Literator - Journal of Literary Criticism, Comparative Linguistics and Literary Studies
ISSN: (Online) 2219-8237, (Print) 0258-2279

Role and Reference Grammar-based grammar pedagogy for complex clauses and voice: A quasi-experiment in Indonesian English as a Foreign Language

Check for updates

Authors:
James E. Lalira[1]
Maya P. Warouw[2]
Jeane Mangangue[3]

Affiliations:
[1]English Language Education, Faculty of Teachers' Training and Education, Christian University of Indonesia in Tomohon, Tomohon, Indonesia

[2]English Department, Faculty of Humanities, Sam Ratulangi University, Manado, Indonesia

[3]English Department, Faculty of Foreign Language, Nusantara University, Manado, Indonesia

Corresponding author:
James Lalira,
jameslalira@fkipukit.ac.id

Dates:
Received: 26 Oct. 2025
Accepted: 21 Apr. 2026
Published: 19 May 2026

How to cite this article:
Lalira, J.E., Warouw, M.P. & Mangangue, J., 2026, 'Role and Reference Grammar-based grammar pedagogy for complex clauses and voice: A quasi-experiment in Indonesian English as a Foreign Language', *Literator* 47(1), a2235. https://doi.org/10.4102/lit.v47i1.2235

Copyright:
© 2026. The Authors. Licensee: AOSIS.
This work is licensed under the Creative Commons Attribution 4.0 International (CC BY 4.0) license (https://creativecommons.org/licenses/by/4.0/).

Read online:
Scan this QR code with your smart phone or mobile device to read online.

This quasi-experimental study examined whether an explicit Role and Reference Grammar (RRG)-based linking sequence contributed to improved accuracy in voice choice and complex-clause packaging among Indonesian EFL learners, as well as more efficient timed performance, and transfer to production academic discourse. Measures included a timed grammaticality judgement task and a sentence-picture matching task analysed for accuracy and reaction time, elicited imitation and elicited production tasks, an academic writing task, and oral retell. ANCOVA models were calculated with pre-test and placement covariates, using type-2 cluster-robust standard errors (CR2). Compared with the control group, the RRG group showed greater gains in voice accuracy, complex-clause accuracy, timed grammaticality judgement task (GJT) accuracy, and reaction times. Sentence-picture matching also demonstrated more accurate actor–undergoer mapping and more rapid responses. These gains were also reflected in writing quality, whereas speech intelligibility improved but did not reach conventional significance ($p = 0.08$). Moderate retention was observed at a delay of 3–4 weeks. Pedagogically, the findings suggest that a linking-first sequence may help learners more explicitly connect semantics, syntax and discourse when making grammatical choices in academic English.

Contribution: This study offers evidence from classroom-based research on the associations among replicable RRG-informed pedagogy and gains in voice, clause packaging, timed processing, writing transfer and short-delay retention in Indonesian EFL.

Keywords: EFL Indonesia; grammaticality judgement; mixed-effects; role & reference grammar; sentence-picture matching.

Introduction

One recurring concern in Indonesian English as a Foreign Language (EFL) scholarship is that grammar teaching often remains strongly form-focused, with limited explicit emphasis on the connections among form, meaning, and discourse roles (Lalira et al. 2024). However, in academic communication, learners must strategically choose voice (e.g. emphasising the undergoer via the passive) and package clauses cohesively using relative, complement, and adverbial constructions (Akesse 2018; Luo 2018). Because the passivisation, relativisation, coordination, and subordination of clauses, and the transfer from sentence-level exercises to writing and speaking, are often affected by persistent errors, the absence of a clear, teachable map from semantic roles to syntactic realisations frequently proves to be a significant stumbling block to student learning.

The Role and Reference Grammar (RRG) provides such a map by relating semantic representations (i.e. predicates and arguments) to syntax through the use of macro-roles (actor vs. undergoer), the Privileged Syntactic Argument (PSA), and a juncture–nexus system for organising clause relations at nuclear, core, and clausal levels (coordination, subordination, co-subordination, etc.). In pedagogical terms, RRG conceptualises grammar as a decision-making process: Learners identify predicate and argument structure, assign macro-roles, select a PSA in relation to information structure, and then realise those decisions through voice choice and clause linking. Therefore, instructors can create meaningful tasks that require students to connect form, meaning, and discourse, rather than simply applying isolated rules. From a typological perspective, Indonesian learners provide a productive yet complex context for examining voice and information-structure choices. Many learners may already have some familiarity with voice alternations and information-structure choices in L1 (e.g. dipassives), which may give them access

Note: Additional supporting information may be found in the online version of this article as Online Appendix 1.

to sensitivity to discourse prominence. On the other hand, English encodes similar functions by auxiliary morphology, past-participle formation, and optional 'by' phrases, which may obscure the communicative motivation for the use of the passive. Without a precise mapping between the two that explains why the undergoer should be foregrounded and how this choice is realised syntactically, learners may fall back on memorised rules they have learned by rote, which are not robust enough to take them to academic genres, such as lab reports, literature reviews, and cause-effect essays, in which clause packaging and information flow are as important as local accuracy.

This study examined whether RRG-based grammar pedagogy was associated with the improvement of the following: (1) voice selection (active vs. passive); and (2) complex-clause packaging (relative, complement, and adverbial clauses) in Indonesian EFL classes. Researchers further sought to determine whether RRG-based instruction was associated with greater processing efficiency and form-meaning integration as measured by timed grammaticality judgement accuracy and reaction time (GJT) and sentence-picture matching (SPM) of actor and/or undergoer mappings across voice conditions. In addition, transfer to production (elicited imitation and elicited production), transfer to academic writing and oral retell performance, and retention at the delayed post-test were investigated.

Despite the emergence of task-based and processing-oriented approaches in L2 grammar teaching, there is little classroom-based evidence for a full-fledged RRG-anchored pedagogy that simultaneously addresses the following: (1) voice as a discourse-functional choice; (2) clause packaging in terms of juncture–nexus relations; and (3) measures of processing alongside measures of production and writing quality, especially in Indonesian tertiary EFL settings in which many learners bring first-language resources for relatively flexible voice and topicalisation. Few studies have included these dimensions in a cohesive and replicable instructional sequence within the ordinary classroom time constraints. Filling this gap is essential, as it bears directly on the viability of a linking-first pedagogy with actor and/or undergoer assignment, PSA selection, and juncture–nexus selection in providing robust learning that generalises from comprehension to production and academic discourse.

