ISSN :2584-2706 DOI: https://doi.org/10.67317/IJMSRT26OCT134

The Effectiveness of Interactive Storytelling in Enhancing Vocabulary Development among Preschool Children in International Schools in Negombo Sri Lanka

G.S.M. Silva

Abstract

The aim of this report is to formulate a critical evaluation of the concept of interactive storytelling for English vocabulary acquisition in international schools in Negombo, Sri Lanka with focus on preschool children. A course of eight weeks includes 3 20-minute sessions per week and the vocabulary will be evaluated using thirty picture naming items. The quantitative demonstration requires 30 simulated records completed from the data provided in the Excel file, but does not report completed fieldwork. The mean scores go up, from 14.87 to 17.43 out of 30, and the mean paired gain is 2.57 points. 20 cases get better, 3 do not change and 7 get worse. Individual variation and category patterns and item transitions are represented in descriptive tables and graphs. An illustrative paired comparison is reported, and there is explicit restriction of inferences from synthetic data. Themes for the teachers are not discerned because teacher narratives and interview responses are not available. A prospective reflexive thematic analysis framework specifies the evidence and interpretive procedures needed to explore implementation experiences. The Workbook offers a helpful analytical demonstration, but claims of effectiveness in education need the actual observation and a better design to support causal interpretation, the report concludes.

Keywords: *Interactive Storytelling; Preschool Vocabulary; Thematic Analysis; Sri Lanka.*

I. INTRODUCTION

1.1 Background and research problem

Vocabulary helps the preschool learner to express needs, to describe experiences and to engage in classroom interactions. In English picture naming, children must identify the meaning they see on the picture and match it with a suitable spoken word. These processes are similar but not the same. Even if a child knows an object in Sinhala or Tamil, he/she might not be able to say its name in English. Hence, this study is on selected English expressive vocabulary and not the English as a whole performance of language development. The suggested setting is the preschools in Negombo that are involved in the project. Home languages, home language instruction and English conversation outside of school can be varied in these settings for children. The level of proficiency at the start of the school year can thus not be assumed from the number of students enrolled in school. It is important to document language exposure and baseline performance to understand any differences that follow. This study does not assume that there is a loss associated with multilingualism or that all classrooms in the study model the same instructional approach.

Interactive storytelling presents an effective way to incorporate vocabulary into circle time. Children can identify pictures, anticipate what will happen, describe actions and participate in the retelling rather than listening all the time. Teachers can explain meanings, enrich answers and return to words in known stories. The pedagogical question is, how does this structured participation aid in a more independent retrieval of the material beyond the immediate activity? Learning may be accompanied by enjoyment, confidence and choral repetition, but only the latter three build vocabulary.

Research can give an "excuse" to evaluate but does not ensure success. Dowdall et al. (2020) found language benefits for all picture book interventions and Noble et al. (2020) reported that there was no child language benefits from an active shared reading comparison. The differences point to the need for instructional comparisons and outcome selection. The effectiveness of a specific programme cannot be confirmed in Negombo based on the other populations' evidence, as local instrument was prepared with thirty words that were deliberately selected.

1.2 Aim objectives and research questions

The goal is to explore the possibility of assessing the effects of an eight-week interactive storytelling program by measuring vocabulary growth and conducting contextual qualitative inquiry. The purposes are to compare pretest and post-test naming performance, to consider variation between participants and across the vocabulary categories, and to detail how teacher accounts might explain implementation experiences. The quantitative question relates to score changes and the qualitative question relate to teachers' description of participation, support and constraints in storytelling.

ISSN :2584-2706 DOI: https://doi.org/10.67317/IJMSRT26OCT134

The following report uses the numbers provided in a simulated workbook to calculate the quantitative procedures. It also creates a prospective thematic analysis framework due to the lack of substantive teacher narratives. There are different evidential statuses for these components: Numerical results describe synthetic records, and qualitative procedures are still proposed. The report thus illustrates the research process, not rather completed work in the field or claims of the children having achieved the improvements calculated.

1.3 Significance and scope

The practical contribution is an aligned assessment and analysis strategy that can be put into practice by teachers with the appropriate approvals and supervision. Thirty picture cards, with consistent scoring and the documentation of session delivery provide a manageable base from which to evaluate the classroom. Narrow scope: performance on taught English words immediately following an intervention. More evidence is needed for general vocabulary development, for retention and for causal effectiveness. These boundaries are used to interpret all the tables and graphs which will follow.

II. LITERATURE REVIEW AND CONPTUAL FRAMEWORK

2.1 Social interaction and instructional processes

The sociocultural approach to learning developed by Vygotsky (1978) justifies the use of supported interaction in learning. When helping children with a story, adults can try to have them use language they don't yet know. A teacher may accept a brief answer, demonstrate a more elaborate expression and prompt another response. The suggested learning process is to engage in the process, and then gradually reduce the amount of help given. The theory does not guarantee a set amount of score increase as a result of any given schedule.

The goal and approach of interactive storytelling should therefore be established by observable behaviours, not by the presence of a story book. Open questions allow verbal contribution, expansions provide more in-depth language models, and waiting time gives opportunities to develop one's language. Repeated retrieval links the story to independent naming. These practices need to be delivered sensitively, since a disjointed delivery of the story may mean that the story is split up and quieter children have fewer opportunities to contribute.

