

Relative Clauses in Reading Schemes: How does Children's Comprehension Align?

Rebecca Hunt

University of Sussex

Abstract: This project investigates the extent to which the Oxford Reading Levels Scheme reflects children's grammatical comprehension with regard to relative clauses. Two key factors are considered: embeddedness and focus. The project adopts two methodologies. Firstly, 40 reading scheme books are analysed, and any relative clauses are coded for embeddedness and focus. This demonstrates that the reading scheme is increasing complexity with level, signified by a higher number of relative clauses, and a greater diversity of them at higher levels. Secondly, an experimental study is conducted with 25 five- and six-year-old participants, in which they are provided with a sentence including a relative clause, and are asked to indicate which of three images matches the sentence. The relative clauses are manipulated according to their embeddedness and focus, to observe if these factors result in differences in comprehension. It is concluded that children comprehend object-embedded, subject focus (OS) relative clauses most easily, followed by subject-embedded, subject focus (SS), object-embedded, object focus (OO) and subject-embedded, object focus (SO). Collectively, these findings indicate that the use of OS and OO relative clauses in the scheme is reflective of children's comprehension, but further integration of SS and SO clauses may benefit children's comprehension abilities. This research is significant for children who have limited exposure to reading outside of school, to aid their grammatical development.

Plain English Abstract: This project investigates the extent to which one UK reading scheme reflects children's grammatical comprehension with regard to four distinct types of relative clause. The project adopts two methodologies. Firstly, all relative clauses were extracted from a sample of 40 reading scheme books. These relative clauses were then coded according to their type. This analysis demonstrates that the reading scheme is increasing in complexity with level, signified by a higher number of relative clauses, and a greater diversity of them at higher levels. Secondly, an experimental study is conducted with 25 five- and six-year-old participants, in which they are provided with a sentence including a relative clause, and are asked to indicate which of three images matches the sentence. They are tested on all four relative clause types, and the data shows that manipulating the clause type alters their understanding of the sentence. When the data from both stages of the project are compared, it suggests that there is a correlation between children's comprehension and frequency in the books for two of the relative clause types, but not for the remaining two. This research is significant for children who have limited exposure to reading outside of school, to aid their grammatical development.

Keywords: relative clauses; child language development; comprehension; reading schemes

1 Introduction

The purpose of this paper is to determine whether a reading scheme is reflective of children's comprehension of relative clauses. To achieve this, three research questions are proposed: (1) What is the distribution of relative clauses across the reading scheme? (2) What, if any, are the differences in children's comprehension of different types of relative clause? (3) Is there a correlation between the reading scheme data and the comprehension data?

Within this study, grammatical comprehension and complexity of the Oxford Owls reading scheme is observed and compared through the lens of relative clauses (RCs). RCs provide information about the noun phrase (NP), such as "that was covered in sprinkles" in "the ice cream, that was covered in sprinkles, began to melt". They were chosen as the focus of this study as the ability to manipulate RCs

into several forms (Brown, 1971; Sheldon, 1977; de Villiers et al., 1979) provides abundant opportunity for a comparative project such as this one. Although it is thought that children have a basic understanding of RCs by the age of five (Kidd, 2011), they are not explicitly taught until year five, when children are aged nine to ten (National Curriculum, 2013). This may therefore suggest that other linguistic input, such as a reading scheme, plays a part in RC acquisition.

At this stage, it is important to speak to the theoretical assumptions made in this article. The scope of this study is a descriptive comparison of the Oxford Owls scheme and RC comprehension, and does not extend to the critical examination of theoretical standpoints regarding the driving forces of language acquisition/development. Nevertheless, it must be acknowledged that the motivation for the study is grounded in the assumption that environmental factors, namely the exposure to a range of reading materials, play some role in children's language development. Thus, suggestions are made as to how the RCs in scheme could be altered to benefit children's comprehension of them; further investigation would of course be required to cement these.

1.1 Reading Schemes

Learning to read is a primary focus in a child's first years at school. The National Curriculum in England states that reading provides students with an opportunity 'to develop culturally, emotionally, intellectually, socially, [and] spiritually' (2013, p. 13). Across the UK, reading schemes are used to track children's reading progress throughout primary school (Oxford Owls, 2022). Studies have shown that children who spend more time reading outside of school have a higher level of reading proficiency, whereas those who read less receive lower reading scores (Anderson et al., 1988; Greaney, 1980)¹⁴.

Investigations into the adequacy of reading schemes have largely been vocabulary based. For example, studies by Hay and Spencer (1998) and Stuart et al. (2003) found that reading schemes develop from easily identifiable words to those that cannot be sight read. Although this is important, there has been a neglect of similar studies focusing on grammatical structures, which are an equally vital part of language development. The only reference to explicit grammar teaching in the year one English programme of study is through affixes (National Curriculum, 2013). The curriculum also states that children's overall language comprehension skills '[draw] from linguistic knowledge' (2013, p. 14) and can be developed by 'reading and discussing a range of stories' (2013, p. 14). This places further importance on the reading scheme to expose children to various grammatical structures, which it is thought will subsequently benefit their comprehension skills.

Within reading schemes, book bands are used to indicate the level that a child is working at. All schemes must match each of their books to one of these bands. In addition to this, some schemes choose to implement their own levels, to create greater distinction within the bands. One such scheme is Oxford Owls, the scheme used in this project¹⁵. In an email response, the Oxford Owls editorial team from the Oxford University Press (OUP) stated that in advancing to a new level, children will encounter 'incremental changes to test difficulty based on a range of features' (2022). Sentences that include a RC are introduced at Oxford level three, when children are four and five years old (OUP editorial team, 2022). However, it should be noted that at this level, 'multi-clause sentences [...] account for just 4%

¹⁴ It is important to recognise that a child's language environment is not necessarily reflective of their social class, and to acknowledge recent work in the US which challenges the idea that children from working-class families show linguistic deficits (e.g., Johnson, 2015). Nevertheless, reading schemes are accessible to all children in the UK through their school, and must therefore provide a level ground to support and encourage language development regardless of the child's language background.

¹⁵ For a visualisation of the Oxford Reading Levels, please see the following link:
<https://cdn.oxfordowl.co.uk/2019/07/19/13/52/18/160/OxfordLevelsAndBookBands.png>

of the total average' (OUP editorial team, 2022). Moreover, information supplied to authors of scheme books directs them to include 'longer and more complex sentences' (Bickler et al., 2003, p. 97) with each level, so it is likely that more RCs are included in higher level books.

To date, only one study can be found accounting for the frequency of RCs in reading schemes. Jardine (1992) aimed to establish the opportunities provided in reading schemes to gain familiarity with a variety of clause types. Across three schemes, RCs accounted for less than 6% of all clause types. Furthermore, RCs occurring after the verb were twenty-five times more frequent than those occurring before the verb. Jardine concluded that the schemes could be integrating more complex clause types to support children's language development. The present study builds on this by providing a more in-depth examination of relative clauses in a reading scheme, and mapping this to children's RC comprehension.

1.2 Relative Clause Comprehension

Two factors are considered in the literature of children's comprehension of RCs: embeddedness and focus. Embeddedness refers to the position of the RC within the sentence (de Villiers et al., 1979). Sheldon (1977), for example, would call a RC *subject-embedded* if it is embedded in the subject NP, and *object-embedded* if in the object NP, as in (1) and (2):

- (1) Subject-embedded: 'The boy that saw the girl hit the man.' (Sheldon, 1977, p. 306)
- (2) Object-embedded: 'The boy hit the man that saw the girl.' (Sheldon, 1977, p. 306)

However, the literature presents some variation when defining these terms. This is evident in example (3) below (Tavakolian, 1978), which indicates that it may not be as straightforward as *subject-* versus *object-embeddedness*. Here, *object-embedded* is not only used for direct objects, but prepositional objects too:

- (3) Object-embedded: 'The lion stands on the duck that bumps into the pig' (Tavakolian, 1978, p. 49)

Further variation in terminology is found in De Villiers et al.'s (1979) paper, where they refer to subject-embedded RCs as *centre-embedded*, and object-embedded RCs as *right-branching*. The tree diagrams in Figures 1 and 2 demonstrate this¹⁶, for two of de Villiers et al.'s example sentences (1979: 500). In Figure 1, the subject-embedded RC breaks up the main clause, 'the cat chased the rat', therefore appearing central in the sentence structure:

¹⁶ The tree diagrams within this paper were created using TreeForm software (Derrick & Archambault, 2010).

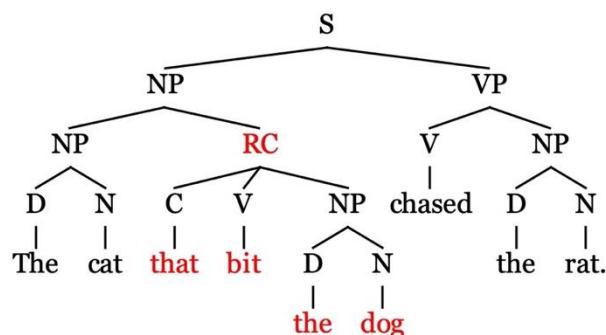


Figure 1: *Subject-embedded/centre-embedded relative clause.*

Contrarily, the object-embedded RC in Figure 2 follows the main clause, 'the cat bit the dog', and branches to the right:

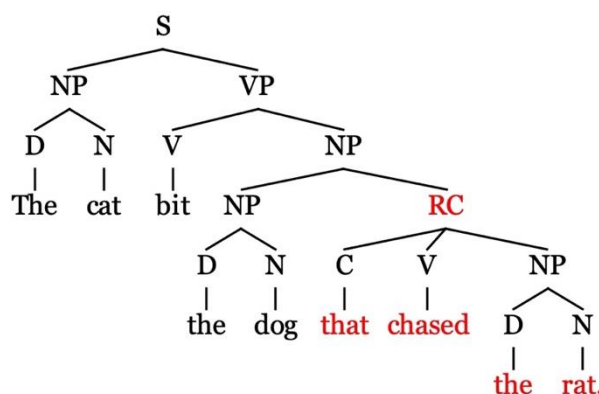


Figure 2: *Object-embedded/right-branching relative clause.*

These diagrams also suggest that an important distinction between subject- and object-embedded RCs is the depth at which the RC is embedded. The object-embedded clause in Figure 2 is embedded deeper in the sentence structure than the subject-embedded clause in Figure 1. This is visualised by it being further down the branches of the tree. Since English has subject-verb-object word order, subject-embedded RCs generally come before the verb, and object-embedded ones follow. Therefore, those preceding the verb, and consequently interrupting the main clause, can be termed *subject-embedded*; those following are *object-embedded*. Because of this link, and the fact that the majority of literature adopts this terminology, it is also used throughout this paper.