Consistent with these objectives, this study addressed the following research questions (RQs):

- **RQ1:** Is RRG-based explicit instruction associated with greater gains than conventional instruction is in passive and/or active voice accuracy?
- **RQ2:** Is RRG-based explicit instruction associated with greater gains than conventional instruction is in complex-clause accuracy (relative, complement, and adverbial clauses)?
- **RQ3:** Is RRG-based explicit instruction associated with higher GJT accuracy and shorter reaction times?
- **RQ4:** Is RRG-based explicit instruction associated with more accurate SPM mapping of actor and/or undergoer across voice conditions?
- **RQ5:** Is RRG-based explicit instruction associated with gains in production (elicited imitation and/or production) and writing and/or speaking quality?
- **RQ6:** To what extent do the identified patterns appear comparable across programme and proficiency profiles, and what do the fidelity records suggest about instructional adherence and student engagement during delivery?

By focusing instruction on meaning-based association rather than on isolated rules, the proposed pedagogy addresses a long-standing applied question: How can learners choose communicatively appropriate forms? If supported, the RRG-based sequence may offer a useful framework for 'grammar-for-writing' courses in response to calls to combine explicit explanations with meaningful tasks and pushed output. Beyond the local relevance of this study, the paper speaks to broader discussions about how explicit teaching and communicative practice may be brought into closer alignment by providing a way to introduce a decision architecture that unifies semantics, syntax, and discourse. In the present study, the comparison does not treat RRG and communicative language teaching (CLT) as mutually exclusive approaches; instead, it compares an RRG-informed, explicit, meaning-sensitive sequence with a more conventional, rule-based instructional sequence in the context of ordinary Indonesian tertiary EFL classrooms.

Literature review

Role and reference grammar principles for pedagogic linking

Role and Reference Grammar describes syntax as a two-way relation between meaning and form. Its linking algorithm 'maps from semantics to syntax and from syntax to semantics' (Van Valin 2008, 2014). This means that grammar is taught by taking learners through a sequence of decisions from predicate-argument structure to sentence form (De Marneffe et al. 2021). In an RRG-based approach, students first identify the semantic predicate of a clause and its argument structure, after which macro-roles (actor vs. undergoer) are assigned to each of these arguments, and the pragmatic subject (PSA, often the grammatical subject) is designated for information focus. Next, they take advantage of the voice system; for instance, passivisation tends to bring the undergoer to the subject (PSA) position to foreground patients or themes and conform to the structure of discourse goals.

Role and Reference Grammar also deals with clause packaging through juncture–nexus constructions: nuclear, core, and clausal clauses may be linked together by coordination, subordination, or co-subordination (including relative clauses, complement clauses, and adverbial clauses) to indicate semantic relations (including causation, time, and manner; Mbuki 2019). Grammatical operators (tense, aspect, mood, and

negation) can be used to mark each clause to express temporal, modal or negative contexts. By conceptualising such elements as a 'teachable map', RRG turns abstract grammatical descriptions into classroom procedures that are carried out step by step. Teachers can present RRG analysis as a sequence of easy steps to find predicate and/or argument, assign macro-roles, select PSA, and apply voice, linking clauses and marking operators so that there is a correspondence between language form and language function (semantic and pragmatic meaning). Role and Reference Grammar combines meaning (semantics) and use (discourse) in grammatical decision-making. Importantly, this pedagogical use of RRG does not position explicit grammar teaching as contrary to communicative language teaching; rather, it offers an explicit route through which grammatical choices can be linked to communicative purpose.

Evidence from recent instructional research

Recent second language acquisition (SLA) literature has supported approaches integrating meaning-based practice and explicit form-meaning mapping processes. For instance, task-based instruction has been associated with positive L2 learning gains. A large-scale meta-analysis by Boers and Faez (2023) reports that task-based programmes can yield beneficial effects on L2 learning outcomes, suggesting that well-designed, structured communicative tasks can successfully scaffold grammar learning. Processing Instruction (PI) studies also suggest that when teachers guide learners' attention to form, they will use target structures faster and more automatically. For example, Henry (2025) reports that learners who received Processing Instruction had better performance than those who were taught with traditional methods on post-test measures of grammatical gender processing and assignment. Processing Instruction can therefore support accuracy and processing speed for challenging grammatical patterns while strengthening form-meaning integration.

Corpus-informed instruction has also been found to enhance learners' grammatical awareness. Jones and Oakey (2024) demonstrate that corpus-based teaching can develop the spoken grammar awareness of learners by making them more aware of repeated forms in context. According to Rodríguez-Fuentes and Swatek (2022), corpus-informed homework materials produced stronger learning outcomes than conventional materials did in a classroom-based study of grammatical construction learning. Together, these studies suggest that corpus-informed tasks can make form-function relations more salient and can support more context-sensitive grammatical choices. In short, corpus-informed tasks can cue target structures in context, raise learners' awareness of usage patterns and support the internalisation of constructions. Longitudinal research on writing demonstrates that syntactic sophistication and constructional control are reliable indicators of L2 development and are responsive to instruction. Crossley (2020) and Nelson (2024) show that more advanced sentence structuring is related to writing development and writing quality. In another study, usage-based complexity measures were better than traditional length-based measures at reflecting writing proficiency (Esfandiari & Ahmadi 2021, 2024; Li & Yang 2023). This finding suggests that as learners receive grammar-focused instruction, indices of syntactic complexity and sophisticated constructions may increase, particularly when writing assignments in which clauses need to be packaged rhetorically. Simultaneously, these lines of evidence are indirect regarding RRG because they promote explicit meaning-oriented teaching overall, but not yet a classroom sequence that is structured specifically in terms of macro-roles, PSA selection, and juncture–nexus choices.

Where role and reference grammar fit and what is missing

Two principles emerge from these approaches: (1) meaningful real-life communication practice must be provided to learners; and (2) explicit mapping between form and meaning must be made available. Role and Reference Grammar complements such traditions because it naturally incorporates semantics, syntax and discourse within the same framework.

By making an explicit connection among conceptual roles, syntactic structure, and discourse status in a single decision space, RRG-based pedagogy could, in principle, provide a coherent roadmap for learners to map meaning onto form (and vice versa) in production and comprehension. In this sense, RRG should not be considered incompatible with CLT; instead, RRG could be seen as one of the ways of making communicative choice more explicit and teachable.

However, there is little evidence on classroom RRG instruction. Few empirical studies have implemented full RRG-based sequences in the L2 classroom, and most data have been gathered from contexts (such as English L1) in which voice and topic are already quite flexible (Collins & Ruivivar 2021; Hall 2022a, 2022b). In many learner settings (including some Indonesian tertiary EFL contexts), learners' L1 may provide resources for passivisation and topicalisation, which still leaves the additive benefit of RRG-based exercises as an open question. Moreover, a few studies have examined measures of processing time (e.g. reaction time in grammaticality judgement or sentence-picture matching tasks) alongside production outcomes (accuracy and complexity in learner speech and/or writing; Román & Gómez-Gómez 2022). Consequently, it is unclear whether RRG-mediated instruction is associated with more efficient form-meaning mapping and enhanced language use or whether any observed gains transfer across outcome types.