2.2 Evidence on shared reading and repetition

Mol et al. (2008) compared dialogic reading to the shared reading of typical passages. Their synthesis will help to explore the role that adult responsiveness and child participation play in vocabulary learning. But there are more one-on-one interactions between parent and child than there are in classroom circle time. This evidence should be considered when applying to a group context, taking into account participation opportunities, teacher workload and group size. Transfer of instruction procedures is not possible without taking the context of their delivery into account.

Flack et al. (2018) examined word learning in shared storybook reading, and how reading practices and exposure characteristics contribute. Their efforts help to record the number of times a target word is seen, not the length of the sessions. It may be beneficial to repeat words that are presented with clear pictures and questions than they are repeated in another word once. The present protocol thus introduces a vocabulary coverage matrix and attendance records and duration of sessions.

Horst et al. (2011) investigated contextual repetition in storybook learning. Their conclusions give justification for revisiting familiar stories, instead of adding new material to each session. Repetition can alleviate the load of 'unfamiliar events' while providing opportunities for continued focus on vocabulary. The three meetings per weekly story proposed in this document do so in the context of classroom practice. They don't replicate the original experiment or come to the conclusion that this adaptation has already been tested.

There is more general evidence that shared picture book interventions can facilitate language learning (Dowdall et al., 2020). Pooled results, however, cover various participants, instruments and delivery arrangements. The effect that has been estimated in the studies cannot predict what to expect on a locally designed naming test. Noble et al. (2019) similarly demonstrate the need to consider study design when interpreting shared reading evidence. Different questions are addressed by positive associations, intervention effects and improvements of active alternatives.

Noble et al. (2020) offer a helpful comparison: the changes in the reading behaviour of caregivers were not found to lead to an increase in language gains compared to an active control. This warns against using implementation success as outcome success. Teachers may engage in interactive routines repeatedly without measurable effects on the chosen measure. Better scores, on the other hand, might be due to normal learning or test familiarity. A well-rounded assessment should address the instruction and the performance separately.

ISSN :2584-2706 DOI: https://doi.org/10.67317/IJMSRT26OCT134

2.3 Measurement and qualitative explanation

The 30 picture assessment measures retrieval of specified English labels. Other factors such as image clarity, familiarity of the task, willingness to respond, and prior experience also may influence performance. Some adjectives like rough or heavy might be more difficult to make a clear description of than concrete nouns. A lower score cannot automatically be interpreted as a lower conceptual knowledge. It requires piloting and continual reminders and the instrument should be clearly identified as a researcher developed instrument not a standardised one.

Thematic analysis is related to meaning in qualitative material, rather than differences in numerical scores. Braun and Clarke (2006) describe a flexible process of engagement with data, coding, theme development and reporting interpretatively. Theoretical assumptions and analytic decisions are key points that must be aligned, as stressed by Braun and Clarke (2022). Teachers' accounts in this study could provide context for the experience and enactment of storytelling. Vocabulary categories and predetermined observation headings are not themes which come out of the participant accounts.

2.4 Conceptual framework

Figure 1 distinguishes the intervention, proposed learning processes and measured outcome. Vocabulary change is illustrated by repeated exposure, supported retrieval, and participation, and supported by age, language exposure, and classroom context. These are processes that are not tested mediators. The diagram would not be automatically accepted by a complementary qualitative strand which would examine teachers' interpretations of the implementation. The structure provides a space for inquiry and offers some room for "other explanations" and "contradictory experiences.

III. METHODOLOGY

3.1 Design and evidence available

Field study is planned, and the evaluation is a single group pretest–post-test evaluation of interactive storytelling. The present analysis only considers the participant records on the SIMULATED sheet of the workbook, which include thirty fake participant records. There are no completed answers in REAL_ENTRY, and no substantive narrative observations in SESSION_LOG. Provenance information: A repeated administrative description identifying a synthetic record is not qualitative evidence. This means that the material is suitable for a quantitative demonstration but cannot be used to complete the thematic analysis.

Teacher reflections and detailed observation notes were not collected, which would be done in the future if the study were extended to a qualitative analysis. A pragmatic focus would focus on numerical measures of performance and contextual accounts on implementation. This extension would be a combined design only after the qualitative data has been generated, analysed and meaningfully integrated. It should not be considered a completed mixed methods study simply because a thematic framework is mentioned in this report.

3.2 Participants and sampling

The proposed population is the children aged four and five of participating international preschools in Negombo. The number of children that can be engaged is a provisional feasibility number, not a representative number or a power number. Classes would be recruited with the permission of schools, and eligible children invited via caregivers. If recruitment, exclusions and missing follow up is achieved, then this would be reported. The 30 synthetic cases and three fictional classroom labels do not constitute actual recruitment or participation in an institutional setting.

Teachers providing the programme would be invited to report qualitatively, and sampling would be based on appropriate classroom experience and differences in implementation contexts. A teacher sample and a teacher saturation claim are not reported. The eventual scope of the qualitative research should be justified in terms of the richness, relevance and diversity of the material. Differences in the scores of the children should raise questions to explore rather than suggest answers; they should not narrow down what teachers explore or explain.