A large proportion of the literature provides evidence suggesting that children initially find object-embedded RCs easier to comprehend. Menyuk (1971, p. 139) concluded that 'operations which disturb the order of SV or SVO appear to be later acquisitions than those which do not', telling us that subject-embedded RCs, those that interrupt the main clause such as the example in Figure 1, may be more challenging for children. Supporting evidence comes from Gaer's (1969) experimental study, in which children aged three to six were shown an image and asked if a given sentence matched correctly. The results showed that children, on average across the ages, understood 66% of object-embedded RCs, in contrast to 61% of subject-embedded RCs. This difference was significant in children aged four, five,

and six. Gaer thus concluded that children find object-embedded RCs easier to comprehend than their subject-embedded counterparts.

In contrast, Lahey's (1974) study found that object-embedded RCs were more difficult for five-year-old children to understand. This study pioneered a new methodology in the study of RC comprehension, the act-out procedure, where children were given a set of toys and asked to act out sentences. This may produce more reliable results than a picture-cued comprehension task, as there are 'no constraints on a child's interpretation' (de Villiers et al., 1979, p. 501). The main focus of this study, however, was on prosody, not embeddedness, so the differing findings may be resulting from less control over extraneous variables. De Villiers et al. (1979) also suggest the conflicting research may be due to studies failing to account for focus alongside embeddedness.

The second factor, *focus*, can be defined as 'the role that the head noun plays in the relative clause' (de Villiers et al., 1979, p. 500)¹⁷. The literature centres on two categories: subject focus and object focus. To determine the focus of a RC, we can look at which element is missing from it. Consider example (4):

(4) The girl that cuddled the cat loved linguistics.

The RC in this sentence, 'that cuddled the cat', is missing a subject. It can therefore be said that the 'relativised [noun phrase] is a subject' (Sheldon, 1977, p. 306) and the RC has subject focus. In (5), below, 'that the cat cuddled' is missing an object; we do not know what the cat cuddled. This RC has object focus.

(5) The girl that the cat cuddled loved linguistics.

A notable study on the effect of focus on RC comprehension was conducted by Brown (1971). Children aged three to five were asked to indicate which of two pictorial stimuli matched a sentence. Although embeddedness was also investigated, focus was the only syntactic variable producing a significant effect. The children produced more correct answers for subject focus RCs than object focus, regardless of the embeddedness of the clause.

The typological model of the Accessibility Hierarchy (Keenan & Comrie 1977) can be linked to the focus of RCs¹⁸. As previously demonstrated, NPs of various types can be relativised, or *focused*. Between languages, there are differences as to which NP positions can be focused. For example, as Song (2002, p. 732) states, 'there are no languages in their sample that cannot relativize on subject although there are languages which can relativize only on subject'. The hierarchy is as follows (where > indicates 'more accessible than'):

Subject > Direct Object > Indirect Object > Oblique > Genitive > Object of
Comparison
(Keenan & Comrie, 1977, p. 66)

¹⁷ It is acknowledged that in recent literature this distinction is known as type of extraction: *subject-extracted* and *object-extracted*. However, *focus* is used in this paper to align with the previous relevant studies.

¹⁸ Other explanations have been proposed. Lau and Tanaka (2021) provide a good summary.

A language that can place focus on a certain NP position is expected to do the same for all preceding positions in the hierarchy. For instance, if a genitive (possessor) can be focused, it is assumed that subjects, direct objects, indirect objects, and obliques can be too.

Whilst this may not immediately elicit connections to child comprehension, there is a growing body of research examining the link between typological models and first language acquisition (Song, 2002; Slobin & Bowerman, 2007)¹⁹. Some studies have suggested that the Accessibility Hierarchy is reflected in the order of which children acquire RCs (De Villiers et al., 1979; Romaine, 1984). Whilst comprehension projects have not explored all NP positions, the hierarchy gives a basis to Brown’s findings that RCs with subject focus were more easily comprehended than those with object focus.

There is an increasing body of research accounting for both features simultaneously. This research discusses four RC types: subject-embedded, subject focus (SS); subject-embedded, object focus (SO); object-embedded, subject focus (OS); and object-embedded, object focus (OO). Table 1 provides an example for each of these types:

Table 1: *Examples of SS, SO, OS and OO relative clauses (de Villiers et al., 1979, p. 500).*

SS	The cat <u>that bit the dog</u> chased the rat.
SO	The cat <u>that the dog bit</u> chased the rat.
OS	The cat bit the dog <u>that chased the rat</u> .
OO	The cat bit the dog <u>that the rat chased</u> .

Research controlling both features has resulted in several conflicting conclusions. Table 2 shows the different difficulty rankings attested in six studies:

Table 2: *Relative clause difficulty rankings concluded from different studies (where > means easier than, and = means equal to).*

Order of difficulty (left=easiest):	Evidenced in:
OS > SS > OO > SO	Smith, 1974.
SS > OS > SO > OO	Tavakolian, 1978.
OO > OS > SO > SS	Romaine, 1984.
OS = SS > OO > SO	De Villiers et al., 1979.
SS > OS > OO > SO	Brown, 1971.
SS > OO > OS > SO	Sheldon, 1977.

There is not one RC type that has consistently been found to be easiest for children to comprehend. However, in four out of six of these studies, the subject focus RCs made up both of the two easiest types. Additionally, only one of the studies found the most difficult type to have subject focus. Regarding embeddedness, the SS type was found easiest in half of the investigations. Despite this, the most difficult RC type in the majority of studies was SO, the other subject-embedded clause. Collectively, this suggests focus may be of more importance than embeddedness in children’s RC

¹⁹ The strongest version of the Accessibility Hierarchy would somewhat counter claims that additional reading input would benefit children’s comprehension. However, these two ideas are not completely incompatible; many factors may be involved.

comprehension. Considering the results from the individual study of each feature, it appears that OS RCs are likely to be the easiest for children to comprehend. The wide range of results may arise from the studies adopting different methodologies, and working with children of different ages and socio-economic backgrounds.

Another direction of research determines whether children's comprehension of RCs is influenced by the inclusion or omission of a relative pronoun. In English, the relative pronouns are *who*, *whom*, *whose*, *which*, and *that* (Aarts, 2014). In RCs with object focus, the relative pronoun is not essential (Trask, 2000), as demonstrated in (6).

- (6) (a) The flowers that the woman grew were pink.
 (b) The flowers the woman grew were pink.

Alternatively, subject focus RCs must include the relative pronoun, as in (7), except in examples such as (8) that contain an auxiliary; in these cases the sentence is still grammatically correct when the pronoun and auxiliary are omitted (Trask, 2000).

- (7) (a) The woman that grew the flowers was happy.
 (b) *The woman grew the flowers was happy.
 (8) (a) The woman that is growing the flowers is happy.
 (b) The woman growing the flowers is happy.

Beaumont's (1982) study investigated the comprehension of RCs in seven-year-olds, based on the inclusion or omission of the pronoun. It was concluded that the inclusion of the pronoun was beneficial in the comprehension of object focus RCs; children understood an average of 1.53 out of 2 object focus RCs including the pronoun, which reduced to 0.79 when the pronoun was removed. However, this effect was not observed in the subject focus RCs. Straying from comprehension, a case study by Slobin and Welsh (1973) discovered that their two-year-old subject found it more difficult to repeat RCs that did not include a relative pronoun than those that did.

1.3 Hypotheses

Considering the background on reading schemes, and the previous literature on comprehension, the research questions can be addressed by a number of hypotheses. Although the main focus of the study is the four RC types, questions regarding relative pronouns are also included, which will contribute to answering the overarching research question of whether the reading scheme is reflective of children's RC comprehension.

RQ1: What is the distribution of relative clauses across the reading scheme?

Hypothesis (1): The reading scheme will increase in relative clause complexity with level.

(1a): There will be an overall increase in number of relative clauses.

(1b): There will be greater diversity of relative clauses at higher levels, in terms of the four types, and the use of relative pronouns.

RQ2: What, if any, are the differences in children's comprehension of different types of relative clause?

Hypothesis (2): The hypothesised hierarchy from easiest to comprehend to most difficult is OS > OO = SS > SO.

(2a): Children will show higher levels of comprehension when tested on object-embedded relative clauses (OS and OO) than subject-embedded relative clause (SS and SO).

(2b): Children will show higher levels of comprehension for relative clauses with subject focus (OS and SS) than object focus (OO and SO).

Hypothesis (3): Children will show lower levels of comprehension for relative clauses omitting the relative pronoun.

RQ3: Is there a correlation between the reading scheme data and the comprehension data?

Hypothesis (4): The use of relative clauses in the reading scheme will reflect children's comprehension levels. This means there will be more of the relative clauses they have good comprehension of, and less of those that they do not comprehend as easily.

2 Reading Scheme Analysis

An in-depth book analysis was conducted to propose an answer to research question (1). This section covers the methodology adopted, and the analysis and discussion of subsequent data.

2.1 Methodology

2.1.1 Materials

This step involved the analysis of 40 books from the Oxford Reading Levels. The sample includes five books each from levels five to twelve, which covers years one to three and thus children aged five to eight. Although this goes beyond the age of the experimental age group (five to six (§3.1.2)), I continued the analysis to level twelve to identify patterns in the emergence of RC types; only analysing the two levels for year one children would not provide opportunity to determine if the scheme was increasing in complexity. Analysing eight levels was the reason why only five books were sampled from each. The data collection process was done manually, so this ensured an accurate sample, whilst remaining realistic about time restraints.

The majority of books were sourced from two locations: a local primary school, and The University of Sussex's Education Department. Six additional books were randomly purchased. Whilst the sample took advantage of easily accessible books, it works well: all the books are matched to an Oxford Reading Level and, to my knowledge, are all currently used in schools. A full list of the books can be found in Appendix one. A preliminary analysis was conducted on five books at level four, which involved reading the books and noting any RCs. As none were found, these books are not included in the final sample.

2.1.2 Procedure

The data collection process began with the counting of all RCs in the 40-book sample. The number of NPs in each book was also counted. This includes proper and common nouns, alongside pronouns such

as *one* and *everything*. Personal pronouns and demonstrative pronouns were not included in this count, as they were not thought to contain RCs in this register of children's literature. This was confirmed through a check of all recorded RCs.

Next, it was noted if each RC contained a relative pronoun. Each RC was then manually coded according to its embeddedness and focus features. The number and proportion of each RC type was determined for each level. The coding system was inspired by the previous literature on children's comprehension of RCs, to ensure the results could be compared with the following experimental activity (§3), as summarised in Table 3:

Table 3: *Book analysis coding system.*

	Embeddedness	Focus
Subject	The RC is embedded in the subject (and precedes the verb).	The RC relativises a subject.
Object	The RC is embedded in the direct object or prepositional object (and follows the verb).	The RC relativises a direct object or prepositional object.