This study fills these gaps by triangulating several outcome measures in an RRG-anchored teaching sequence in Indonesian EFL contexts. The research is based on timed comprehension exercises (GJT and reaction times), mapping accuracy tasks, and analysis of learners' written and spoken output. This holistic approach will help to determine how (and whether) RRG-based instruction is associated with timed performance, mapping precision, and the quality of written and spoken production in a real classroom setting.

Research methods and design

Sample and setting

Eighty ($N = 80$) Indonesian undergraduates were recruited from two higher-education institutions (a Christian university and a health science college) across eight programmes: English, Economics, Pharmacy, Biology, Theology, Nursing Diploma, Nursing Degree, and Health Administration. The sample was numerically balanced across programmes with 10 students from each programme (five in the RRG condition and five in the control condition) to avoid over-representation of any single programme; however, this balance should be interpreted as numerical distribution rather than as equivalence in prior English exposure or opportunity to learn English. All participants were regular enrollees in the target courses. Students who had spent extended time in English-speaking countries or whose identified language or learning conditions made timed language tasks inappropriate were not included.

The participants were Indonesian L1 speakers with regional languages as additional languages, and their proficiency was approximately in the Common European Framework of Reference for Languages (CEFR) B1–B2 range. Assignment to the RRG and control conditions was made at the level of the programme-based class sections to minimise across-group contamination (cluster-wise assignment). Each of the eight programmes contributed two sections, one assigned to the RRG condition and one to the control condition, yielding 16 sections in total (eight RRG and eight controls). Sections were paired on the time, scheduling, and availability of instructors. The study was conducted at the participating institutions after approval for data collection had been granted through the relevant institutional review process. Written informed consent was obtained from all participants. The data were anonymised prior to analysis and stored on encrypted drives.

Data collection

The research design was a pre-test/post-test/delayed/post-test quasi-experimental design: pre-test–8 sessions–post-test (within 48 h)–delayed post-test (approximately 3–4 weeks later). In this study, equal dosage referred to equivalent instructional time and coverage of target grammatical areas, rather than to identical pedagogical procedures. A 20–25-min mini-lecture introducing the key linking concepts (macro-roles, PSA, juncture–nexus relations, and relevant operators) and the task cycle (input → noticing → guided practice → pushed output) was delivered to the RRG group. The tasks involved explicit role-linking and clause-packaging choices (e.g. marking actor or undergoer, PSA choice, active or passive choice to fit information structure, and relative, complement or adverbial relation choice).

The control group received rule-based explanations and transformation exercises, but these lessons did not provide an explicit form–meaning–discourse linking sequence and did not systematically require actor–undergoer or PSA analysis. The GJT and SPM were administered on computers in silent classrooms. Each trial was conducted in a regular order: fixation (500 ms) → stimulus (sentence + response options) → response window (self-paced, within 6 s for the GJT; 8 s for the SPM) → intertrial interval (800 ms – 1000 ms). Accuracy and reaction time (RT, in ms) were recorded on a trial-by-trial basis. The blocks were randomised, and the order of the items was randomised within the blocks. Short breaks were given every 10 items. The instructors participated in a three-session training programme (RRG basics, lesson flow, and scoring or rubrics). A teaching kit (slides, handouts, SPM figures, EI and/or EP prompts, and rubrics) ensured the uniformity of delivery. Instructional adherence (covering the RRG targets) and student engagement (on-task time and participation) in class sections were recorded by using a fidelity checklist. Deviations and contextual notes (e.g. changing rooms or technical downtime) were recorded to facilitate transparency in procedures and an exploratory understanding of implementation consistency. Representative lesson outlines and sample instructional materials for both conditions were provided in Appendix 1 to support replicability.

Data analysis

Two raters scored EI or EP, writing, and retelling independently, and disagreements were resolved through discussion and, where necessary, consultation with a third rater. Inter-rater agreement for a stratified subset of annotated categories was estimated by using Cohen's κ. For timed tasks, reaction time (RT) trials below 200 ms or above 3 standard deviation (SD) from a participant's mean were excluded prior to analysis, and flagged response records were checked against the raw logs. Missing data (< 3%) were handled by using pair-wise deletion for descriptives and maximum-likelihood in mixed models. We summarised the CEFR-based proficiency indicators, placement scores, and vocabulary measures to assess baseline comparability between the groups. Because subgroup cell sizes were limited and section-level implementation data were descriptive, programme and/or proficiency variation and fidelity indicators were treated as contextual information, rather than as the foundation of formal predictive modelling. For descriptive inspection only, change scores (Δ = post − pre) were computed. For each outcome, we estimated ANCOVA models of the following form (see Equation 1):

$$\text{Post}_{ij} = \beta_0 + \beta_1 \text{Group}_i + \beta_2 \text{Pre}_{ij} + \beta_3 \text{Placement}_{ij} + \varepsilon_{ij}, \qquad \text{[Eqn 1]}$$

with CR2 cluster-robust standard errors (Huang & Li 2022; Lee & Pustejovsky 2024; Pfaffermayr 2023) to account for class-level clustering and the small number of clusters. We reported the adjusted mean differences, 95% confidence intervals (CIs), and standardised effect sizes. For the GJT, we fit mixed-effects models with fixed effects Group × Time × Structure and random intercepts for participants (with by-participant random slopes for structure added where

convergence permitted). Retention was assessed by using delayed post-test differences with the same covariate structure. All analyses were scripted in R/Python (code and outputs are provided in Appendix 1). De-identified summary outputs and APA-style table templates were also provided to facilitate independent verification. In the ANCOVA specification, *X* denotes treatment group (RRG vs. Control), and *C* denotes the covariate (pre-test score; Brown et al. 2023; Cook et al. 2020; Liu & Maxwell 2020). ANCOVA assumptions include independence of observations, normality of residuals, homogeneity of error variance, linearity between the covariate and outcome within each group, and homogeneity of regression slopes (Knief & Forstmeier 2021). We checked these assumptions using diagnostic plots and tests. To address the issue of clustering, we used cluster-robust (sandwich) variance estimates, with Bell McCaffrey degrees-of-freedom adjustment (CR2) to improve small-cluster inference. This approach is appropriate even for a low number of clusters and provides more accurate *p*-values than standard or HC3 adjustments do.