3.3 Assessment and intervention procedures.

The assessment consists of a total of thirty picture cards with ten nouns, ten verbs and ten adjectives. For each answer in English, a maximum of one mark will be awarded, for a total of 30 marks. Developmental pronunciation and grammatical variants are accepted which are recognisable. Variants and pronunciation are agreed and recognisable. If the answer to an item offered is incorrect or if there is no response to an item, the score for the item is 0. Items refused and/or interrupted or not administered will not be counted. Only when each of the 30 answers is correct will a total be calculated.

Assessment is given one on one from a series of neutral prompts, two practice pictures and up until ten seconds per item. No translation or initial sound/correction feedback is offered by teachers. The same pictures and scoring guide are used at the post-test, but with a different fixed sequence of presentations. The experts check the clarity of the pictures and their appropriateness for the sample culture, and then test them in a small sample of the culture. If it is allowed, independent rescoring, which is based on the consistency of the scores, also examines the consistency of the scores.

ISSN :2584-2706 DOI: https://doi.org/10.67317/IJMSRT26OCT134

The intervention will take place for 8 weeks, 3 days a week, 20 minutes of intervention per day, during circle time. There are 24 sessions for a total of 480 minutes. Each session takes about 3 minutes to introduce, 12 minutes to interactive story and 5 minutes to retrieve/re-tell. Five words are introduced a week per week in Weeks 1-6 with all words being consolidated in Weeks 7-8. Attendance, length, word coverage and deviations should be made from the planned schedule.

3.4 Quantitative analysis

Analysis covers record identifiers, binary item values and consistency of item totals and stored totals. Means, standard deviations, medians and ranges are provided as a summary for pretest, post-test and paired gains. Gains are computed for each record prior to aggregation. The scores are compared descriptively because the category sizes are not equal, which does not mean these categories are of equal difficulty. Individual trajectories and item response transitions vary around an overall average.

A paired t test, confidence interval and standardised paired effect are only calculated as a statistical demonstration. They have assumed the differences between the cases to be independent, and that the differences are sufficiently behaved. The class labels are not real and therefore do not contain any actual clustering information, and the actual classrooms might not be independent. Inferential calculations are not valid on synthetic records for population claims. The paired effect is the ratio between the mean difference and the standard deviation, and is specifically designated as dz.

3.5 Proposed thematic analysis

The qualitative extension would be based on reflexive thematic analysis with an experiential approach, where teacher accounts are considered as situated interpretation. The researcher would follow the procedures outlined by Braun and Clarke (2006) and re-read the text, code key parts, formulate potential themes, check the coherence of the themes, edit their meanings, and create an evidence-based narrative. These phases would not be process-driven, but rather iterative. An analytical journal dated with information about interpretation decisions and relationship between the researcher and the participating teachers would be developed.

First sensitising questions would be about participation opportunities, vocabulary support, language negotiation and constraints on delivering the language. They would draw the attention of the reader, but not act as "finding" points. Explanations that contradict and unsuccessful accounts of sessions would still have analytical significance. The themes would be each have a meaning at the core with contextualised extracts. “No quotations, frequencies or theme claims can be provided from the present workbook as there is no narrative material that fulfils the need for the same”.

As with Braun and Clarke (2022), reflexive interpretation would not be described as an activity of finding a single correct coding solution to a task. Supervisory discussion might involve questioning assumptions and provide greater interpretation without explicitly stating intercoder agreement as a quality criterion. This is different to independent checking of binary assessment scores. This difference should not be ignored when applying procedures that have been developed for numerical reliability to qualitative meaning making.

3.6 Ethical considerations

Institutional approval, school permission, informed parental consent and continuous child assent is required for fieldwork. Participation should not have an impact on educational access and/or grading. Children can take breaks or stop anytime without penalty. Different consent, confidentiality and careful removal of classroom information is required for teacher accounts. Records are recorded in the presence of participant codes, limited access and an authorized retention period. The presence of synthetic data and the preparation of this demonstration report may not be construed as granting any of these permissions.

IV. RESULTS OF THE SIMULATED ANALYSIS

4.1 Data audit and descriptive profile

Table 1 identifies materials that are available and others that are still needed. There are thirty valid binary cases (binary pretest and binary post-test cases). Totals are recalculated and match workbook. There are no assessment entries, nor are there any qualitative session narratives. This separation is crucial as the analytical functioning of the simulated observations does not provide a picture of response rate or a successful field recruitment or absence of missingness in a real study.

The age of children in the invented records ranges from 48 to 70 months with a mean of 59.87 and a standard deviation of 7.43. The average number of sessions attended is 20 and ranges from 16 to 24. These values are only for the constructed dataset. They are not able to confirm the exposure to storytelling, programme fidelity or the demographic profile of children going to international schools in Negombo.

4.2 Overall vocabulary change

The pretest mean is 14.87 points and the post-test mean is 17.43 points as shown in table 2. The standard deviation for the corresponding values is 6.15 and 4.46. The median rises from 16 to 18 and the ranges observed expand from 4–23 to 9–25. The average gain (mean) is 2.57 points, the standard deviation is 3.81, and the median gain is 3 points. These calculations are in the context of increased average performance, and they do not show an intervention effect.