This system produces the four RC types: subject-embedded, subject focus (SS), subject-embedded, object focus (SO), object-embedded, subject focus (OS), and object-embedded, object focus (OO). Initially, all RCs that followed the verb were to be classified as object-embedded, thus including those embedded in grammatical constituents other than the direct or prepositional object²⁰. To maintain consistency between features, the object focus category also included the alternative constituents. However, initial analyses suggested that these categories were too large, and thus the methodology was adapted to ensure accurate analysis. Since the object categories in the comprehension study (upcoming in §3) were limited to direct and prepositional objects only, it was decided that the book analysis would also adopt this system. Because of this, 54% of the 159 recorded RCs could not be classified into one of the four categories. As analysis was also conducted on embeddedness and focus independently, there are differing total values for the two features: 81 (51%) RCs could be analysed for embeddedness, and 143 (90%) for focus. Whilst this reduces the data sample significantly, it is important to ensure an accurate and direct comparison can be drawn between the two stages of the study. There is certainly scope for further study to address RCs embedded in/relativised by the additional constituents. Throughout the analysis and discussion, where *other* is mentioned, this refers to RCs that did not meet the criteria for *subject* and/or *object* in terms of embeddedness and/or focus. For instance, RCs that were object embedded, but had 'other' focus, were not included. RCs that were not embedded in a sentence were also marked as 'other' for embeddedness. Additionally, 21 RCs were not embedded in a sentence, such as (9):

(9) 'Someone who didn't wipe their feet!' (Level 8: Hunt, 1990. *The Flying Carpet*)

Such cases could therefore not be categorised into one of the four SO, SS, OS or OO types, and also not analysis for embeddedness. They are, however, still included in the analysis of focus.

²⁰ This included subject predicative complements, existential sentences, pseudo-clefts, and adverbials.

Examples of the four RC types, taken from the books, are given in (10)-(13):

- (10) SS: 'Though anybody who wants to listen to horrid animal noises must be mad' (Level 11: Coldwell, 2010. *Amy the Hedgehog Girl*).
 (11) SO: 'The vase you're throwing balls at is worth seven hundred pounds!' (Level 7: Hunt, 1988. *The Old Vase*).
 (12) OS: 'He even loved doughnuts that had nothing special about them at all' (Level 10: Murtagh, 2017. *Dragon Doughnuts*).
 (13) OO: 'They finally found the bit they were looking for' (Level 10: Bradman, 2010. *Cuckoo Trouble*).

2.1.3 Data analysis

The first stage of analysis was to calculate the percentage of NPs containing a RC, for each book and then for each level, and the scheme overall. Both mean and median rates were observed. The mean was initially beneficial to identify any trends in the data, but the variability within levels meant further analysis was necessary. For example, within the level five books, four contained no RCs, and one contained four. Due to the small sample size, it wouldn't be justified to class this as an outlier, but calculating a median ensured the data was represented accurately.

To prepare the data for quantitative analysis, the number of each RC type was counted for the entire sample, and each level. As some of the analysis required a relative measure, a percentage was calculated to show the proportion of all RCs made up by each type; further percentages were then calculated for each level of the reading scheme. Embeddedness and focus were also considered separately, so equivalent calculations were made of subject- and object-embedded RCs, and then subject and object focus RCs. Finally, the number of RCs containing, and omitting, the relative pronoun, were counted, and relative percentages were calculated.

2.2 Analysis and Discussion

In the 40 books analysed, 8,186 NPs were recorded, 159 of which contained a RC. Across levels, this meant the average percentage of NPs including a RC was 2%. As Figure 3 shows, however, the higher-level books contain higher numbers of RCs than lower levels.

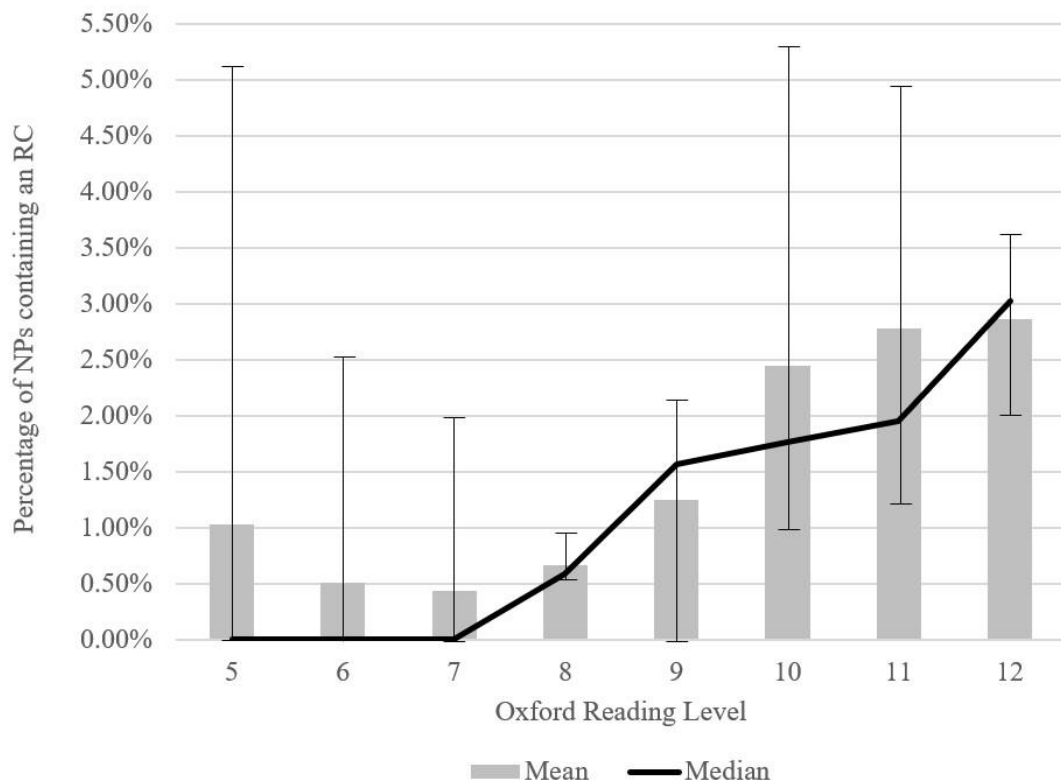


Figure 3: *The mean and median percentage of noun phrases containing a relative clause at Oxford Reading Levels 5–12, with range bars showing the ranges.*

Whilst the error bars display considerable variation within each level, there remains a strong positive correlation between reading level and percentage of NPs including a RC. A Spearman's Rho Correlation shows this to be a statistically significant²¹ finding: for the mean, $r_s=0.833$, $p=0.010$; for the median, $r_s=0.976$, $p<0.001$.

²¹ All statistical tests were conducted using SPSS.

2.2.1 Relative Clause Types

Figure 4 shows the frequency of each of the four RCs types across levels:

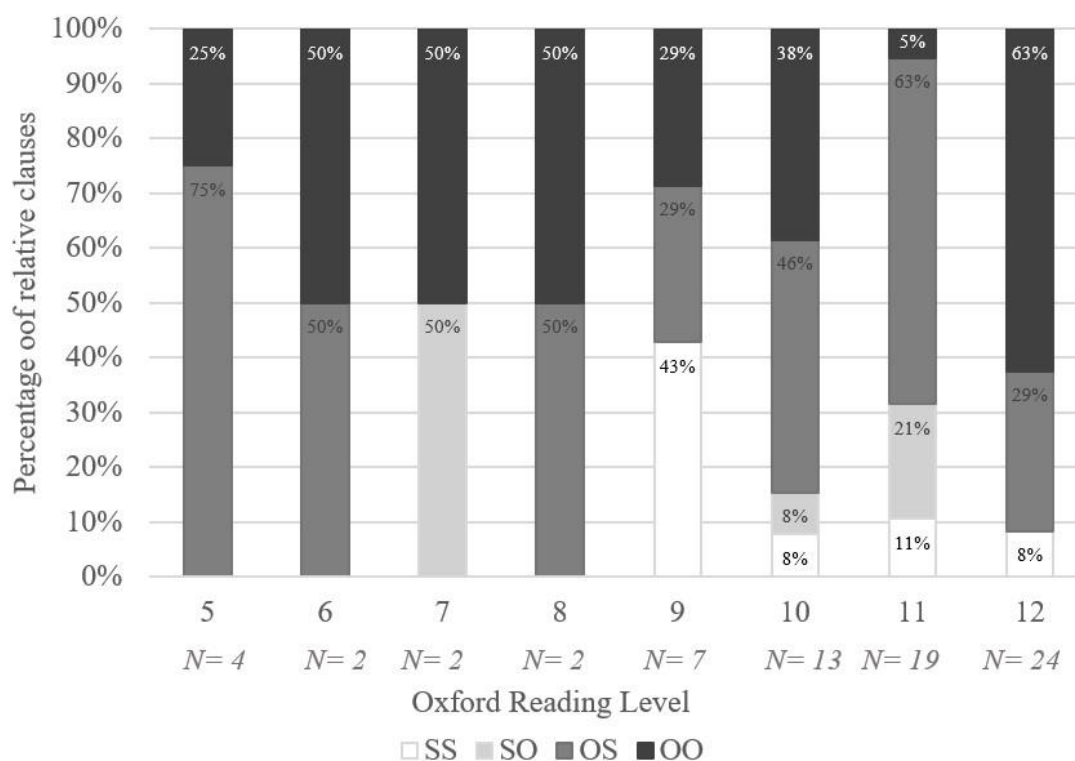


Figure 4: The distribution of SS, SO, OS, and OO relative clauses across Oxford Levels 5–12.

The higher levels demonstrate a greater variety of RCs, however only levels ten and eleven have examples of all four types. Object-embedded RCs, OS and OO, appear first, and remain predominant in all levels except seven. The subject-embedded RCs, SO and SS, are introduced at levels seven and nine respectively, but neither are consistently common across levels. Since some levels have very few RCs (levels six, seven, and eight have only two each), it is therefore useful to also examine the distribution of types across the levels as a whole. This data is shown in Figure 5:

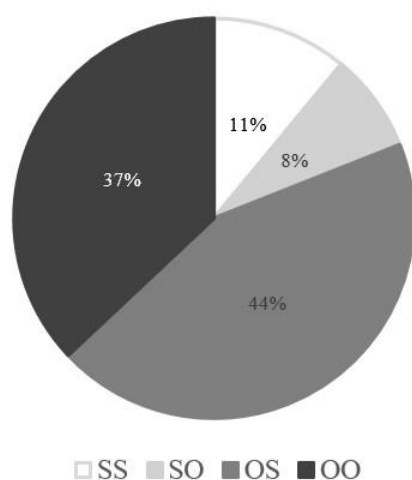


Figure 5: *The distribution of SS, SO, OS, and OO relative clauses within Oxford Levels 5–12 (N= 73).*

RCs coded as OS were most common across the levels, followed by OO, SS, and finally SO. Object-embedded RCs make up a majority of 81%. There also appears to be a preference for subject focus RCs; 55% of RCs have this feature.