As a robustness check, we used linear mixed-effects models with random intercepts for class to address the issue of nested data and found similar effect estimates. Effect sizes were calculated as Hedges' g for adjusted mean differences, which were bias-corrected for small sample sizes (Brydges 2019; Lin & Aloe 2021). Where timed contrasts were described descriptively, Cohen's d is reported to make them easy to interpret, but the main index of effect size to use in adjusted between-group comparisons is Hedges g. For interpretation, values around 0.2 were considered small, around 0.5 moderate, and 0.8 or above large. As the research was based on a clustered quasi-experimental design, the results were viewed with the necessary caution regarding selection effects, instructor effects, and unmeasured between-class differences (Gopalan, Rosinger & Ahn 2020; Handley et al. 2018; Waddington et al. 2017).

Results

Baseline comparability

Descriptive comparisons indicated that the RRG and control groups were comparable on observed baseline measures, including placement, vocabulary, and CEFR distribution (Table 1). Regarding RQ6, the available descriptive evidence did not indicate an apparent imbalance in baseline characteristics between the two conditions in terms of programme and proficiency profiles; however, the study was not conducted with sufficient power to sub-analyse the subgroups. No ceiling effects were observed, and the baseline distributions showed adequate variance for detecting change.

Internal consistency of the measures

The internal consistency of multi-item measures was acceptable to high. Cronbach's α and McDonald's ω were computed for each subscale (Table 2). All α and ω values exceeded 0.70, indicating adequate internal consistency across subscales (Trabelsi et al. 2024).

Inter-rater reliability

Two annotators coded a subset of sentences for juncture–nexus relations and information structure (topic and/or focus). Cohen's κ indicated substantial agreement for these categories (Table 3). According to Demers et al. (2021) guidelines, κ values in the 0.61–0.80 range reflect substantial inter-rater agreement.

Primary outcomes

Adjusted ANCOVA results for the core primary outcomes are presented in Table 4. Results from ANCOVA (CR2 SEs) revealed significant group differences favouring the RRG condition in voice accuracy and complex-clause accuracy. For example, voice accuracy favoured the RRG condition with an adjusted difference of +11.05 (standard error [SE] = 3.12), 95% CI (4.95, 17.15), $p < 0.001$, $g = 1.057$, indicating a large effect. Complex-clause accuracy showed an adjusted difference of +9.47 (SE = 2.89), 95% CI (3.77, 15.17), $p = 0.002$, $g = 0.85$. Writing quality improved by +8.95 points in the RRG condition, a significant gain over the control (95% CI [4.57, 13.33], $p < 0.001$, $g = 0.82$). The speaking intelligibility gains (+4.06) did not reach conventional significance ($p = 0.08$, $g = 0.41$). The RRG condition also showed higher accuracy and shorter response times on the timed GJT and SPM tasks. Timed results are provided in more detail in 'Secondary outcomes (elicited imitation or elicited production and sentence-picture matching)' section to prevent duplication. Taken together, these patterns were consistent with RQ1–RQ5. Retention effects at the delayed post-test were generally in the same

TABLE 1: Baseline comparability.

Measure	RRG			Control		
	M	SD	*n* out of 20	M	SD	*n* out of 20
Placement (0–100)	61.5	8.4	-	61.1	8.6	-
Vocabulary (0–60)	42.3	6.1	-	41.9	6.5	-
CEFR B1/B2	-	-	20	-	-	20

RRG, role and reference grammar; M, mean; SD, standard deviation; CEFR, common European framework of reference for languages.

TABLE 2: Internal consistency of the measures (α/ω per subscale).

Subscale	Cronbach's α	McDonald's ω
Voice accuracy (active or passive)	0.85	0.87
Clause packaging (relative)	0.79	0.81
Clause packaging (complement)	0.82	0.85
Clause packaging (adverbial)	0.78	0.80
Writing quality (combined)	0.88	0.90
Speaking intelligibility	0.83	0.85

TABLE 3: Inter-rater reliability: Cohen's κ for juncture–nexus and information structure.

Annotation category	Cohen's κ
Juncture–nexus (coordination)	0.78
Information structure (topic or focus)	0.69

TABLE 4: ANCOVA results for the core primary outcomes (adjusted mean differences, standard error, 95% confidence interval, *p*-value, and *Hedges'* g).

Outcome	Adjusted mean difference (RRG-Control)	SE	95% CI	*p*-value	Hedges' *g*
Voice accuracy	+11.05	3.12	4.95, 17.15	< 0.001	1.05
Complex-clause accuracy	+9.47	2.89	3.77, 15.17	0.002	0.85
Writing quality (score)	+8.95	2.19	4.63, 13.27	< 0.001	0.82
Speaking intelligibility	+4.06	2.29	-0.51, 8.63	0.08	0.41

RRG, role and reference grammar; SE, standard error.

direction but with smaller magnitudes; a summary is reported in the main text, with fuller structural breakdowns provided in Appendix 1.

Secondary outcomes (elicited imitation or elicited production and sentence-picture matching)

In elicited imitation and elicited production (EI and/or EP), the RRG group produced the target structures more accurately than the control group under task constraints did. This pattern was also reflected in writing-related transfer indicators, in which the RRG group showed stronger control of passive voice and complex-clause use. In sentence-picture matching (SPM), the RRG group also showed a modest but consistent advantage in actor–undergoer mapping accuracy alongside faster responses, and this phenomenon was consistent with the timed results in the summary in 'Primary outcomes' section. The same directional pattern was observed at the delayed post-test for these secondary outcomes, although the values were lower than at the immediate post-test. Because these measures supplement the main ANCOVA outputs, only the major descriptive trends are condensed here; expanded values at the structure level are shown in Appendix 1 (Tables A1–1 – A1–5).

Practical interpretation of scale differences

The scores reported above range from 0 to 100 on percentage or point scales, in which higher values indicate greater accuracy or better performance. For example, voice and clause accuracy scores are reported as percentages, so an adjusted difference of +11 points in voice accuracy corresponds to an 11-point advantage on the 0–100 scale. Writing quality was rated on a rubric with the same overall scale, and an adjusted difference of +8.95 points likewise indicates a meaningful improvement in academic writing performance. By conventional benchmarks, effect sizes around 0.2 are typically interpreted as small, around 0.5 as moderate, and 0.8 or above as large.

On this basis, the observed effects for voice accuracy and writing quality fall within a large range. In practical terms, differences in the range of approximately 9–11 points on a 0–100 scale suggest learning gains that are substantial enough to be educationally noticeable, rather than trivial fluctuations attributable only to routine practice or measurement noise. However, these interpretations should be read in light of the clustered quasi-experimental design and denote practically meaningful differences associated with the RRG condition, rather than definitive causal effects.

Discussion

This section interprets the main findings of a quasi-experiment on role and reference grammar-based grammar pedagogy in the Indonesian EFL context. The findings are linked to six research questions (RQ1–RQ6), with RQ1–RQ5 addressed through the main outcome analyses, and RQ6 treated more cautiously as an exploratory implementation question. The results are synthesised with recent work on meaning-oriented grammar instruction and used to clarify the study's novelty, as well as its theoretical, practical, and methodological implications.