ISSN :2584-2706 DOI: https://doi.org/10.67317/IJMSRT26OCT134

The mean accuracy increases by 8.56 percentage points from 49.56% to 58.11%. This is not a percentage increase of general vocabulary, but a change on the thirty-item instrument. Overall, the number of correct responses in the nine hundred item opportunities rises from 446 to 523. The difference of 77 responses when compared to the mean gain multiplied by 30 cases is a clear check for aggregation and reporting.

Figure 2 shows the parallel trajectories to preserve the variation. The 20 records improve, 3 are the same and 7 get worse, which is 66.7% of the records that improve, 10.0% that stay the same and 23.3% that get worse. The gains are between negative five and plus eight points. When reporting is reduced while progress is made, this leads to the impression of average progress when it should not. The synthetic distribution is used to demonstrate why both central tendency and the spread of individual change need to be evaluated in a classroom.

4.3 Statistical demonstration and category patterns

For the illustrative paired comparison, $t(29) = 3.69$; $p < .001$; mean difference (MD) = 1.14–3.99 points (95% confidence interval – CI). The standardised paired effect is $dz = 0.67$. These outputs are based on the specified model assumptions and summarise the values of the models. The simulated cases are not statistical evidence and the interval is not a range of uncertainty about the children in Negombo.

The three sets of words in each category are each worth 10 points and are compared in Table 4 and in Figure 3. Noun means increase from 5.23 to 5.73, verb means from 5.37 to 6.23, and adjective means from 4.27 to 5.47. The corresponding gains are 0.50, 0.87 and 1.20. The highest descriptive increase is seen with adjectives, but they also have the lowest starting level and so offer the greatest room for improvement. The differences between categories are not indicators of differences in effectiveness of instruction.

Substantive interpretation is also inappropriate in the simulation procedure. The workbook has lower initial probabilities for responses to adjective questions and calculates post-test results based on assumed probabilities of retention and acquisition. Items that are initially marked wrong are added to the generation rule for attendance. The patterns that emerge can be affected by the assumptions. The differences in the categories are therefore a construction and in themselves cannot substantiate theory on storytelling, attendance or the relative difficulty for the word classes in English.

4.4 Items transitions and qualitative evidence

Table 5 breaks down response transitions by the 900 paired item opportunities. The number of correct responses that are retained over both occasions is 373, while responses that are incorrect on both occasions is 304. There are another 150 incorrect-to-correct changes and 73 correct-to-incorrect changes. The difference between their two is 77. These are opportunities within an item, not nine hundred independent participants—descriptive interpretation.

Figure 4 shows all of the target words, not just positive ones. Rough has the greatest net gain with ten to eighteen responses and bridge has the greatest net loss, from sixteen to twelve responses. The aggregate count for Feathers is unchanged and the same applies to the other descriptions of boat, dry and short. It should be noted that unchanged counts do not imply that the same children answered the questions the same way on both times. There should be individual response transitions to differentiate between stable knowledge and offsetting acquisitions and losses.

No data of participant themes are reported. An evidence requirement matrix is given to help link potential inquiry areas to the material required for thematic interpretation; this is found in Table 6. Participation and language negotiation are not found themes, but rather areas of enquiry. There are no contextual extracts provided in the workbook for the teachers to analyse their meanings. The evidence gap would not be addressed by using numerical gains as themes or by making up quotations.

If true teacher notes or interview transcripts are available, then integration could involve comparing the interpretation(s) with the observed score patterns. For example, explanations for unequal opportunities for speaking could be explored together with uneven trajectories. Agreement would not establish the causation, disagreement would warrant analysis, not removal. The current report shows where integration would fit in-place and keeps any qualitative claims open until the gathering of the needed material.

V. DISCUSSION AND CONCLUSION

5.1 Interpretation and relationship to literature

The simulated analysis shows that it is possible to have a seemingly small average increase, but a large variance. A mean gain of +2.57 points is a measurable difference in the scores obtained in the constructed test but seven records go down. This contrast further emphasizes the need to look at distributions and individual trajectories. It does not excuse the child identification of successful or unsuccessful children, however, since the records are fabricated and there are no narrations describing the assessment behaviour or the learning experiences.

The direction of the numbers does not support the idea of vocabulary acquisition by means of supported interaction, but that is only a compatibility not a confirmation. Vygotsky's (1978) view supports the importance of studying guided participation, but does not

ISSN :2584-2706 DOI: https://doi.org/10.67317/IJMSRT26OCT134

support the determination of learning from synthetic scores. Only when a set of independent observations were made relating classroom processes to assessment outcomes would the framework become an empirically informative framework. What is currently known are still hypothesized mechanisms.
There is a literature, too, that warns against hasty generalizations. Dialogue, exposure and repetition are called for from the work of Mol et al. (2008), Flack et al. (2018) and Horst et al. (2011). However, Noble et al. (2020) show that the switch to interactive reading does not necessarily lead to further language benefits over an active alternative. The actual test of this will require an evaluation of the educational value provided by the storytelling beyond the ordinary provision and not a pretest–post-test increase.

5.2 Thematic contribution and limitations
The proposed qualitative strand may examine teacher conceptions of participation, their management of multilingual participation and their flexibility in storytelling in the face of constraints in the classroom. These accounts would help describe meaning and context, and would accompany score descriptions. Their contribution is not just to slap on thematic headings to statistical results, but to be detailed and to interpret. If there are no narratives, the report will not be able to identify whether confidence and/or peer support and/or teacher scaffolding defined participants' experiences.