Observing embeddedness and focus individually may further support these patterns. Object-embedded RCs are more common at all levels, except level seven in which there is an equal number of subject-embedded RCs (Figure 6). This trend is most prevalent in level twelve, where there are over six times as many object-embedded RCs as subject-embedded RCs.

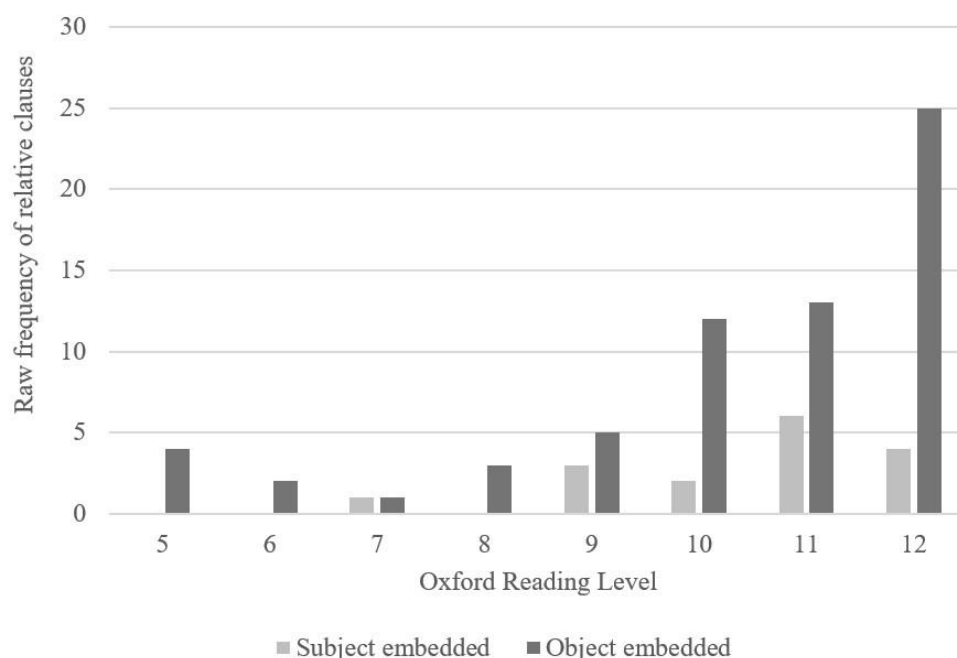


Figure 6: *The number of subject- and object-embedded relative clauses across Oxford Levels 5–12.*

Overall, object-embedded RCs accounted for 80% of all recorded RCs, with subject-embedded RCs accounting for the remaining 20%. This finding is proved statistically significant with a Spearman's Rho Correlation ($r_s=0.696$, $p=0.003$).

In relation to focus, the data is not as clear. Levels six, seven, eight, and twelve have a higher number of object focus RCs, whereas levels five, nine, ten and eleven have a more subject focus RCs, as pictured in Figure 7:

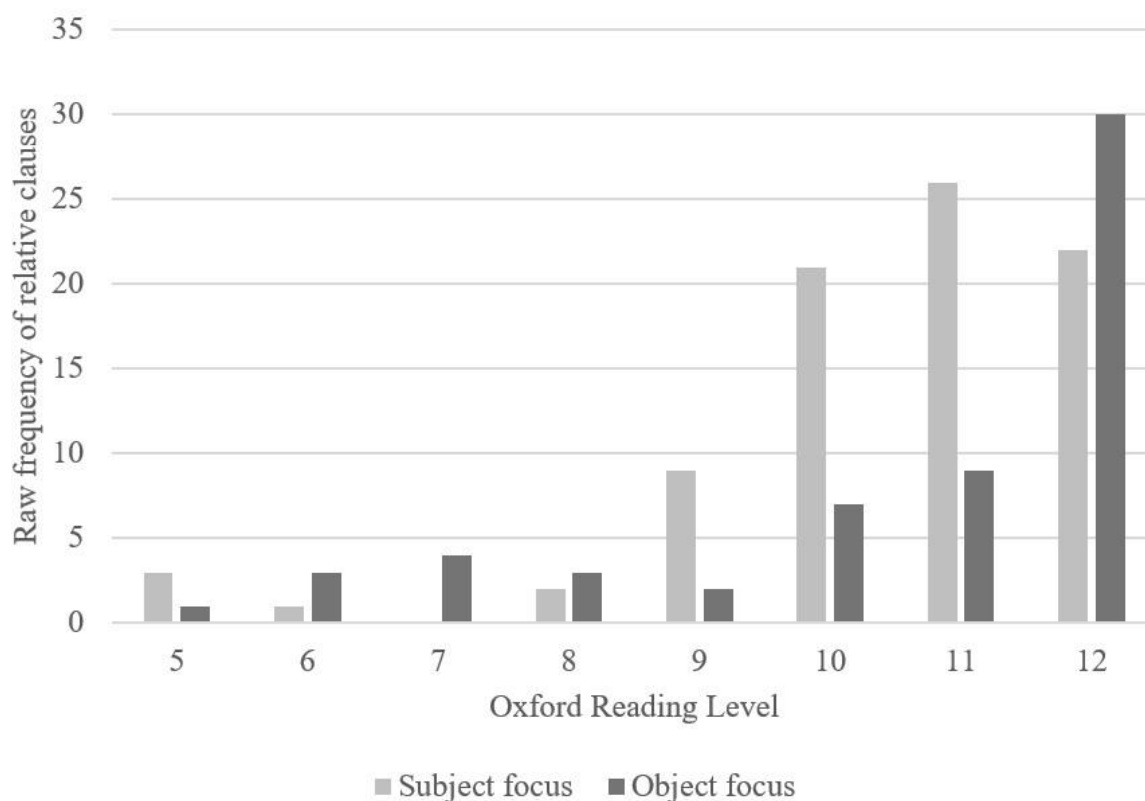


Figure 7: The number of subject and object focus relative clauses across Oxford Levels 5–12.

Insight into the distribution of the two focus types for all levels collectively is therefore necessary to determine which is more frequent. The data shows that subject focus accounts for more RCs (59%) than object focus (41%), by 18%, although this is not statistically significant (Spearman's Rho Correlation, $r_s=0.041$, $p=0.880$).

Following the analysis of features both collectively and individually, it can be concluded that object-embedded, subject focus RCs, OS, are most frequent in the books, and SO is the least. This supports the first hypothesis, 'the reading scheme will increase in relative clause complexity with level', as there is a general increase in the number of RCs with level, and more RC types are included at higher levels.

2.2.2 Relative Pronouns

The analysis of relative pronouns begins with the SO and OO RCs, as these are the categories in which the pronoun is consistently optional. This reveals that there is an overwhelming tendency to omit the relative pronoun. In 100% of the SO clauses, no relative pronoun was recorded. In the OO RCs, only 14% of them included a relative pronoun. As the SO and OO types were less frequent in the books, a

further calculation was made including the SS and SO clauses, to put the results into perspective. In the RCs that fit into the four specific categories, there were nearly twice as many RCs including a relative pronoun (66%) than omitting it (34%).

To address if RCs omitting the relative pronoun become more frequent at higher levels, Figure 8 was created. This shows the percentage of all RCs at each level including and omitting the relative pronoun.

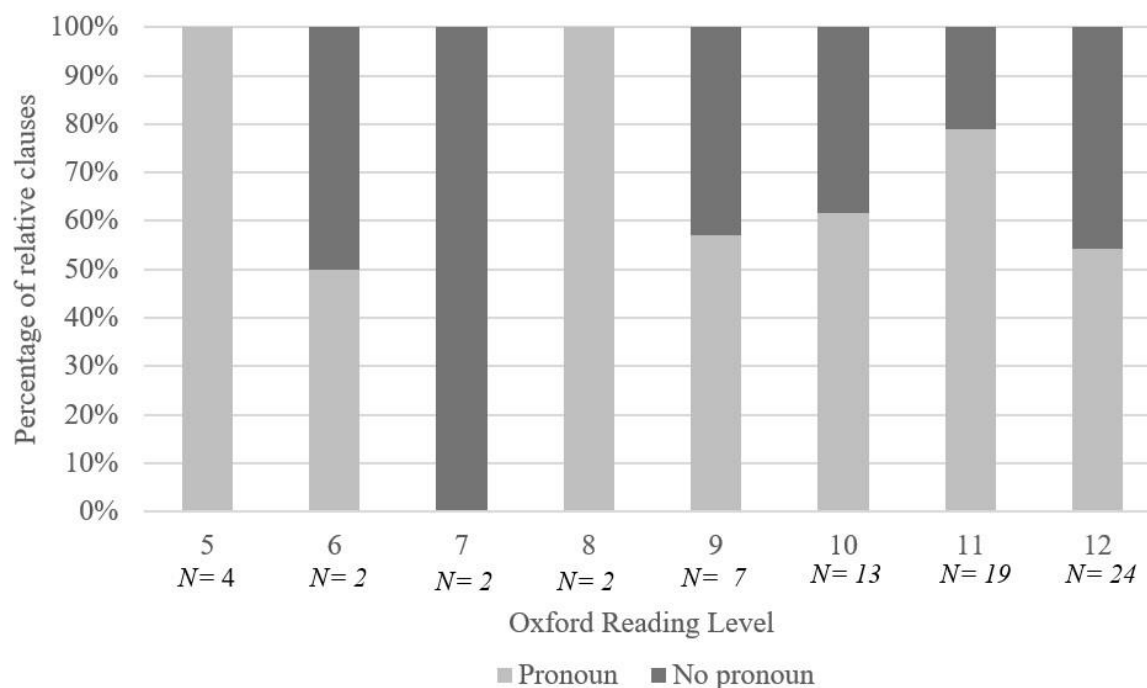


Figure 8: *The percentage of relative clauses at Oxford Levels 5–12 including and omitting the relative clause.*

Although the majority of levels contain more RCs including the pronoun, there is not a pattern regarding the emergence of those omitting the pronoun; it appears quite coincidental. This provides an uncertain answer to hypothesis (1b), that higher levels will integrate more diversity relating to the use of pronouns. However, these figures work with a very small sample size, and so further research with a larger sample would be necessary to provide a more accurate representation.

