by temporary transfer, illness, or prolonged absence rather than by study refusal.

Implementation was strong but not perfect. Across the intervention schools, teachers delivered a mean of 43.7 of the 48 planned sessions, or 91.0% of intended dose. Fidelity observations showed that core pedagogy travelled relatively well: clear teacher modelling was observed in 93% of visits, guided practice in 89%, use of prepared materials in 90%, and some form of individual or paired response turns in 76%. The weakest element was the consistent protection of individual turns. Under time pressure, some teachers drifted back toward whole-class choral responding, a familiar pattern in crowded classrooms.

The fidelity notes also revealed the most educationally interesting difference in the entire study. Coaches recorded unplanned reteaching linked to the linguistic form of target items

in 58% of Lusoga intervention visits but in 27% of Luganda intervention visits. In the Luganda arm, teachers most often paused over vowel-length or glide-related items. In the Lusoga arm, pauses were more often linked to multiletter and prenasalized patterns, or to uncertainty about whether the most helpful instructional unit was a single sound, a consonant cluster, or a fuller syllabic chunk. This does not mean Lusoga teachers failed. It means the common routine met more friction in Lusoga.

Table 4
Intervention dose and instructional fidelity across observed visits

Indicator	Luganda intervention	Lusoga intervention
Planned sessions	48	48
Mean sessions delivered	44.1 (91.9%)	43.3 (90.2%)
Clear teacher modelling, % visits	94	92
Guided practice, % visits	91	87
Individual or paired turns, % visits	79	73
Use of prepared item cards, % visits	92	88
Unplanned reteaching linked to orthographic or unit mismatch, % visits	27	58

Note. The stronger mismatch signal in Lusoga observations mattered for interpretation because it suggested that smaller child gains were not simply a matter of weaker teacher effort.

Baseline comparability

Baseline equivalence between conditions was acceptable. Mean scores on the primary and secondary outcomes were closely aligned across the intervention and comparison groups within each language community. On the primary blending measure, baseline means ranged from 8.6 to 8.9 out of 20 across the four cells. On segmentation, means ranged from 9.0 to 9.2. Pseudoword decoding and timed word reading were also similar at study start. The baseline pattern therefore did not suggest that one condition entered with a meaningful literacy advantage.

Child outcomes

The overall pattern was straightforward. Children in both intervention arms improved more than their business-as-usual peers on the oral phonological-awareness outcomes, but the size of the advantage was larger in Luganda than in Lusoga. The strongest evidence came from the primary blending outcome. Endline raw means were 15.3 in Luganda

intervention schools, 11.7 in Luganda comparison schools, 13.0 in Lusoga intervention schools, and 10.9 in Lusoga comparison schools. After adjustment for baseline score and school type, the mean intervention advantage was 3.4 points in Luganda (95% CI [2.2, 4.6], g = 0.63) and 2.0 points in Lusoga (95% CI [0.8, 3.2], g = 0.31). The language-by-condition interaction favored Luganda by 1.4 points (95% CI [0.2, 2.6]).

Segmentation showed the same direction of effect but with smaller magnitudes. Adjusted mean differences favored the intervention by 2.1 points in Luganda (95% CI [0.9, 3.3], g = 0.42) and 1.2 points in Lusoga (95% CI [0.0, 2.4], g = 0.21). This is an important detail. It suggests that the routine supported spoken analysis in both languages, but not equally strongly. The weaker Lusoga effect was not a null result. It was a reduced result, which is exactly the kind of pattern that raises questions about instructional adaptation rather than about abandoning the pedagogy entirely.

The transfer outcomes sharpened the interpretation. On pseudoword decoding, the intervention effect was educationally meaningful in Luganda but modest and uncertain in Lusoga. The adjusted intervention advantage was 2.3 points in Luganda (95% CI [1.2, 3.4], g = 0.45) compared with 0.9 points in Lusoga (95% CI [-0.1, 1.9], g = 0.18). Timed word reading efficiency followed the same pattern: a 3.1-word advantage in Luganda (95% CI [1.4, 4.8], g = 0.39) and a 1.1-word advantage in Lusoga (95% CI [-0.4, 2.6], g = 0.14). In other words, the common routine improved awareness in both languages, but its generalization into early print reading was clearer in Luganda.