With a pre-test/post-test/delayed/post-test design and triangulated measures including a timed grammaticality judgement task with reaction times, a sentence-picture matching task, elicited imitation and elicited production, academic writing, and oral retell, and covariate controls, the study provides a useful basis for evaluating whether a linking-first pedagogy can yield gains in accuracy, processing efficiency, and transfer. The discussion below interprets these findings in relation to the six research questions and the wider literature on meaning-oriented grammar instruction.

Voice accuracy

ANCOVA with CR2 adjustments indicated that the RRG condition outperformed the control condition with an adjusted difference of approximately 11 points, and an effect size of approximately $g = 1.05$, which is large and pedagogically meaningful. Interpretively, this pattern is consistent with the hypothesis that the choice of Privileged Syntactic Argument and/or subject in the framework of RRG is based on the macrorole assignment to the actor or undergoer, which may help learners align voice choice with discourse function, for example, foregrounding the undergoer using the passive for the purpose of realising thematic focus. In practical terms, learners appeared to move beyond the blind application of transformational rules. They instead appeared to follow a decision workflow, identifying predicate and argument, assigning macro-roles, selecting the PSA, and realising voice choices and related operators.

The larger improvements in voice accuracy than in some other components suggest that when the mapping from meaning to form is organised as a clear sequence of choices, learners may move more efficiently from rule memorisation to more meaningful form selection. From a typological perspective, learners may benefit from functional parallels available in Indonesian, such as dipassives. Even so, the RRG framework may provide a formal bridge between L1

discourse motivations and English morphology and syntax, including auxiliary + past-participle construction, and an optional by-phrase, so that the passive becomes a communicative choice, rather than a difficult form to avoid.

Accuracy in complex clauses, relative or complement, and adverbial clauses

The improvement in clause-packaging accuracy, with an adjusted difference of approximately 9.5 points and *g* of around 0.85, suggests that juncture–nexus relations, coordination, and subordination at the nuclear, core, and clausal levels can be taught explicitly within the same decision framework. Although the effect was slightly smaller than that for voice accuracy, the effect size was large enough to be substantively important.

From a cognitive perspective, clause packaging likely places a greater demand on learners who must decide which semantic relation to express (causality, time, condition, etc.) and must mark the relation and combine the structural information. For that reason, progress in this domain may be somewhat slower even when instruction is effective. The RRG curriculum, which guides learners from the need to express a particular semantic relation to its grammatical realisation as a relative, complement, or adverbial clause, appears to support movement from sentence-level manipulation towards paragraph-level academic discourse.

This finding aligns with research based on corpus and task, focusing on exposure to frequent form-function pairings and principled construction choices in authentic situations. The difference here is that the RRG organises those choices within a common linking structure, so that the relations of clauses are not taught as a disjoint list but as part of an integrated decision architecture.

Timed GJT performance: Accuracy up and RT down

The timed GJT accuracy increased by approximately 11 points (d = 0.98). It was accompanied by a reduction in reaction time of approximately 95 ms (d = 0.62), a pattern consistent with more efficient task performance and the early routinisation of the linking sequence. From a processing-oriented perspective, instruction that directs attention to meaning-bearing cues – who is doing what, who is affected, and how inter-clausal relations are mapped onto grammatical form – may support faster and more accurate decision-making. Under repeated practice, when learners maintain the linking steps rather than relying only on surface-rule checking, they may experience less difficulty in semantic-syntactic integration and may therefore make grammatical choices with lower processing demands. The RT effects are understandably smaller than the accuracy effects because speeded performance may lag gains in explicit accuracy. Nevertheless, the regularity in both metrics supports the argument that RRG pedagogy was associated not only with higher scores but also with more efficient sentence processing under timed conditions.

Sentence-picture matching: Actor–undergoer mapping and response efficiency

SPM accuracy improved by approximately 5.5 points, with *g* around 0.71, and reaction times were reduced. This pattern is consistent with stronger mapping of semantic roles onto event representations, especially for passives. The SPM task may be interpreted as a useful stress test for the comprehension of propositions. When macro-roles and PSA are understood, learners can choose the correct image without lexical translation of the sentence. This pattern is important because it may reflect stronger conceptual representations rather than only task-specific strategies. Sensitivity to prominence alignment, including the undergoer functioning as the PSA in passives, seems to have grown, which may help explain later production choices in both writing and speaking. In this sense, RRG may be understood as offering an explicit link between form and function across receptive and productive modes, including reading, listening, speaking, and writing.

Transition to production, elicited imitation and elicited production, and the quality of writing and speech

For elicited imitation, elicited production, and academic writing, the RRG group showed stronger control of passives and complex clauses, with better coherence and information linkage than the control group did. This finding is consistent with research suggesting that construction-sensitive measures of writing may be more responsive to pedagogical intervention than crude length-based indices are. When learners are required to make discourse-level decisions about what to topicalise or focus on and how to realise inter-clausal relations, paragraph-level rhetoric control may improve (Bodger 2025). In contrast, speaking intelligibility showed a positive numerical trend but did not reach conventional significance (p = 0.08). There are at least three plausible methodological explanations for this observation. Firstly, oral retell tasks tend to suppress passive and relative-clause use, because fluency pressure often leads learners to default to simpler constructions in real time, especially under time pressure.

Secondly, the dosage of eight sessions may be sufficient for planned writing but insufficient for automatising complex constructions in spontaneous speech. Thirdly, the intelligibility rubric captures global intelligibility, including segmental and prosodic factors, and message completeness, which are not governed only by passive choice or clause relations. Taken together, these points suggest that the oral domain may require longer or more targeted practice before gains of the same magnitude become visible. Further designs that extend oral output, such as guided retells with relative or complement clauses, may help extend these effects into spoken performance.

Retention at the delayed post-test

Moderate but meaningful retention suggests that gains were still evident 3–4 weeks after instruction, although some decay was observed relative to the immediate post-test. This pattern is consistent with the consolidation of recently learned skill

sequences. A newly learned linking algorithm may stabilise more effectively when it is periodically retrieved and applied in meaningful tasks. Curricular sequencing should therefore provide spaced retrieval and cumulative tasks that require voice and juncture decisions across genres, such as laboratory reports, literature reviews, and cause-and-effect essays, paired with brief self-explanations of form-function decisions. In this way, the representational trace built during the initial intervention may be maintained and gradually strengthened through continued use.