3 Comprehension Experimental Study

To address research question (2), an activity was conducted with 25 five- and six-year-olds to determine their comprehension of the different RC types.

3.1 Methodology

3.1.1 Materials

The first step in the activity design was to generate the sentences. These were based on the same definitions as used in previous literature, with RCs embedded in the subject being labelled *subject-embedded*, and those in the object, *object-embedded*. A similar distinction was made between *subject* and *object* focus. To remain consistent with previous studies (Tavakolian, 1978), the *object* categories also included prepositional objects, giving the sentences more diversity and making them more engaging for the children. Eight sets of sentences were created. Each set consisted of one control sentence, containing no RC, and one sentence containing each of the RC types. For the SO and OO types, a no pronoun option was also included. Each set of sentences contained the same nouns and the same verbs. Animals, such as *sheep*, *dogs* and *cats* were chosen as the nouns for all sentence sets as they are easily recognisable for the children. Likewise, the chosen actions, including *jumping*, *standing*, and *barking*, could be clearly illustrated, to reduce ambiguity.

Whilst Lahey's (1974) act-out procedure is thought to be 'methodologically the most sound' (de Villiers et al., 1979: 503), this required adaptation due to ethical complications associated with the recording of children. Without such recordings, the results would lack physical evidence. Thus, a picture-cued comprehension task, inspired by Brown (1971) and Gaer (1969) was developed. Three options were given instead of two, to counter Gaer's (1969) limitation of children having a 50/50 chance of being correct each time, which could skew results.

Each test sheet consisted of the three images displayed below a written sentence, as in Figure 9:



Figure 9: Example comprehension test sheet.

For each test sentence, one image included an animal not mentioned, one was the correct answer, and one showed a slight alteration depending on the RC type. For example, in Figure 9, the difference between the second and third images is which animal is wearing the hat.

From the previously created eight sentence sets, four groups of materials were formed. Each consisted of ten sentences: two controls, and two of each of the RC types. The material groups contained

a range of nouns and verbs to ensure that one particular stimuli, or set of nouns/verbs, was not producing a certain effect. In the SO and OO conditions, one sentence of each type included the relative pronoun, and one omitted it; this is the reason behind having two subsets, as the sentence including the pronoun alternated. An example group of stimuli sentences is given in Table 4:

Table 4: *Example group of materials.*

	Set 1a	Set 1b
SS	The cow that jumped over the cat stood on the table.	The cow that jumped over the cat stood on the table.
	The dog that barked at the sheep wore the hat.	The dog that barked at the sheep wore the hat.
SO	The horse that the goat sat on ate the grass.	The horse the goat sat on ate the grass.
	The pig the mouse sprayed rolled in the mud.	The pig that the mouse sprayed rolled in the mud.
OS	The rabbit sang to the chicken that flapped over the cat.	The rabbit sang to the chicken that flapped over the cat.
	The donkey jumped over the dog that wore the scarf.	The donkey jumped over the dog that wore the scarf.
OO	The frog taught the cake that the fish baked.	The frog taught the cake the fish baked.
	The snake hissed at the tower the fox built.	The snake hissed at the tower that the fox built.
Control	The cat jumped over the cow.	The cat jumped over the cow.
	The dog barked at the sheep.	The dog barked at the sheep.

3.1.2 *Participants*

A small pilot was conducted to confirm that the methodology was appropriate, and to determine if any changes needed to be made before beginning the main study. The participants were five L1 English-speaking children, aged five to seven, living locally. The participants in the main study were 25 five- and six-year-olds: twelve girls and thirteen boys. This age was chosen to ensure the children were independent readers, but still developing their grammatical knowledge. All participants were L1 English speaking pupils of a Year One class of a primary school in a working-class area of Devon. Please note this data collection process abided by the ethical policy of the University of Sussex and was fully approved by the ethics board. Prior to beginning the activity, I explained the process to each child and made sure they were aware they did not have to complete the task, and that it was not a test.

3.1.3 *Procedure*

After a brief acquainting period, the activity began. Each sentence was read aloud, and the child was asked to indicate which picture matched, by ticking the corresponding box. If the correct box was ticked, it was taken that the child comprehended the sentence. It is acknowledged that the sentences were both

written and read aloud, meaning conclusions cannot be drawn specific to reading or listening comprehension. Initially the sentences were to be read aloud; the children would not have the sentence written in front of them. However, when I conducted the experiment (Spring 2022), Covid-19 remained a concern, and so I wore a mask and sat at a distance from the children to minimise risk to both me and the children²². Alongside classroom noise, this raised doubt on whether the children would be able to hear me. Thus, it was decided for both written and oral stimuli to be included.

After completing the activity with the first pilot participant, a small change was made. Two additional test stimuli were added to the beginning of the study, to act as a buffer, ensuring the children understood the task. The data from these was not recorded for analysis. The pilot confirmed that this activity was appropriate for five- and six-year-old children. The data of three pilot participants who were this age (and were not the first pilot participant) was included in the overall data sample. Following the pilot, there were no additional changes made to the experimental paradigm, the children were not tested twice, and no data analysis was conducted on the pilot data alone. The remaining pilot participants were seven years old, and so their data was not included.

Children in the main study were individually called out to a table just outside the classroom to complete the tasks, throughout the afternoon of a normal school day. The process took five minutes per child, meaning their learning was minimally disrupted.

3.1.4 Data analysis

Prior to analysis, all data was input onto a spreadsheet. The number of correct responses to each sentence type were recorded for each child. The data of any participants who got one, or both, of the control sentences wrong, was discarded, as it could not be confirmed that they fully understood the task. This meant that data from seven participants was not analysed. For the remaining children, a mean average was generated for the number of correct responses to each RC type, including the no pronoun conditions. Additional calculations were made to determine the percentage of participants getting each number of RC types correct. For example, the raw data was input into the following formula to determine what percentage of participants answered correctly to all four subject-embedded RCs.

$$\frac{\text{Number of participants getting all subject – embedded RCs correct}}{\text{Total number of participants}} \times 100$$

Similar calculations were made for the number of participants getting three, two, one, and zero answers correct, and this was repeated for all RC types.

3.2 Analysis and Discussion

Overall, children correctly understood an average of 1.46 out of two RCs per type ($SD = 0.65$), demonstrating that by the age of five or six, they have a good comprehension of these clauses.

²² I also only went into the school after confirming I was Covid-negative.

3.2.1 *Embeddedness*

Table 5 shows the distribution of answers for subject- and object-embedded RCs. For example, 5% of participants correctly responded to only one of the subject-embedded RCs, meaning they incorrectly responded to the remaining three.

Table 5: *Percentage of participants who gave each number of correct responses, comparing subject- and object-embedded relative clauses (N= 21).*

Embeddedness type	Number of correct responses (percentage of participants)				
	0	1	2	3	4
Subject	5%	5%	38%	24%	29%
Object	0%	5%	14%	38%	43%

Notably, 43% of participants correctly comprehended all four of the object-embedded RCs, in contrast to only 29% correctly responding to all of the subject-embedded RCs. Furthermore, only 19% of children got fewer than three object-embedded RCs correct, as opposed to 48% for subject-embedded. These percentages suggest that children found object-embedded RCs easier to comprehend than subject-embedded RCs. This is supported by the data in Figure 10, which shows the average number of correct answers for each embeddedness type.

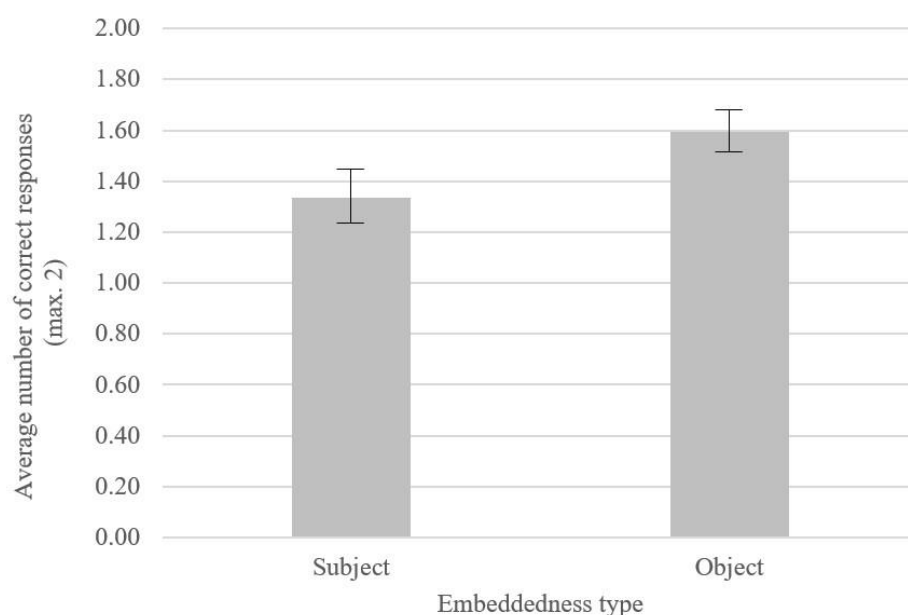


Figure 10: *The average number of correct responses for subject- and object-embedded relative clauses (max. 2), with standard error bars (subject-embedded = 1.222–1.444; object-embedded = 1.511–1.679).*

Children's scores for object-embedded RCs ($M=1.595$, $SD=0.544$) were 0.27 higher than subject-embedded ($M=1.333$, $SD=0.721$) RCs. An independent t-test does not show this as statistically

significant ($t(82)=-1.879, p=0.064$), and therefore cannot provide support for hypothesis (2a) 'Children will show higher levels of comprehension when tested on object-embedded relative clauses (OS and OO) than subject-embedded relative clauses (SS and SO)'. Nevertheless, the finding is suggestive of a difference in understanding in line with this hypothesis. A potential explanation for this stems from Menyuk's (1971) observation (§1.2) that clauses which disrupt the SVO word order are acquired later than clauses that don't. The Anti-Interruption Hypothesis (Slobin, 1971) argues that this disruption is what makes subject-embedded clauses harder for children to comprehend.

3.2.2 Focus

Almost half of the children got all four subject focus RCs correct, as Table 6 shows. For object focus, only 14% answered all questions correctly.