Even if there is a real collection of data, there would be several limitations. It is not possible to distinguish between the influence of the intervention and maturation in one group design, or between the influence of the intervention and ordinary instruction or repeated testing. Expectancy risks are posed by teacher involvement in delivery and also in scoring. One limitation of convenience recruitment is the lack of transferability and the dependence in the classroom makes inferences difficult. Selected words are assessed immediately after instruction; the assessment does not measure other words or other environments of language development, retention or spontaneous use.

The current demonstration has a more basic boundary, that is, every filled participant record is simulated. Its statistical outcomes do not assess either school provision or the effectiveness of the programme. The observed pattern is partly due to the assumption used in the simulation and there is no empirical qualitative strand available. The above restrictions do not invalidate the analytical exercise; instead, they clearly indicate the proper place for the analytical exercise: preparation for transparent data collection, analysis and reporting.

5.3 Recommendations and conclusion
Prior to fieldwork, teachers are encouraged to test the cards and determine acceptable answers and set a reasonable timeframe for their delivery. The session should record ‘real time', attendance and individual speaking opportunities. Expectancy concerns would be reduced by independent assessment, if possible. A more robust later study would involve a similar group of classes getting ordinary story, several classes in each condition and a later assessment. These add-ons would help in interpretation, but not in any specific way.

There should be teacher interviews that elicit specific descriptions of sessions, such as times when children did not behave as expected. Questions should focus on what happened and how the teacher interpreted it, and what other explanations might be possible. It would be great to have context, contradictory accounts and changes over weeks recorded so that it could be interpreted more richly. These materials are anonymous and shared via classroom codes, so that careful comparisons can be made to other patterns of assessment, without identifying individual children and without suggesting causations.

To wrap up, the workbook is designed to be repeated in order to demonstrate quantitative vocabulary analysis, and to reveal the need for authentic qualitative material. Thematic approach provides a pathway to explore teacher experiences after having narratives. Claims of effectiveness need to be based on gathered data. Reporting improvements, declines, uncertainty and contextual interpretations together would lead to a more valuable contribution to preschool education than would treating a favourable simulated average as a completed research result.

REFERENCES

[1]. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
[2]. Braun, V., & Clarke, V. (2022). Conceptual and design thinking for thematic analysis. *Qualitative Psychology, 9*(1), 3–26. https://doi.org/10.1037/qup0000196
[3]. Dowdall, N., Melendez-Torres, G. J., Murray, L., Gardner, F., Hartford, L., & Cooper, P. J. (2020). Shared picture book reading interventions for child language development: A systematic review and meta-analysis. *Child Development, 91*(2), e383–e399. https://doi.org/10.1111/cdev.13225
[4]. Flack, Z. M., Field, A. P., & Horst, J. S. (2018). The effects of shared storybook reading on word learning: A meta-analysis. *Developmental Psychology, 54*(7), 1334–1346. https://doi.org/10.1037/dev0000512
[5]. Horst, J. S., Parsons, K. L., & Bryan, N. M. (2011). Get the story straight: Contextual repetition promotes word learning from storybooks. *Frontiers in Psychology, 2*, Article 17. https://doi.org/10.3389/fpsyg.2011.00017

ISSN :2584-2706 DOI: https://doi.org/10.67317/IJMSRT26OCT134

[6]. Mol, S. E., Bus, A. G., de Jong, M. T., & Smeets, D. J. H. (2008). Added value of dialogic parent–child book readings: A meta-analysis. *Early Education and Development, 19*(1), 7–26. https://doi.org/10.1080/10409280701838603
[7]. Noble, C., Cameron-Faulkner, T., Jessop, A., Coates, A., Sawyer, H., Taylor-Ims, R., & Rowland, C. F. (2020). The impact of interactive shared book reading on children's language skills: A randomized controlled trial. *Journal of Speech, Language, and Hearing Research, 63*(6), 1878–1897. https://doi.org/10.1044/2020_JSLHR-19-00288
[8]. Noble, C., Sala, G., Peter, M., Lingwood, J., Rowland, C., Gobet, F., & Pine, J. (2019). The impact of shared book reading on children's language skills: A meta-analysis. *Educational Research Review, 28*, Article 100290. https://doi.org/10.1016/j.edurev.2019.100290
[9]. Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press.

Table 1

Evidence audit of the supplied workbook

Material	Availability	Permitted use
SIMULATED	30 complete invented cases; 1,800 binary responses	Demonstration of pretest–post-test analysis
REAL_ENTRY	0 completed item responses	Template for future observations
SESSION_LOG	0 substantive narrative entries	No basis for thematic findings
Administrative notes	Synthetic record labels only	Provenance; not participant accounts

Note. Source: Storytelling_SIMULATED_Data_and_Real_Entry.xlsx. Blank real-data and narrative fields were not replaced or imputed.

Table 2

Descriptive vocabulary scores for simulated records

Measure	n	Mean	SD	Median	Range
Pretest	30	14.87	6.15	16	4 to 23
Post-test	30	17.43	4.46	18	9 to 25
Paired gain	30	2.57	3.81	3	-5 to 8

Note. Pretest and post-test scores range from 0 to 30; gain = post-test − pretest. All records are simulated. Accuracy means are 49.56% and 58.11%, respectively. No real population estimates are implied.