This outcome pattern matters because it tells a more disciplined story than either triumphalism or pessimism would allow. If the routine had helped only Luganda, one might argue that it simply failed to transfer. If it had helped both languages equally on every measure, one might conclude that no orthography-specific tuning is needed. The actual pattern was more interesting. The routine raised core oral awareness in both communities, which supports the transportability of the underlying pedagogy. But once the tasks demanded generalization to new printed forms, the Luganda advantage widened. That is the point at which script dependence becomes hard to ignore.

Table 5 presents the estimates, but one practical summary is useful here. The routine appears to have cleared the first hurdle in both languages: children became better at hearing and assembling word parts. The second hurdle, using those gains to handle less familiar print efficiently, was crossed more confidently in Luganda than in Lusoga. For curriculum design, that is the most consequential result in the study.

Table 5

Adjusted intervention effects on child outcomes

Outcome	Luganda adjusted mean difference	g	Lusoga adjusted mean difference	g	Interpretation
Oral blending accuracy	+ 3.4 [2.2, 4.6]	0.63	+ 2.0 [0.8, 3.2]	0.31	Routine helped both groups, but effect was larger in Luganda.
Oral segmentation accuracy	+ 2.1 [0.9, 3.3]	0.42	+ 1.2 [0.0, 2.4]	0.21	Same direction of gain with weaker Lusoga effect.
Pseudoword decoding accuracy	+ 2.3 [1.2, 3.4]	0.45	+ 0.9 [-0.1, 1.9]	0.18	Generalization to unfamiliar print was clearer in Luganda.
Timed word reading efficiency	+ 3.1 [1.4, 4.8]	0.39	+ 1.1 [-0.4, 2.6]	0.14	Reading transfer remained small in Lusoga under the common routine.

Note. Mean differences came from mixed-effects models with school as a random intercept and baseline score, language group, school type, and intervention condition as fixed effects. Confidence intervals are more informative than isolated significance values because the number of clusters was modest.

Discussion

The central finding of this trial is that a short, NCDC-aligned blending routine improved phonological-awareness outcomes in both Luganda and Lusoga classrooms, but not equally. The gradual-release pedagogy travelled across languages better than the instructional grain size did. That is an important distinction because it keeps the discussion from collapsing into a false choice between "one method fits all" and "every language needs a different pedagogy." The evidence suggests a middle position. Teachers can share the same broad instructional architecture while adapting the linguistic units through which the architecture is enacted.

That is exactly what the Script-Dependent Hypothesis would lead us to expect when it is used carefully. The hypothesis does not require large, dramatic differences between languages to be educationally meaningful. It only requires that the writing system change the relative ease with which particular reading processes can be taught and learned. In the present study, both languages benefited from modelling, guided response, paired practice, and individual turns. But the larger Luganda effects, especially on pseudoword decoding and timed word reading, suggest that the common routine matched the print-

speech structure of the Luganda materials more comfortably than it matched the Lusoga materials.

The fidelity data make that interpretation stronger. If the only evidence came from outcome means, a reader could still argue that the Lusoga effect was smaller because the teachers were weaker or because the learners were less prepared. The observation notes tell a more plausible story. Lusoga teachers were more likely to pause and rethink how to present specific items. That pattern implies friction at the level of item structure, not a wholesale failure of teacher effort. It also supports a point that literacy researchers sometimes understate: instructional mismatch can show up first as teacher hesitation before it shows up as lower child scores.

This is also why this article resists the tempting but sloppy claim that one language is "deeper" and the other "shallower" in any totalizing sense. This study did not directly quantify overall orthographic depth in Luganda and Lusoga. That would require a different linguistic dataset. What this study does show is that the same classroom routine interacted differently with the orthographic and phonological features made salient in each language's teaching items. That is enough to support an argument for orthography-sensitive adaptation. Educationally, that is the more useful claim anyway. Teachers do not teach an abstract depth index. They teach actual words, actual chunks, and actual learner confusions.