Table 6: *Percentage of participants who gave each number correct responses, comparing subject and object focus relative clauses (N= 21).*

Focus type	Number of correct responses (percentage of participants)				
	0	1	2	3	4
Subject	0%	0%	14%	38%	48%
Object	0%	14%	33%	38%	14%

Furthermore, 47% of participants got fewer than three object focus RCs correct, compared to 14% for subject focus RCs. This implies that subject focus was easier for the children to comprehend, which is supported by the data displayed in Figure 11:

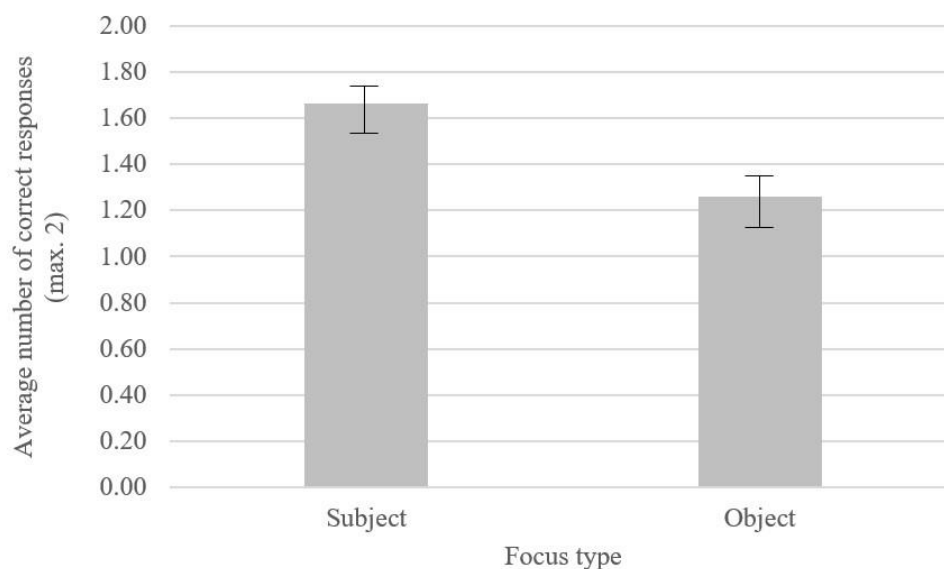


Figure 11: *The average number of correct responses for subject and object focus relative clauses (max. 2), with with standard error bars (subject focus = 1.586–1.748; object focus = 1.154–1.370)*

The number of correct responses was 0.4 higher for subject focus ($M=1.667$, $SD=0.526$) compared to object focus ($M=1.262$, $SD=0.701$). An independent t-test shows this as significant ($t(82)=$, $p=0.004$).

This conclusion is in line with the Accessibility Hierarchy (Keenan & Comrie, 1977), and Brown's (1971) finding that subject focus RCs were easier for children to comprehend (§1.2). The findings thus support hypothesis (2b), 'children will show higher levels of comprehension for relative clauses with subject focus (OS and SS) than object focus (OO and SO)'. Collectively, these results support a proposition by Slobin (1971), which suggests that subject focus RCs are easiest to comprehend because the underlying word order stays constant, unlike in object focus RCs.

3.2.3 Combined Relative Clause Types

Considering the features together, the OS type elicited the highest average number of correct responses, as displayed in Figure 12:

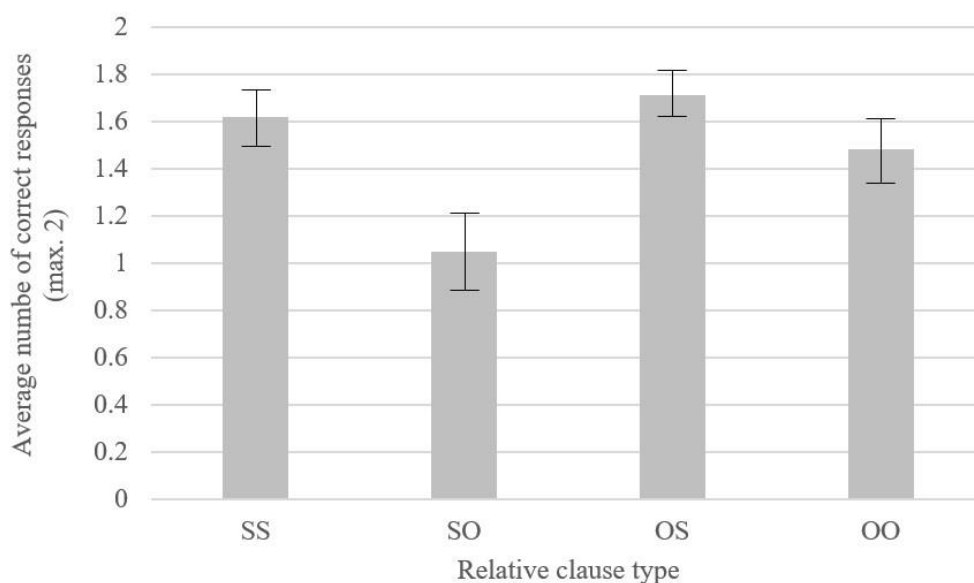


Figure 12: The average number of correct responses for the four relative clause types (max. 2), with standard error bars (SS = 1.490–1.748; SO = 0.887–1.209; OS = 1.613–1.815; OO = 1.345–1.607)

The OS type is closely followed by SS, then OO, and finally SO. The two subject RCs have the highest comprehension levels, suggesting that focus may be of more importance in RC comprehension than embeddedness. An ANOVA shows the differences in number of correct responses between RC types to be significant ($F=4.951$, $p=0.003$). A post-hoc Tukey analysis revealed that the significant differences were between SS and SO ($p=0.016$), and SO and OS ($p=0.003$). The SO RC therefore produced significantly lower correct responses than SS and OS, but not OO. Additionally, the correct responses were significantly higher for subject focus than object focus ($F=8.970$, $p=0.004$), but the embeddedness distinction was not statistically different ($F=3.531$, $p=0.064$). These findings provide further evidence for focus having a superior effect to embeddedness.

The hypothesised hierarchy of the RC types children would find easiest to comprehend is compared to the experimental findings in Table 7:

Table 7: *Hypothesised and concluded hierarchy of ease.*

Hypothesised hierarchy:	OS > OO = SS > SO
Concluded hierarchy:	OS > SS > OO > SO

Children's comprehension matched most of the hypothesised hierarchy. The only difference is a specification between OO and SS; the latter is shown to be easier. The suggested explanation of focus being more important in comprehension than embeddedness would fit this conclusion. Additionally, whilst no significant difference was identified between OO and SO, the conclusion that OO is easier to comprehend than SO is drawn from the raw data, in addition to the previous findings of object focus being easier comprehend than subject focus.

3.2.4 Relative Pronouns

The final variable to consider is the role of the relative pronoun. The SO and OO categories were first compared collectively, as the SS and OS sentences did not have a 'no pronoun' condition. The results of this show that there is an increase in comprehension of 0.02 for the no pronoun condition (0.64) in comparison to the pronoun condition (0.62). However, this is not a large enough difference to conclude that the inclusion or omission of the pronoun has an impact on comprehension. A different pattern emerges when analysing SO and OO RCs separately, as in Figure 13.

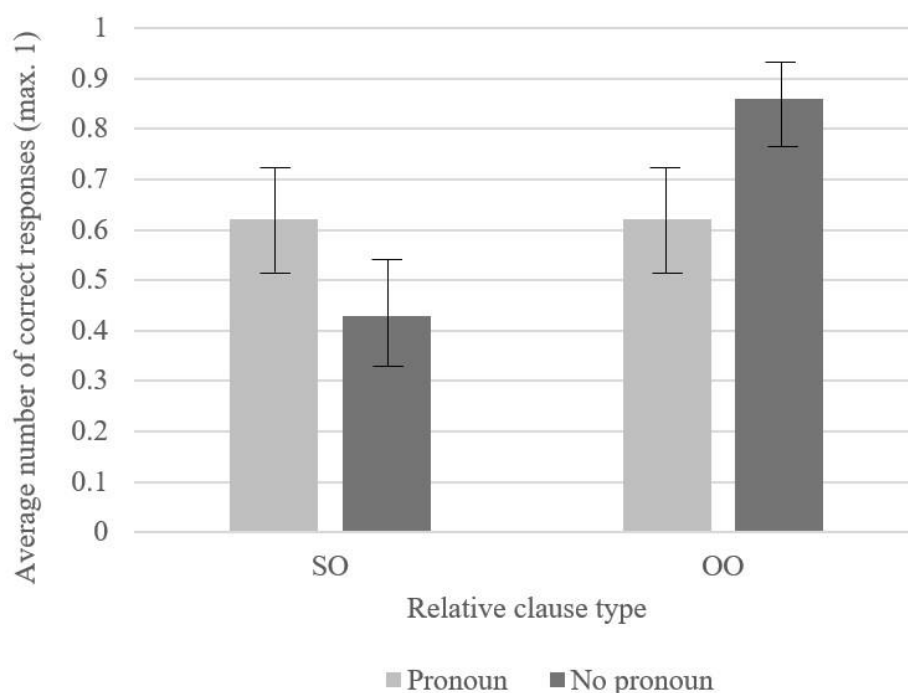


Figure 13: *The average number of correct responses for SO and OO relative clauses separately, dependent on the inclusion or omission of a relative pronoun (max. 1), with standard error bars (SO pronoun = 0.510–0.728; SO no pronoun = 0.318–0.540; OO pronoun = 0.510–0.728; OO no pronoun 0.779–0.935).*

The children found SO RCs easier to comprehend when they included the relative pronoun; the average score for this condition was 0.19 higher than the omitted condition. Contrarily, comprehension levels of OO RCs were higher when the relative pronoun was missing, by an average of 0.24. In fact, OO RCs with an omitted pronoun elicited the joint highest percentage of correct responses, with 86% of prompts receiving a correct response. This directly contrasts Beaumont's (1982) findings that this type was most difficult for children (§1.2). The difference between the SO no pronoun ($M = 0.429$, $SD = 0.507$) and the OO no pronoun ($M = 0.857$, $SD = 0.359$) conditions is statistically significant for this sample of children ($t(40) = -3.162$, $p = 0.003$). However, this cannot be generalised as the investigation of pronouns was a secondary objective of my study and consequently each child was only tested on two RCs without a pronoun: one SO and one OO.