Table 3

Illustrative paired comparison using simulated values

Statistic	Value	Definition
Mean paired difference	2.57 points	Average post-test minus pretest
95% model-based interval	1.14 to 3.99 points	Mean difference under paired t assumptions
Paired t statistic	3.69	Degrees of freedom = 29

ISSN :2584-2706 DOI: https://doi.org/10.67317/IJMSRT26OCT134

Two-sided p value	< .001	Calculated p = .000926
Standardised paired effect dz	0.67	Mean difference / SD of differences

Note. Educational calculation only. Assumes independent case differences and an adequately behaved difference distribution. Not adjusted for classroom clustering. These statistics cannot establish a real intervention effect.

Table 4

Simulated vocabulary category scores

Category	Items	Pretest mean	Post-test mean	Mean gain
Nouns	10	5.23	5.73	0.50
Verbs	10	5.37	6.23	0.87
Adjectives	10	4.27	5.47	1.20

Note. Each category is scored out of 10, with 30 synthetic cases. Comparison is descriptive. Simulation rules set lower starting probabilities for adjective items; these values do not demonstrate category-specific teaching effects.

Table 5

Paired item response transitions in the simulated dataset

Pretest response	Post-test response	Item opportunities	Percentage
Correct	Correct	373	41.4%
Incorrect	Incorrect	304	33.8%
Incorrect	Correct	150	16.7%
Correct	Incorrect	73	8.1%

Note. Denominator = 900 paired item opportunities (30 cases × 30 items), not 900 independent children. Net increase = 150 − 73 = 77 correct responses. Percentages may differ slightly from 100% after rounding.

Table 6

Prospective thematic inquiry and evidence requirements

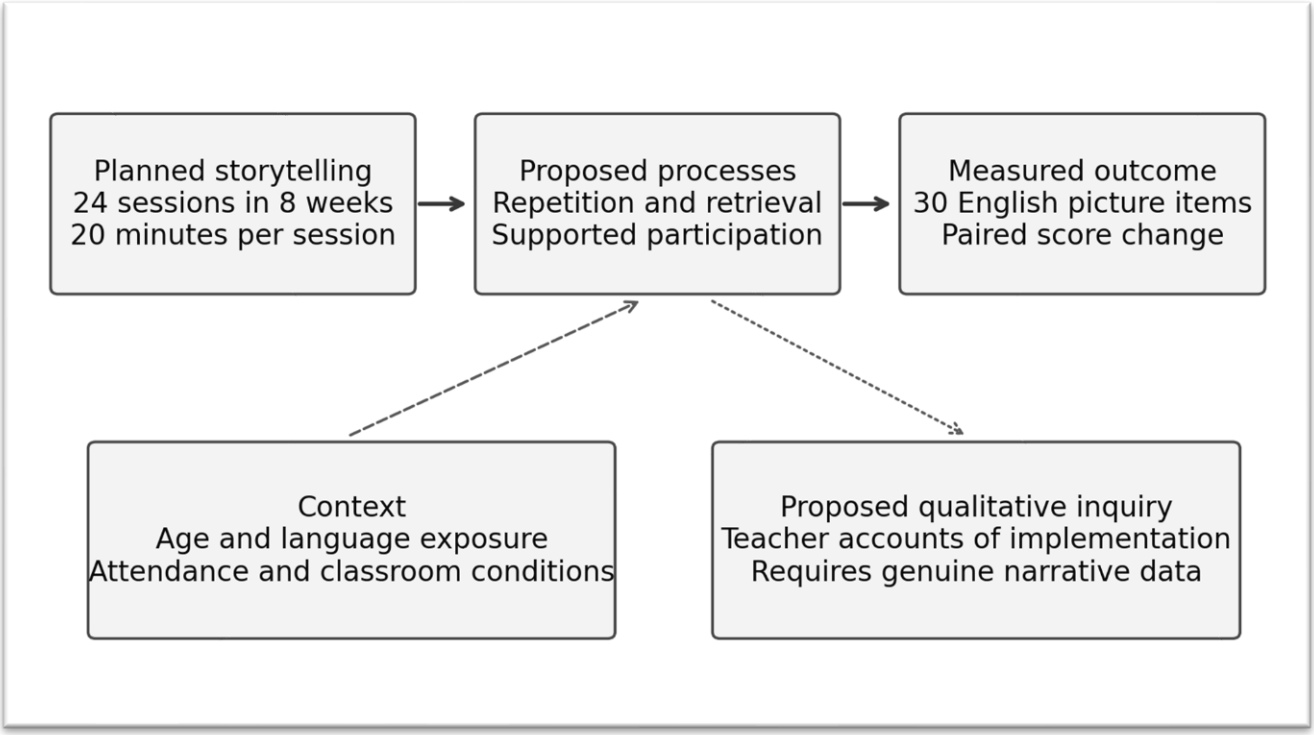
Inquiry area	Required qualitative material	Current status
Participation opportunities	Detailed episodes showing invitations, responses, silence and teachers' interpretations	No narratives available
Vocabulary support	Accounts of prompts, repetition and support, including unsuccessful attempts	No narratives available
Language negotiation	Contextual accounts of English and home-language responses and teacher decisions	No narratives available
Delivery constraints	Teacher accounts of time, materials, interruptions and adaptations	No narratives available

Note. These are sensitising areas for future inquiry, not empirical themes or a fixed coding result. Genuine transcripts or detailed notes are required. No invented quotations, theme frequencies or saturation claims are presented.