Possible explanatory mechanisms and broader contribution

Three explanatory patterns appear especially relevant. Firstly, RRG offers an explicit decision architecture that requires learners to settle meaning, macro-roles, PSA, and inter-propositional relations before choosing form, interrupting the habit of answering form with form and replacing the habit with selecting form from meaning and discourse needs. Secondly, the same linking sequence can operate across receptive and productive modes. Because the linking path is the same in reception and production, work in one mode, such as SPM or timed GJT, may prepare some of the conceptual work needed in other modes, including writing and speaking. Thirdly, the pedagogy packages discourse decisions explicitly. Positioning voice and complexification in an information-structuring logic may help learners to foreground the undergoer when appropriate, control information flow, and construct coherent paragraphs. Taken together, these patterns may help explain why the largest effects arose for voice, which is a relatively local PSA-related decision with consequences for information structure, followed by clause packaging, which is cognitively heavier. In contrast, oral outcomes may require more dosage and scaffolding to elicit target constructions without compromising fluency.

The results add to growing evidence that grammatical structures taught through meaningful tasks that emphasise the relationship between form and meaning can produce durable learning benefits beyond those typically associated with transformational drills. The timed GJT pattern is broadly consistent with findings from processing instruction when the focus of attention on meaningful information facilitates the early processing of grammatical decisions.

Gains in clause packaging and writing quality are consistent with longitudinal results showing that constructional sophistication is instruction-sensitive and correlated with proficiency development (Crossley 2020; Deng, Uccelli & Snow 2022). The advantage of SPM suggests that the enhancement of the actor–undergoer representations is unlikely to be explained only by familiarity with the tests used but may instead reflect the internalisation of form-function relations. The contribution of the present study lies in bringing these strands together within a single explicit and replicable RRG-based linking framework, thereby extending task-based and processing-oriented traditions into a more explicit classroom decision architecture.

Compared with previous research in classrooms, this study has at least four contributions. Firstly, the study triangulates outcome types that are seldom integrated into one design by testing voice, clause packaging, timed GJT with RT, SPM, EI, EP, writing, and retention in a realistic eight-session classroom protocol in Indonesian EFL. Secondly, it proposes a replicable pedagogy based on standardised linking sequences and teaching kits and rubrics that can be adapted by instructors without requiring advanced specialist training. Thirdly, it focuses on discourse function, teaching of voice choice, and clause packaging as decisions about information, rather than as entries in rule catalogues, and shows transfer to academic writing. Fourthly, it applies CR2-adjusted cluster-robust analysis within a quasi-experimental comparison across study programmes, which provides a stronger basis for inference than an unmanaged group comparison does while still requiring caution about causal interpretation.

Together, these contributions are especially relevant in the Indonesian EFL context and suggest that RRG can function not only as a typological description but also as a promising pedagogical framework when translated into a classroom decision workflow. These results are consistent with the view that PSA selection is a discourse-driven choice. Selecting the PSA, often the same as the grammatical subject, depends on information-structuring needs; therefore, voice choice is not cosmetic but a result of information packaging. From this perspective, passive errors may not be purely morphological; they may also reflect difficulty in aligning focus requirements with syntactic structures. Likewise, junctures and nexuses provide a formal language for expressing propositional relations that provide a basis for cohesion and coherence. The integration of semantics, syntax, and discourse provides a mediating stance in the debates over explicit versus communicative instruction, which have been running for decades (Li & Zhang 2022; Mirzaei et al. 2021).

Targeted explicitness, which is explicit about how meaning is converted into form, is compatible with and can even enhance meaningful tasks that require using those decisions. Thus, the RRG may be viewed as an operational middle ground between decontextualised rule teaching and communicative practice. The grammar-for-writing curriculum should incorporate the concept of voice and clause complexity into the role of information packaging across various academic genres, such as summaries, problem-solution texts, and cause-effect analyses. A stable decision sequence can then be taught, namely: (1) identifying the predicate (PRED) and argument (ARG); (2) assigning macro-roles; (3) selecting the PSA; (4) realising voice; (5) choosing juncture and nexus relations; (6) marking operators; and (7) finishing with a brief discourse review. Classwork may progress from guided noticing with short SPM or timed GJT probes to guided practice with the help of decision cards with focus and semantic relations to be expressed, and then to pushed output through paraphrase or rewrite using relative, complement, and adverbial

constructions. Assessment rubrics should focus on the quality of decisions and not merely surface form (e.g. the suitability of voice to informational focus or the suitability of relations between clauses to rhetorical purposes) and direct feedback towards reasoned choices.

Thus, recycling should be combined with regular (weekly) voice audits and microtasks that work on rotating clause relationships across topics. Short RRG-oriented professional-development modules for instructors, together with think-aloud modelling of the decision path while composing, may help establish this approach as a sustainable classroom culture. Based on the present findings, repeated practice with similar decision sequences across contexts and genres appears to be a plausible route for consolidating these gains.

Suggestions for future research

This study had some limitations. With respect to RQ6, the available evidence on programme and/or proficiency variation and implementation fidelity was descriptive rather than inferential, so no strong subgroup or predictive claims are made here. The clustered quasi-experimental design did not permit individual randomisation, although CR2 adjustments and baseline comparability in placement, CEFR level, and vocabulary knowledge helped reduce concerns about major observable bias. The eight-session dosage was adequate to produce large effects on voice and writing but may not be optimal for spontaneous speech; thus, future work should include more staged oral output tasks. The instruments used here estimated specific processing facets, and future studies could introduce self-paced reading, eye-tracking, or dual-task paradigms to help probe deeper automatisation. Medium-term retention was measured at 3–4 weeks and did not address semester-long durability, which may require booster sessions at regular intervals. Future research should verify potential heterogeneity of effects in larger, multi-campus samples across fields and proficiency levels, preferably using designs with adequate power to conduct subgroup, interaction, and implementation-fidelity analyses. Finally, the matching between oral assessment and task design is of interest. Because passives and relative clauses are often licensed by discourse conditions, oral tasks such as information-gap retells may increase the sensitivity of intelligibility measures to form-function choices (De La Torre García, Ainciburu & Buyse 2021; Jiang & Hyland 2023).

Programme development can occur in several ways. A spiral model can recycle voice and clause relations across different text types, laboratory reports, literature reviews, and discussions of results with progression from guided to autonomous choices. Short decision cards, flow diagrams, and decision-support materials for RRG can be used for warm-ups with mini-tasks lasting between 5 and 7 min. Formative assessment can be built into very short GJT or SPM probes used as daily checks, with visualisation of RT and accuracy trends so that learners can track their processing gains over time. Finally, an open repository of teaching kits, rubrics, and annotation sheets could potentially increase the consistency of implementation across campuses in Indonesia. These suggestions are motivated by the observed pattern that performance improved across modalities when linking decisions were explicitly practised and assessed.