To determine the overall effect of the pronoun, average scores were taken from all RCs with a relative pronoun: this consisted of all SS and OS RCs, alongside the SO and OO RCs that retained the pronoun. This was compared to the SO and OO no pronoun condition (Figure 14):

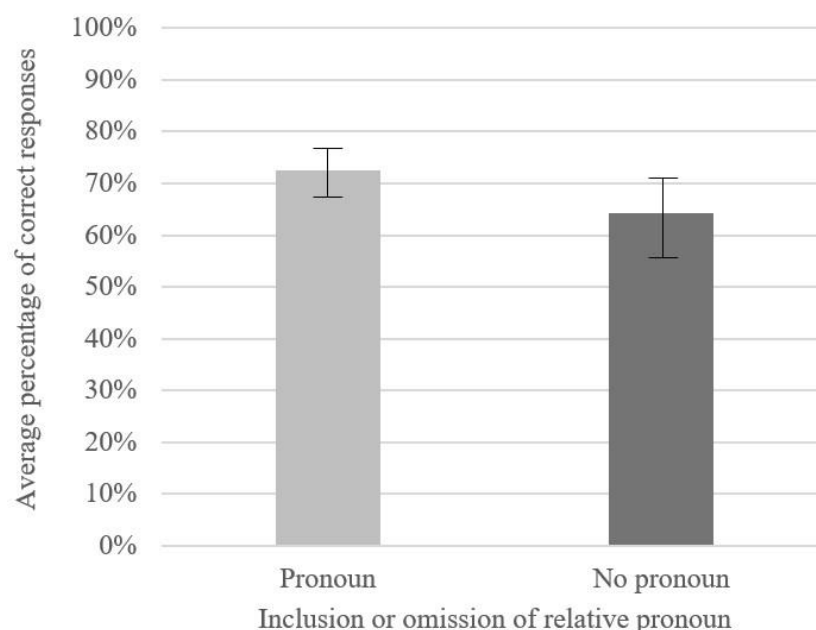


Figure 14: *The average percentage of correct responses for relative clauses including and omitting the relative pronoun, with standard error bars (pronoun = 68.19%–77.05%; no pronoun = 56.80%–71.77%).*

This shows that comprehension was 9% higher for RCs including a relative pronoun than those omitting it. Unlike the other figures, this displays percentages of correct answers, because the conditions were not equal in size. Children were only tested on two RCs without a pronoun, as opposed to six which included the pronoun. Considering all relative pronoun related data, it is therefore not accurate to confirm hypothesis (3), ‘children will show lower levels of comprehension for relative clauses omitting the relative pronoun’, however there is evidence of a slightly higher understanding of RCs that include the pronoun. To determine if these results withstand, the study would need repeating with additional control of relative pronouns.

4 Comparative Discussion

To determine if the reading scheme is reflective of children's RC comprehension and thus answer the overarching research question, the findings of both methodologies need to be compared.

In the book data, object-embedded RCs were four times as common as subject-embedded RCs (§2.2.1). In the comprehension activity, children found the object-embedded RCs easier to understand, with an average score 0.27 higher than subject-embedded RCs (§3.2.1). This demonstrates that, in terms of embeddedness, the scheme books are reflective of children's comprehension. To include a more challenging element that may benefit children's comprehension, the scheme could start integrating more subject-embedded RCs at an earlier level.

A similar correlation is evident between the RC frequencies and comprehension scores in terms of focus. In the comprehension activity, children produced 20% more correct responses for subject focus RCs than object focus RCs (§3.2.2). Subject focus RCs were also 18% more frequent than object focus RCs in the analysed books (§2.2.1). This highlights that the scheme appears to be mirroring children's comprehension abilities.

A Spearman's Rho Correlation was conducted to see if this reflection remains when RC features are combined to produce the four types. This used the total frequencies of each RC across all studied levels, and the average number of correct comprehension responses, as displayed in Table 8:

Table 8: *The frequency of each relative clause type compared to the mean number of correct responses.*

	SS	SO	OS	OO
Frequency	8	6	32	27
Mean number of correct responses (max. 2)	1.62	1.05	1.71	1.48

This was not shown to have a statistically significant correlation ($r_s=0.800$, $p=0.200$). However, the reason for this is evident in Figure 15:

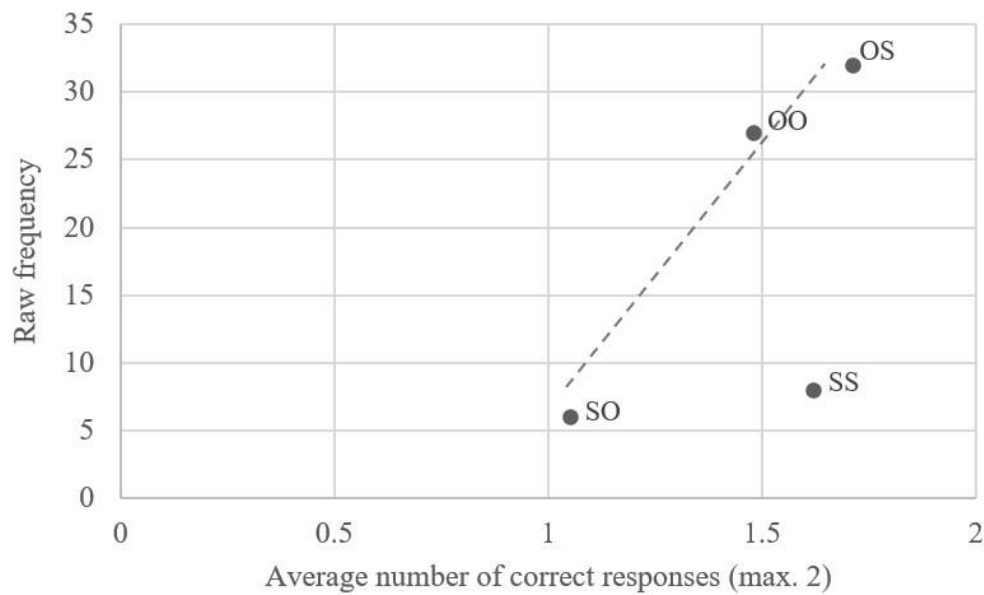


Figure 15: Average number of correct responses for each relative clause compared to the frequency of each type within Oxford Levels 5–12. The line represents what would be the overall trend, excluding the SS clauses.

There is a positive correlation shown for the OS, OO, and OS RC types. Yet, the SS type does not follow this trend. Children demonstrated strong comprehension of this type, so it was expected that they would be more frequent in the reading books than was found. The reading scheme could apply this finding by including a higher number of SS RCs in their books to act as consolidation. This imbalance between frequency and comprehension level could suggest that the reading scheme does not have a major influence on children's comprehension of RCs. If this was the case, the comprehension scores would be considerably lower for SS clauses. This is, however, a very ambitious claim to make, as there are multiple factors that could be influencing children's comprehension. It would be useful for any replications of this study to record children's linguistic environments at home, to see if this trend differs between those reliant on these schemes for a high proportion of their reading exposure, and those receiving additional reading stimuli.

Putting the SS outlier aside, a connection is established between the RC types that children are exposed to, and their comprehension of them. For the SO, OO, and OS types, the more frequent they are in the books, the higher the average number of correct responses. Despite correlation not equating to causation, it may be beneficial to include additional SO clauses in the scheme, as these had the lowest comprehension levels. If higher frequencies are the cause of increased comprehension, this will provide more opportunities for children to develop their understanding. Alternatively, in the case that the scheme is not responsible for any comprehension effects, including more SO RCs will still act as consolidation.

The final feature to consider is the relative pronoun. Both sets of data were somewhat inconclusive. RCs including the relative pronoun were 32% more frequent in the book data than those without, and also elicited a slightly increased comprehension performance. Within the SO and OO categories, only 14% included a relative pronoun. This is reflected in the OO comprehension condition, where results were significantly higher when the pronoun was excluded. However, for SO RCs, the

omission of the relative pronoun reduced correct responses. The inconsistency in these results makes it difficult to determine the relationship between the books and comprehension regarding relative pronouns. As stated, this was a secondary objective of the study, so further research may clarify results.

5 Conclusion

In answer to the main research question, the Oxford Reading Levels 5–12 are largely reflective of five- to six-year-old children's comprehension of RCs. Despite this, the results must remain situated in my data, and therefore cannot be generalised to all children, such as speakers of another language or of a different age, or to any other reading scheme.

As predicted, the analysed books increase in RC complexity with level, signified by the overall increase of RCs, and the increased diversity of RC types at higher levels. In terms of comprehension, the hypothesis was also supported, with children understanding object-embedded and subject focus RCs more easily than subject-embedded and object focus RCs respectively. This is realised in the concluded hierarchy of OS > SS > OO > SO. Although the findings regarding pronouns were not explicit, RCs without pronouns were still evident in the books, and children did demonstrate a good level of understanding.

The general alignment of the RCs in the reading schemes with the comprehension results shows that the scheme is providing RC complexity at a level appropriate for the children; this means they are provided with enough examples of the types they already understand in order to consolidate their knowledge, whilst also integrating types they do not yet fully grasp, to develop their comprehension. This is definitely the case for the OS and OO types, however it remains arguable if additional sentences with SO and SS RCs should be included in the scheme books. For children without a rich linguistic environment at home, this addition would ensure that they are being exposed to an adequate variety of RC types, to extend their grammatical development.

It is necessary to acknowledge the overarching limitations of this study to understand the restrictions of its application. Firstly, it is not certain that the children have read the particular books that were analysed; therefore, if data was collected from those that they had read, the conclusions may be different. Additionally, it may have been more beneficial to have analysed a greater number of books from fewer levels, to achieve more accurate averages. However, the variation within levels should not be so great that the sample is not representative. Another limitation is that the statistical tests, whilst verifying connections, should not be taken as the sole reason for a conclusion, because of the small sample sizes.

Whilst this study has been successful, the constraints of an undergraduate project have resulted in multiple remaining questions. Consequently, there are many avenues for further research that may answer these. If there was a greater interest in children's comprehension of RCs, there is space for further investigation on the effect of relative pronouns. In particular, it would be useful to test whether comprehension levels are dependent on which relative pronoun is used; are there differences between *who*, *that*, and *which*, for example. Additionally, a more extensive review of whether the inclusion of a pronoun in different RC types aids comprehension, would be beneficial to evaluate the findings of this project.

To give the conclusions of this study more context, future research could investigate adults' comprehension of these RC types, to establish what the goal is for children, and at what age they tend to reach adult-like competence. Similarly, an analysis of literature written for adults would be useful to put the numbers of RCs in the reading scheme books into context.

RCs were used in this study as a way of determining grammatical complexity. Thus, future projects could investigate other grammatical structures in reading schemes to increase reliability; if similar trends are found with structures such as anaphora, inflectional morphology, or tense changes, this would support that grammatical complexity is increasing as a whole, rather than just for RCs.