ISSN :2584-2706 DOI: https://doi.org/10.67317/IJMSRT26OCT134

Figure 1

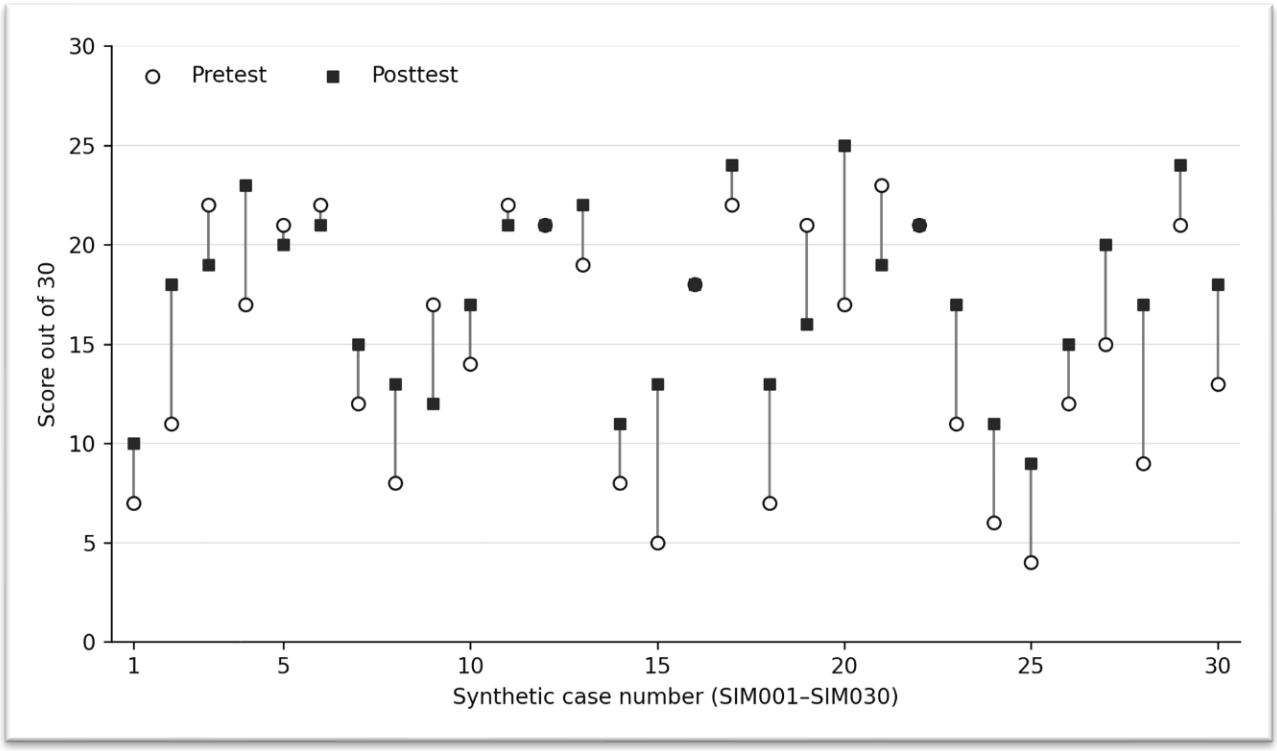
Conceptual framework and proposed qualitative extension



Note. Solid arrows express proposed theoretical relationships, not estimated causal effects. Dashed and dotted arrows locate contextual influence and future qualitative inquiry. No mediating process or thematic finding has been established.