Taken together, the results suggest that explicit, meaning-oriented RRG-based grammar instruction was associated with large gains in voice selection, clause packaging and writing quality alongside converging improvements in timed GJT performance and more stable semantic mapping in SPM. Transfer to speech was more limited and appears to require stronger task design and/or greater dosage. Theoretically, the findings are consistent with the view that PSA functions as an information-structuring choice and that juncture and nexus are useful tools of discourse coherence. Practically, the RRG appears to provide a teachable, assessable, and replicable decision architecture that may help bridge explicit instruction and communicative practice. In the Indonesian tertiary EFL context, in which academic literacy is an important instructional goal, the present findings suggest that grammar should be taught as a planned mapping from meaning to form, rather than through a purely rules-based sequence.

Assessment should focus not only on surface correctness but also on the quality of grammar decisions, and recycled practice in meaningful tasks across genres should be built into instruction. With sustained recycling and appropriate task design, short-term retention may develop a foundation for longer-term academic development, although that longer trajectory still requires empirical confirmation.

Conclusion

This study indicated that the explicit use of RRG-based pedagogy was associated with stronger performance in voice choice and clause packaging, in ways that were sensitive to communicative intent. By making semantics-syntax relations more explicit, the RRG-informed sequence was associated with higher accuracy and more efficient timed performance than the comparison condition was. Transfer was clearer in writing than in speaking, and moderate retention was observed over an interval of 3 to 4 weeks. These findings suggest that explicit grammar instruction and communicative practice need not be treated as opposing approaches, because an RRG-informed sequence can provide a coherent decision architecture for mapping meaning to form. Given the strong effect sizes observed for several key outcomes, RRG-based approaches merit further attention in curriculum design, although replication across contexts remains necessary. Future research should investigate the long-term effects and the way this pedagogy can be integrated with other forms of classroom interaction across different L2 contexts.

Acknowledgement

The author would like to express sincere gratitude to the students and lecturers of Universitas Kristen Indonesia Tomohon and STIKES Bethesda Tomohon for their participation and support during the implementation of this pedagogical quasi-experiment.

Competing interests

The authors declare that they have no financial or personal relationships that may have inappropriately influenced them in writing this article.

CRediT authorship contribution

James E. Lalira: Conceptualisation; Formal analysis; Investigation; Methodology; Project administration; Writing – original draft; Writing – review & editing. Maya P. Warouw: Conceptualisation; Formal analysis; Investigation; Methodology; Project administration; Writing – original draft; Writing – review & editing. Jeane Mangangue: Conceptualisation; Resources; Supervision; Writing – review & editing. All authors reviewed the article, contributed to the discussion of results, approved the final version for submission and publication, and take responsibility for the integrity of its findings.

Ethical considerations

This study involved human participants and was conducted after approval had been obtained from the Institutional Ethics Committee of the Christian University of Indonesia in Tomohon. The ethical clearance number is 102/LPPM-UKIT/X/2025. Written informed consent was obtained from all participants, and the data were anonymised prior to analysis.

Funding information

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

Data availability

The data that support the findings of this study are available on request from the corresponding author, James E. Lalira.

Disclaimer

The views and opinions expressed in this article are those of the authors and are the product of professional research. It does not necessarily reflect the official policy or position of any affiliated institution, funder, agency, or that of the publisher. The authors are responsible for this article's findings and content.

References

Akesse, P., 2018, *The forms and functions of passive constructions in Ghanaian newspaper editorials*, University of Cape Coast, Cape Coast.

Akesse, P., 2018, *The forms and functions of passive constructions in Ghanaian newspaper editorials*, University of Cape Coast.

Bodger, F., 2025, 'Gateways to a different world of meaning: Expanding theme use in primary-aged children's writing', *Written Communication* 42(4), 860–893. https://doi.org/10.1177/07410883251346403

Boers, F. & Faez, F., 2023, 'Meta-analysis to estimate the relative effectiveness of TBLT programs: Are we there yet?', *Language Teaching Research* 30(3), 1525–1543. https://doi.org/10.1177/13621688231167573

Brown, S., Song, M., Cook, T.D. & Garet, M.S., 2023, 'Combining a local comparison group, a pretest measure, and rich covariates: How well do they collectively reduce bias in nonequivalent comparison group designs?', *American Educational Research Journal* 60(1), 141–182. https://doi.org/10.3102/00028312221136565

Brydges, C.R., 2019, 'Effect size guidelines, sample size calculations, and statistical power in gerontology', *Innovation in Aging* 3(4), igz036. https://doi.org/10.1093/geroni/igz036

Collins, L. & Ruivivar, J., 2021, 'Research agenda: Researching grammar teaching and learning in the second language classroom', *Language Teaching* 54(3), 407–423. https://doi.org/10.1017/S0261444821000070

Cook, T.D., Zhu, N., Klein, A., Starkey, P. & Thomas, J., 2020, 'How much bias results if a quasi-experimental design combines local comparison groups, a pretest outcome measure and other covariates?: A within study comparison of preschool effects', *Psychological Methods* 25(6), 726–746. https://doi.org/10.1037/met0000260

Crossley, S.A., 2020, 'Linguistic features in writing quality and development: An overview', *Journal of Writing Research* 11(3), 415–443. https://doi.org/10.17239/jowr-2020.11.03.01

De La Torre García, N., Ainciburu, M.C. & Buyse, K., 2021, 'Morphological complexity and rated writing proficiency: The case of verbal inflectional diversity in L2 Spanish', *ITL – International Journal of Applied Linguistics* 172(2), 290–318. https://doi.org/10.1075/itl.20009.del

De Marneffe, M.-C., Manning, C., Nivre, J. & Zeman, D., 2021, 'Universal dependencies', *Computational Linguistics* 47(2), 255–308.

Demers, K., Morin, C., Collette, L., DeMont, R., 2021, 'Moderate to substantial inter-rater reliability in the assessment of cranial bone mobility restrictions', *Journal of Alternative and Complementary Medicine* 27(3), 263–272. https://doi.org/10.1089/acm.2020.0325

Deng, Z., Uccelli, P. & Snow, C., 2022, 'Diversity of Advanced Sentence Structures (DASS) in writing predicts argumentative writing quality and receptive academic language skills of fifth-to-eighth grade students', *Assessing Writing* 53, 100649. https://doi.org/10.1016/j.asw.2022.100649

Esfandiari, R. & Ahmadi, M., 2021, 'Syntactic complexity measures and academic writing proficiency: A corpus-based study of professional and students' prose', *Journal of AsiaTEFL* 18(3), 745–763. https://doi.org/10.18823/asiatefl.2021.18.3.1.745

Esfandiari, R. & Ahmadi, M., 2024, 'Large and fine-grained complexity measures and writing quality among Iranian EFL learners', *Journal of Language and Translation* 14(3), 273–287.