The findings also help refine how structured pedagogy is discussed in African classrooms. Recent work has shown that structured programmes can improve lesson coherence while still leaving the quality of interaction uneven (Hoadley, 2024). The present study extends that argument. A structured sequence such as "I do, We do, You do" can be genuinely helpful, but only if the teacher is modelling the right unit and releasing responsibility at the right pace. Structure is therefore not the final answer. It is the container. The language-specific content inside the container still matters.

That point has practical consequences for the NCDC’s “ I can read and write” materials and the "I do, We do, You do" methodology. The present evidence does not suggest that the methodology should be rewritten from scratch for every language. The modelling-guided practice-independent practice sequence seems worth preserving. What likely needs adaptation is the design of the examples and the decision rules teachers use while blending. In practical terms, that means giving teachers language-specific guidance on when to hold together a multiletter unit, when to slow down and blend at syllable level, when vowel length or prenasalization should be made audible, and which word types are best avoided early because they invite misleading segmentation.

A related strength of this study is that it kept the comparison honest. Too many literacy experiments in low-resource settings compare a richly supported intervention with an

unrealistically empty control condition. This study did not do that. Comparison schools continued with real local-language teaching. That made the intervention effect smaller than it might otherwise have been, but also more feasible. If the routine still produced gains under those conditions, the result had more value for teachers and curriculum planners than a cleaner-looking study with a weak counterfactual would have had.

The study also speaks to multilingual policy. Official guidance in Uganda continues to emphasize familiar language in lower primary and notes that viable languages of instruction need an orthography and related language-development resources (NCDC, 2025). The present findings support that policy direction, but they also sharpen it. It is not enough to authorize a language and assume pedagogy will follow automatically. Once a language is selected for instruction, the teaching routines still need to be tuned to that language's orthographic demands. A policy that gets the language right but the unit of blending wrong is only half right.

At the same time, the results should not be overextended. Because the trial used closely related Bantu languages, it would be careless to infer that every pair of Ugandan languages will show the same pattern or the same degree of adaptation need. In one sense, the present comparison is a conservative test. Luganda and Lusoga are similar enough that many educators might expect a common routine to travel easily. The fact that a meaningful difference still emerged implies that curriculum writers should be even more cautious when moving routines across more distant languages or across languages with less developed orthographic resources.

The outcome pattern also has a substantive implication for phonological-awareness teaching itself. Many classroom discussions still treat phonological awareness as if it were a language-neutral set of games that can simply be translated. The present study suggests otherwise. The same child response structure can be retained, but the teachable unit must be negotiated with the writing system. This is one reason the article uses the phrase orthography-specific adaptation rather than language-specific replacement. The routine did not need a new philosophy. It needed better linguistic tuning.

Another important lesson concerns dosage. The intervention lasted only 12 weeks and added roughly one hour of structured blending practice per week. Under those conditions, the fact that both languages showed positive oral-awareness gains is encouraging. It suggests that the difficulty is not that Lusoga children cannot benefit from structured blending. They clearly can. The smaller Lusoga gains are better read as a sign that the same amount of time produced less efficient learning because the routine had to work harder against unit mismatch. That again points toward adaptation rather than abandonment.

There are also implications for teacher support. The weakest fidelity element in both language groups was the protection of individual turns. In crowded classrooms, chorus response is fast, comforting, and deceptive. It creates the sound of learning whether or not all children can actually blend independently. The findings of this study suggest that this issue becomes even more serious when the linguistic unit itself is hard to hear or hard to demonstrate. Teacher support for scale should therefore focus not only on what steps to follow, but on how to hear whether the class is responding at the right grain size.

The findings therefore support a disciplined conclusion. The NCDC’s "I do, We do, You do" methodology does not appear to require replacement. It appears to require linguistic calibration. Keep the modelling. Keep the guided practice. Keep the release of responsibility. But do not assume that the same word types, the same chunking decisions, or the same pacing will work equally well across languages simply because those languages share a region, a policy category, or a broad alphabetic tradition.