Figure 2

Individual paired scores in the simulated dataset

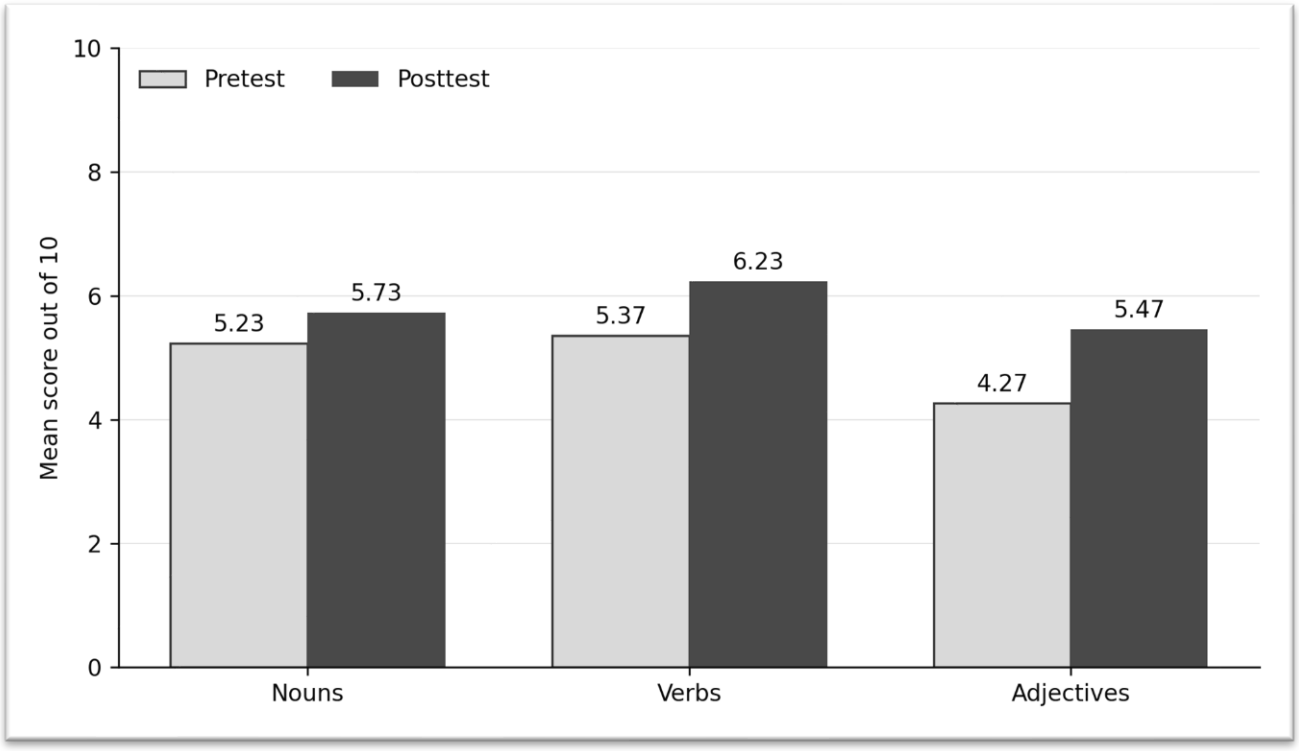


Note. Open circles indicate pretest scores; filled squares indicate post-test scores. Each connector belongs to one synthetic case. Twenty improve, three remain unchanged and seven decline. Source: SIMULATED worksheet.

ISSN :2584-2706 DOI: https://doi.org/10.67317/IJMSRT26OCT134

Figure 3

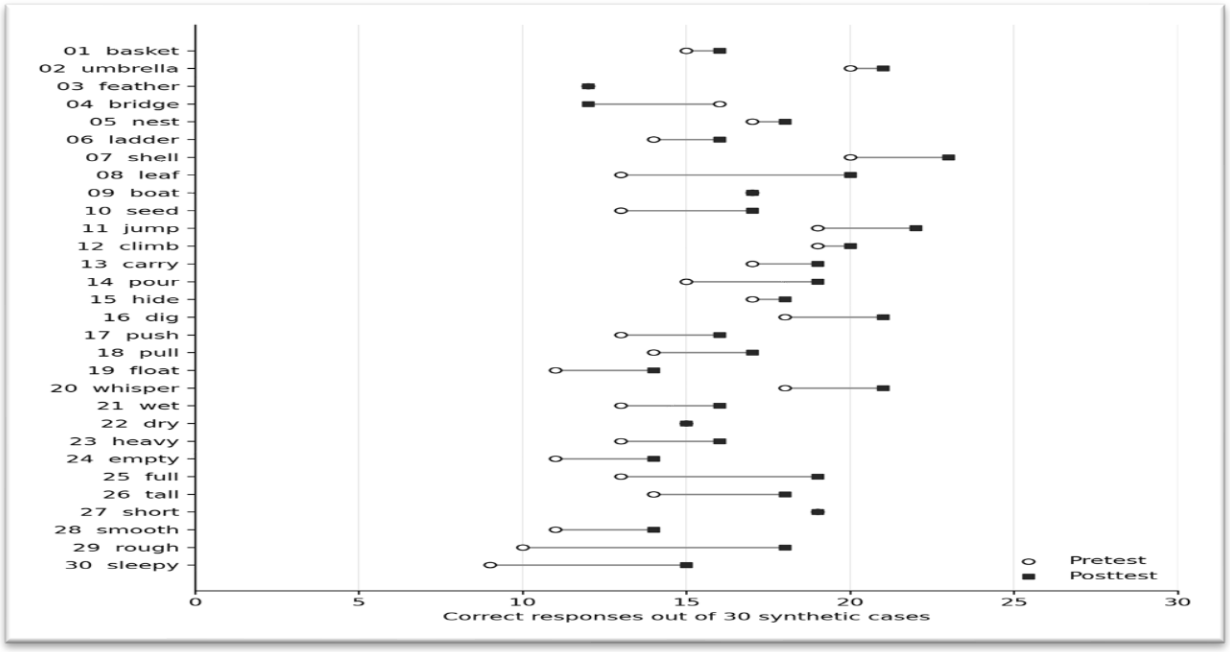
Mean category scores in the simulated dataset



Note. Each category contains ten items. Bars show descriptive means for 30 invented cases, not findings from participating schools. No uncertainty bars or inferential category comparisons are implied.
Source for quantitative tables and graphs: calculations from the supplied workbook's SIMULATED worksheet. Item totals and gains were independently recomputed from binary responses and reconciled with the stored totals.

Figure 4

Target word performance across simulated assessment occasions



Note. Each item has 30 paired responses. Open circles show pretest counts; filled squares show post-test counts. Aggregate counts can conceal offsetting individual changes. Source: SIMULATED worksheet.