Gopalan, M., Rosinger, K. & Ahn, J.B., 2020, 'Use of quasi-experimental research designs in education research: Growth, promise, and challenges', *Review of Research in Education* 44(1), 218–243. https://doi.org/10.3102/0091732X20903302

Hall, J.K., 2022a, 'L2 classroom input and L2 positionally sensitive grammars: The role of information-seeking question sequences', *Modern Language Journal* 106(S1), 113–131. https://doi.org/10.1111/modl.12751

Hall, J.K., 2022b, 'L2 classroom interaction and its links to L2 learners' developing L2 linguistic repertoires: A research agenda', *Language Teaching* 55(1), 100–115. https://doi.org/10.1017/S0261444820000397

Handley, M.A., Lyles, C.R., McCulloch, C. & Cattamanchi, A., 2018, 'Selecting and improving quasi-experimental designs in effectiveness and implementation research', *Annual Review of Public Health* 39, 5–25. https://doi.org/10.1146/annurev-publhealth-040617-014128

Henry, N., 2025, 'The effects of processing instruction on the acquisition and processing of grammatical gender in German', *Language Teaching Research* 29(4), 1426–1457. https://doi.org/10.1177/13621688221096368

Huang, F.L. & Li, X., 2022, 'Using cluster-robust standard errors when analyzing group-randomized trials with few clusters', *Behavior Research Methods* 54(3), 1181–1199. https://doi.org/10.3758/s13428-021-01627-0

Jiang, F. & Hyland, K., 2023, 'Changes in research abstracts: Past tense, third person, passive, and negatives', *Written Communication* 40(1), 210–237. https://doi.org/10.1177/07410883221128876

Jones, C. & Oakey, D., 2024, 'Learners' perceived development of spoken grammar awareness after corpus-informed instruction: An exploration of learner diaries', *Tesol Quarterly* 58(3), 1138–1165. https://doi.org/10.1002/tesq.3305

Knief, U. & Forstmeier, W., 2021, 'Violating the normality assumption may be the lesser of two evils', *Behavior Research Methods* 53(6), 2576–2590. https://doi.org/10.3758/s13428-021-01587-5

Lalira, J.E., Pangemanan, Y.A.T., Scipio, J.E., Lumi, S., Merentek, T.C. & Tumuju, V.N., 2024, 'Evaluating the impact of AI tools on grammar mastery: A comparative study of learning outcomes', *Voices of English Language Education Society* 8(3), 701–713. https://doi.org/10.29408/veles.v8i3.27856

Lee, Y.R. & Pustejovsky, J.E., 2024, 'Comparing random effects models, ordinary least squares, or fixed effects with cluster robust standard errors for cross-classified data', *Psychological Methods* 29(6), 1084–1099. https://doi.org/10.1037/met0000538

Li, D. & Zhang, L., 2022, 'Exploring teacher scaffolding in a CLIL-framed EFL intensive reading class: A classroom discourse analysis approach', *Language Teaching Research* 26(3), 333–360. https://doi.org/10.1177/1362168820903340

Li, Y. & Yang, R., 2023, 'Assessing the writing quality of English research articles based on absolute and relative measures of syntactic complexity', *Assessing Writing* 55, 100692. https://doi.org/10.1016/j.asw.2022.100692

Lin, L. & Aloe, A.M., 2021, 'Evaluation of various estimators for standardized mean difference in meta-analysis', *Statistics in Medicine* 40(2), 403–426. https://doi.org/10.1002/sim.8781

Liu, Q. & Maxwell, S.E., 2020, 'Multiplicative treatment effects in randomized pretest-posttest experimental designs', *Psychological Methods* 25(1), 71–87. https://doi.org/10.1037/met0000222

Luo, H., 2018, *Passive voice usage in undergraduate STEM textbooks*, College of Education and Human Performance, University of Central Florida, Orlando, FL, viewed 3 October 2025, from http://purl.fcla.edu/fcla/etd/CFE0007047.

Mbuki, K., 2019, *Clause chaining in Kisukuma: A role and reference grammar approach*, University of Nairobi, Nairobi.

Mirzaei, A., Naseri, F., Jafarpour, A. & Eslami, Z., 2021, 'ZPD-based mediation of L2 learners' comprehension of implicatures: An educational praxis framework', *Lodz Papers in Pragmatics* 17(1–2), 127–152. https://doi.org/10.1515/lpp-2021-0007

Nelson, Jr., R., 2024, 'Using constructions to measure developmental language complexity', *Cognitive Linguistics* 35(4), 481–511. https://doi.org/10.1515/cog-2023-0062

Pfaffermayr, M., 2023, 'Cross-sectional gravity models, PPML estimation, and the bias correction of the two-way cluster-robust standard errors', *Oxford Bulletin of Economics and Statistics* 85(5), 1111–1134. https://doi.org/10.1111/obes.12553

Rodríguez-Fuentes, R.A. & Swatek, A.M., 2022, 'Exploring the effect of corpus-informed and conventional homework materials on fostering EFL students' grammatical construction learning', *System* 104, 102676. https://doi.org/10.1016/j.system.2021.102676

Román, P. & Gómez-Gómez, I., 2022, 'Changes in native sentence processing related to bilingualism: A systematic review and meta-analysis', *Frontiers in Psychology* 13, 757023. https://doi.org/10.3389/fpsyg.2022.757023

Trabelsi, K., Saif, Z., Driller, M.W., Vitiello, M.V. & Jahrami, H., 2024, 'Evaluating the reliability of the athlete sleep behavior questionnaire (ASBQ): A meta-analysis of Cronbach's alpha and intraclass correlation coefficient', *BMC Sports Science, Medicine and Rehabilitation* 16(1), 1. https://doi.org/10.1186/s13102-023-00787-0

Van Valin, Jr., R.D., 2008, *Investigations of the syntax–semantics–pragmatics interface*, John Benjamins Publishing Company, Amsterdam.

Van Valin, Jr., R.D., 2014, 'Role and Reference Grammar', in A. Carnie, Y. Sato & D. Siddiqi (eds.), *The Routledge handbook of syntax*, pp. 579–603, Routledge, London.

Van Valin, R.D., 2014, 'Role and reference grammar', in *The Routledge handbook of syntax*, pp. 579–603, Routledge.

Waddington, H., Aloe, A.M., Becker, B.J., Djimeu, E.W., Hombrados, J.G. & Tugwell, P., 2017, 'Quasi-experimental study designs series – Paper 6: Risk of bias assessment', *Journal of Clinical Epidemiology* 89, 43–52. https://doi.org/10.1016/j.jclinepi.2017.02.015