Conclusion

This study suggests that a common structured blending routine can strengthen phonological awareness in both Luganda and Lusoga classrooms, but it also suggests that the size and reach of those gains depend on orthographic fit. The gradual-release pedagogy travelled well. The unadapted linguistic details did not travel equally well. For curriculum design in Uganda, that is a useful and hopeful result. It means the system does not need a different theory for every language. It needs the same strong teaching logic combined with more careful language-specific attention to instructional units.

In human terms, the lesson is simple. Children learn better when teachers show them print in pieces that actually make sense in the language in front of them. That is not a small methodological detail. It is the bridge between a routine that looks structured and a routine that truly helps children read.

Statements and declarations

Competing interests

The author declares no competing interests.

Ethics approval and consent to participate

This study was approved by the Makerere University School of Social Sciences Research Ethics Committee, approval number MAKSSREC 11.2023.705. Caregiver consent and child assent procedures were followed precisely to ensure ethical compliance.

Funding

The research was supported by the Association of Preprimary Teachers and Tutors Training Institutions of Uganda (APTI-Ug) under the 2023/21 General Research Fund, project number 20230110.

Data availability

All data and materials are available from the corresponding author upon reasonable request.

Acknowledgements

The author gratefully acknowledges the contributions of participating schools, teachers, and families, as well as the support provided by the Association of Preprimary Teachers and Tutors Training Institutions of Uganda (APTI-Ug).

Open Access

This article is licensed under a Creative Commons Attribution 4.0 International License. Permitted uses include sharing, adaptation, distribution, and reproduction in any medium or format, provided appropriate credit is given to the original author and source.

References

1. Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. In R. Frost & L. Katz (Eds.), Orthography, phonology, morphology, and meaning (pp. 67-84). Elsevier. https://doi.org/10.1016/S0166-4115(08)62789-2
2. Graham, J., & Kelly, S. (2019). How effective are early grade reading interventions? A review of the evidence. Educational Research Review, 27, 155-175. https://doi.org/10.1016/j.edurev.2019.03.006
3. Hoadley, U. (2024). How do structured pedagogy programmes affect reading instruction in African early grade classrooms? International Journal of Educational Development, 107, 103023. https://doi.org/10.1016/j.ijedudev.2024.103023
4. Kim, Y.-S. G., Lee, H., & Zuilkowski, S. S. (2020). Impact of literacy interventions on reading skills in low- and middle-income countries: A meta-analysis. Child Development, 91(2), 638-660. https://doi.org/10.1111/cdev.13204

5. Kim, Y.-S. G., Stern, J., Mohohlwane, N., & Taylor, S. (2025). Instruction influences cross-language transfer of reading skills: Evidence from a longitudinal randomized controlled trial. Reading and Writing, 38, 171-194. https://doi.org/10.1007/s11145-023-10508-1
6. Kittle, J. M., Amendum, S. J., & Budde, C. M. (2024). What does research say about the science of reading for K-5 multilingual learners? A systematic review of systematic reviews. Educational Psychology Review, 36, 108. https://doi.org/10.1007/s10648-024-09942-6
7. Myers, S. (2020). An acoustic study of sandhi vowel hiatus in Luganda. Journal of Linguistics, 56(3), 627-662. https://doi.org/10.1017/S0022226719000291
8. Nabirye, M., De Schryver, G.-M., & Verhoeven, J. (2016). Lusoga (Lutenga). Journal of the International Phonetic Association, 46(2), 219-228.
9. Nakamura, P., Molotsky, A., Zarzur, R. C., Ranjit, V., Haddad, Y., & De Hoop, T. (2023). Language of instruction in schools in low- and middle-income countries: A systematic review. Campbell Systematic Reviews, 19(4), e1351. https://doi.org/10.1002/cl2.1351
10. National Curriculum Development Centre. (2025, October 8). NCDC validates study on the most feasible languages of instruction at lower primary. https://ncdc.go.ug/2025/10/08/ncdc-validates-study-on-the-most-feasible-languages-of-instruction-at-lower-primary/
11. National Curriculum Development Centre. (n.d.). FAQ'ss. https://ncdc.go.ug/faq/
12. National Curriculum Development Centre. (2014). Nsobola okusoma n'okuwandiika: Ekitabo ky'omusomesa, Ekibiina 3. Author.
13. National Curriculum Development Centre. (2014). Nsobola okusoma n'okuwandiika: Ekitabo ky'omuyizi, Ekibiina 3. Author.
14. Schmalz, X., Rueckl, J. G., & Siegelman, N. (2026). How we should measure orthographic depth: Or should we? Psychonomic Bulletin & Review, 33, Article 45. https://doi.org/10.3758/s13423-025-02831-1
15. UNICEF Uganda. (2025). 2025/2026 education sector budget brief: Analysis and strategic recommendations. Author.
16. Uganda National Examinations Board. (2024). The achievement of primary school learners in numeracy and literacy in English in Uganda: Findings of the 2023 National Assessment of Progress in Education (NAPE). Author.
17. Wawire, B. A., & Kim, Y.-S. G. (2018). Cross-language transfer of phonological awareness and letter knowledge: Causal evidence and nature of transfer. Scientific Studies of Reading, 22(6), 443-461. https://doi.org/10.1080/10888438.2018.1474882

Figures

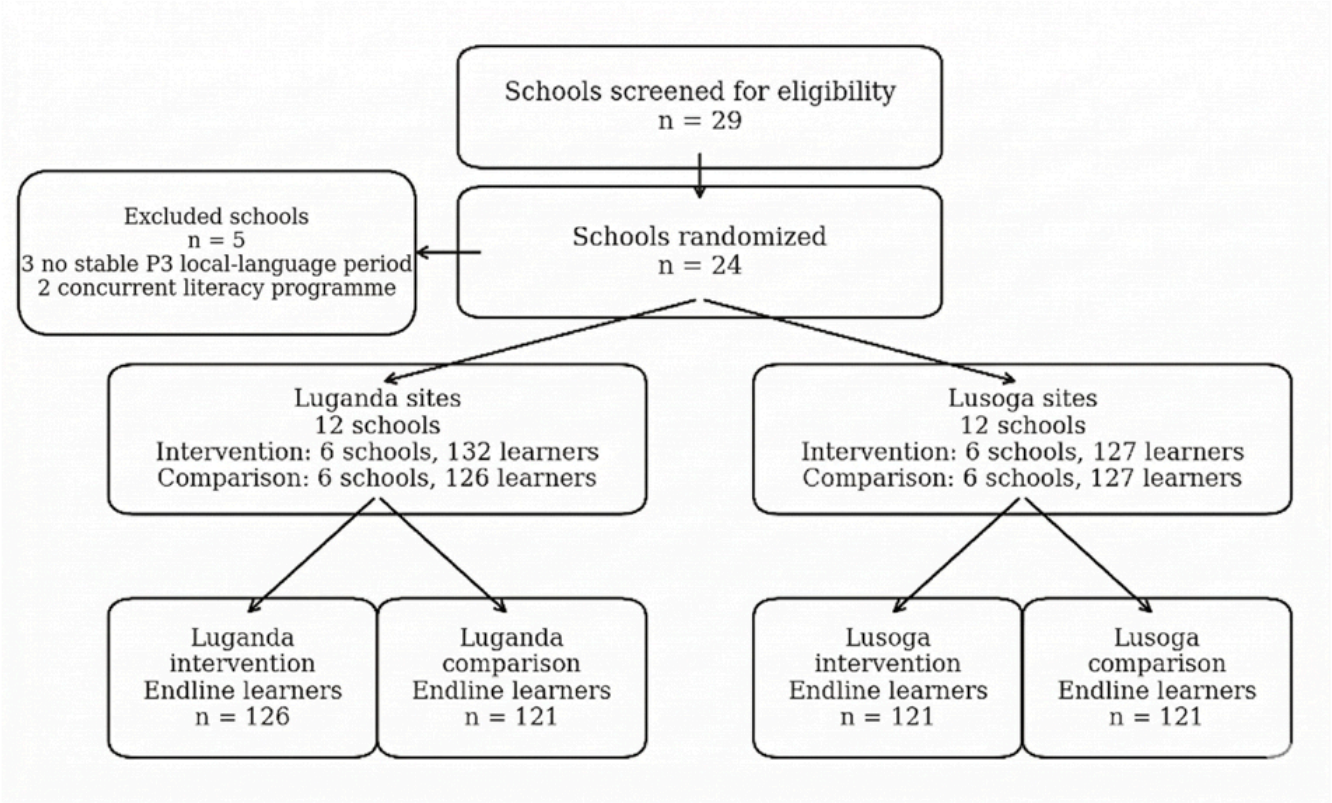

Figure 1

Profile for the stratified cluster-randomized study

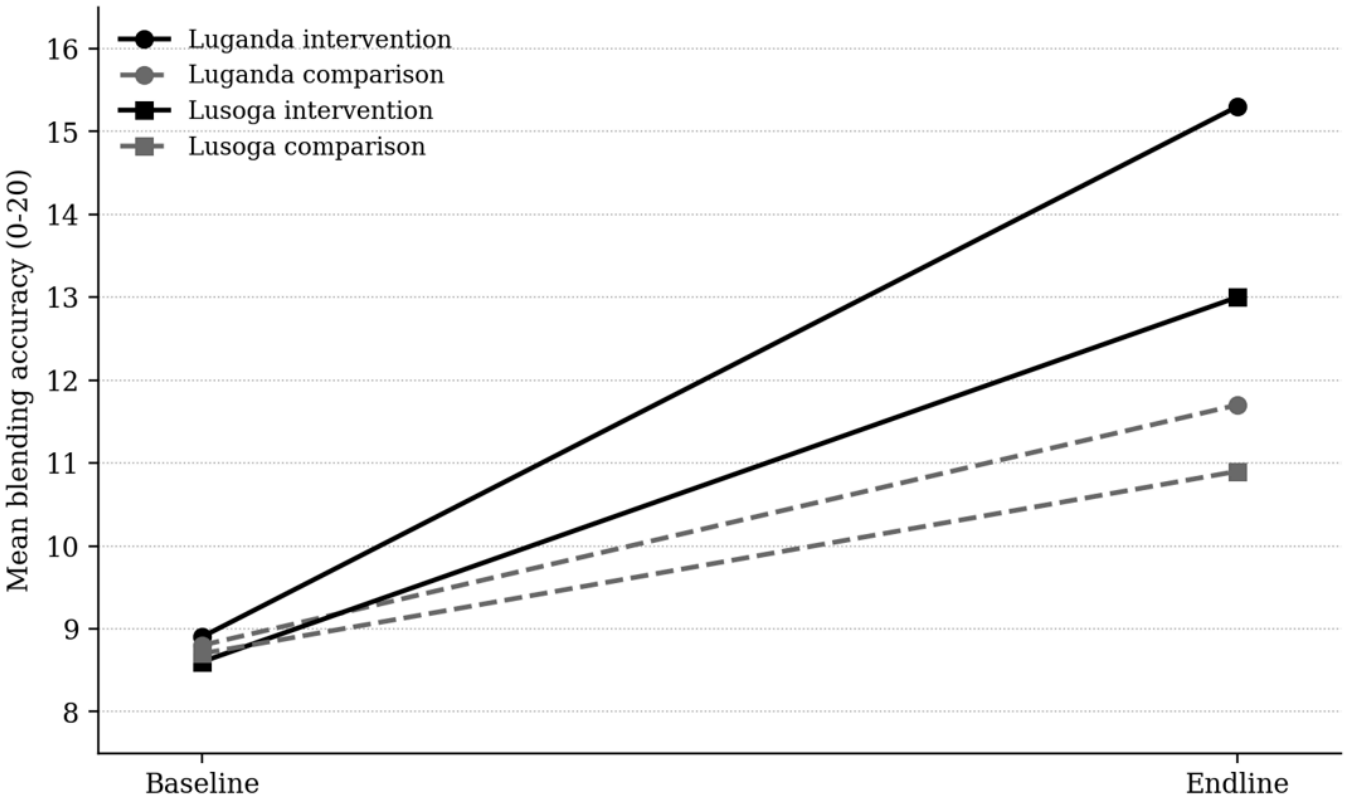

Figure 2

Mean blending accuracy at baseline and endline by study arm