This study was inspired by the literature regarding children's sentence comprehension abilities, and expanded to observe the reading scheme, so the categories were based on previous research, which is not completely reflective of the RCs found in scheme books. If a future comparative project was to be conducted, it would be essential to ensure the experimental sentences are realistic of what children are exposed to in their reading. This would likely involve testing children on RCs embedded in and focused on SPCs, adverbials, and existential sentences, alongside any additional categories evidenced in the books. A wider sample of books is likely to reveal more RC types, so future comparative studies should conclude analysis of the books, before using their results surrounding RCs to determine their experimental conditions. This would ensure direct comparisons can be drawn.

6 References

- Aarts, B. (2014). *The Oxford Dictionary of English* (2nd ed) Online. Oxford: University Press.
Available at:
<https://www.oxfordreference.com/view/10.1093/acref/9780199658237.001.0001/acref-9780199658237> [Accessed 14/05/22]
- Anderson, R., Wilson, P., & Fielding, L. (1988). Growth in Reading and How Children Spend Their Time Outside of School. *Reading Research Quarterly*, 23(3), 285–303.
- Beaumont, C. (1982). Reading Relative Clauses. *Journal of Research in Reading*, 5(1), 29–42.
- Bickler, S., Baker, S., & Hobsbaum, A. (2003). *Book Bands for Guided Reading* (3rd ed). University of London: Institute of Education.
- Brown, H. (1971). Children's Comprehension of Relativized English Sentences. *Child Development*, 42(6), 1923–1936.
- Derrick, D., and Archambault, D. (2010). TreeForm: Explaining and Exploring Grammar through Syntax Trees. *Literary and Linguistic Computing*, 25(1), 53–66.
- De Villiers, J., Tager Flusberg, H., Hakuta, K., & Cohen, M. (1979). Children's Comprehension of Relative Clauses. *Journal of Psycholinguistic Research*, 8(5), 499–518.
- Gaer, E. (1969). Children's Understanding and Production of Sentences. *Journal of Verbal Learning and Verbal Behaviour*, 8(2), 1923–1936.
- Greaney, V. (1980). Factors Related to Amount and Type of Leisure Time Reading. *Reading Research Quarterly*, 15(3), 337–357.
- Hay, I., and Spencer, R. (1998). Initial Reading Schemes and Their High Frequency Words. *Australian Journal of Language and Literacy*, 1(3), 222–233.
- Jardine, J. (1992). Opportunities Provided by Reading Schemes for Becoming Familiar with a Range of Clause Types. *Educational Psychology*, 12(3–4), 275–289.
- Johnson, E. J. 2015. Debunking the “Language Gap”. *Journal for Multicultural Education*, 9(1), 42–50.
- Keenan, E., and Comrie, B. (1977). Noun Phrase Accessibility and Universal Grammar. *Linguistic Inquiry*, 8(1), 63–99.
- Kidd, E. (2011). *The Acquisition of Relative Clauses: Processing, Typology and Function*. Amsterdam; Philadelphia: John Benjamins Publishing Company.

- Lahey, M. (1974). Use of Prosody and Syntactic Markers in Children's Comprehension of Spoken Sentences. *Journal of Speech and Hearing Research*, 17(4), 656–688.
- Lau, E., and Tanaka, N. 2021. The Subject Advantage in Relative Clauses: A Review. *Glossa: A Journal of General Linguistics*, 6(1), 34.
- Menyuk, P. (1971). *The Acquisition and Development of Language*. New Jersey: Prentice-Hall.
- (The) National Curriculum in England: Key Stages 1 and 2 Framework Document. (2013). Department for Education. Available online at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/425601/PRIMARY_national_curriculum.pdf [Accessed 14/05/22].
- Oxford Owls Online. (2022). *Oxford Reading Tree & Oxford Levels*. Available at: <https://home.oxfordowl.co.uk/reading/reading-schemes-oxford-levels/oxford-reading-tree-levels/> [Accessed 14/05/22]
- Oxford University Press, Primary Literary Editorial Team. (2022). Email to this author. Received 25/04/22.
- Romaine, S. (1984). Relative Clauses in Child Language, Pidgins and Creoles. *Australian Journal of Linguistics*, 4(2), 257–281.
- Sheldon, A. (1977). On Strategies for Processing Relative Clauses: A Comparison of Children and Adults. *Journal of Psycholinguistic Research*, 6(4), 305–318.
- Slobin, D. (1971). Developmental Psycholinguistics. In Dingwall, W (Ed). *A Survey of Linguistic Science* (pp.267–316). Stamford, Connecticut: Greylock Publishers.
- Slobin, D., and Bowerman, M. (2007). Interfaces Between Linguistic Typology and Child Language Research. *Linguistic Typology*, 11(1), 213–226.
- Slobin, D., and Welsh, C. (1973). Elicited Imitation as a Research Tool in Developmental Psycholinguistics. In Ferguson, C, A., and Slobin, D (Eds). *Studies of Child Language Development* (pp. 485–497). New York: Holt, Rinehart, and Winston.
- Smith, M. (1974). Relative Clause Formation Between 29–36 months: A Preliminary Report. *Papers and Reports on Child Language Development*, 8, 104–110.
- Song, J. (2002). Linguistic Typology and Language Acquisition: The Accessibility Hierarchy and Relative Clauses. *Language Research*, 38(2), 729–756.
- Stuart, M., Dixon, M., Masterson, J., & Gray, B. (2003). Children's Early Reading Vocabulary: Description and Word Frequency Lists. *British Journal of Educational Psychology*, 73(4), 585–598.
- Tavakolian, S. (1978). The Conjoined Clause Analysis of Relative Clauses. *University of Massachusetts Occasional Papers in Linguistics*, 3, 37–48.
- Trask, L. (2000). *Dictionary of English Grammar*. London: Penguin.

7 Appendices

7.1 Appendix One: Book Data Sources

Level 4 (preliminary analysis):

- Burchett, J., and Vogler, S. 2009. *The Play Park*. Oxford: University Press.
- Burchett, J., and Vogler, S. 2011. *Tom, Dad and Colin*. Oxford: University Press.
- Hunt, R. 1997. *Everyone Got Wet*. Oxford: University Press.

Hunt, R., and Young, A. 2010. *No Tricks, Gran!* Oxford: University Press.

Hunt, R., and Young, A. 2010. *Painting the Loft*. Oxford: University Press.

Level 5:

Donaldson, J. 2006. *The Cinderella Play*. Oxford: University Press.

Donaldson, J. 2013. *Best Friends*. Essex: Pearson Education Limited.

Doyle, M. 2010. *Silly Jack and the Dancing Mice*. Essex: Pearson Education Limited.

Hawes, A. 2010. *Make a Wind Vane*. Oxford: University Press.

Llewellyn, C. 2009. *On the Wing*. Oxford: University Press.

Level 6:

Brandon, T. 2009. *Max the Detective*. Oxford: University Press.

Burchett, J., and Vogler, S. 2012. *Swoopie Mischief*. Oxford: University Press.

Donaldson, J. 2006. *Tara's Party*. Oxford: University Press.

Hunt, R. 1988. *The Dump*. Oxford: University Press.

Llewellyn, C. 2009. *Turn it Off!* Oxford: University Press.

Level 7:

Hunt, R. 1987. *Red Planet*. Oxford: University Press.

Hunt, R. 1987. *The Broken Roof*. Oxford: University Press.

Hunt, R. 1988. *The Old Vase*. Oxford: University Press.

Morgan, M. 2009. *Tiger's Discovery*. Oxford: University Press.

Rayner, S. 2009. *Hide and Cheat*. Oxford: University Press.

Level 8:

Burchett, J., and Vogler, S. 2010. *Attack of the Centipede*. Oxford: University Press.

Hunt, R. 1990. *The Flying Carpet*. Oxford: University Press.

Hunt, R. 1990. *The Rainbow Machine*. Oxford: University Press.

Hunt, R. 1990. *Victorian Adventure*. Oxford: University Press.

Poulton, M. 1988. *The Emergency*. Oxford: University Press.

Level 9:

Morgan, M. 2009. *The Deadly Boomslang*. Oxford: University Press.

Oakden, D. 1988. *Danger at Sea*. Oxford: University Press.

Penrose, J. 2008. *Pirates*. Oxford: University Press.

Poulton, M. 1988. *The Photograph*. Oxford: University Press.

Powling, C. 2009. *The Thing in the Cupboard*. Oxford: University Press.

Level 10:

Blank, A. 2009. *Human Body Adventures*. Oxford: University Press.

Bradman, T. 2010. *Cuckoo Trouble*. Oxford: University Press.

Bradman, T. 2010. *Underwater Adventure*. Oxford: University Press.

Middeton, H. 2009. *Let's Form a Band!* Oxford: University Press.

Murtagh, C. 2017. *Dragon Doughnuts*. Oxford: University Press.

Level 11:

Cartwright, P. 2010. *Awesome Animal Adventures*. Essex: Pearson Education Limited.

Coldwell, J. 2014. *Amy the Hedgehog Girl*. Oxford: University Press.

Ralphs, M. 2014. *Space Jump*. London: Collins.

Sparkes, A. 2014. *Scratch's Bad Reputations*. Oxford: University Press.

White, D. 2017. *Teeny Tiny Aliens and the Great Big Pet Disaster*. Oxford: University Press.

Level 12:

Bradman, T. 2014. *Making a Stand*. Oxford: University Press.

Clarke, Z. 2015. *Big Weather*. Oxford: University Press.

Gates, S. 2014. *The Lie Detector*. Oxford: University Press.

Gowar, M. 2008. *A Matter of Life and Death*. Oxford: University Press.

Knapman, T. 2017. *Extinct*. Oxford: University Press.

7.2 Appendix Two: Full List of Test Sentences

Set	Clause type	Sentence	
1	SS	The cow that jumped over the cat stood on the table.	SS1
	SO	The cow that the cat jumped over stood on the table.	SO1
		The cow the cat jumped over stood on the table.	SO1
	OS	The cat jumped over the cow that stood on the table.	OS1
	OO	The cat jumped over the table that the cow stood on.	OO1
		The cat jumped over the table the cow stood on.	OO1
	Control	The cat jumped over the cow.	Control1
2	SS	The dog that barked at the sheep wore the hat.	SS2
	SO	The sheep that the dog barked at wore the hat.	SO2
		The sheep the dog barked at wore the hat.	SO2
	OS	The dog barked at the sheep that wore the hat.	OS2
	OO	The dog barked at the hat that the sheep wore.	OO2
		The dog barked at the hat the sheep wore.	OO2
	Control	The dog barked at the sheep.	Control2
3	SS	The goat that ate the grass sat on the horse.	SS3
	SO	The horse that the goat sat on ate the grass.	SO3
		The horse the goat sat on ate the grass.	SO3
	OS	The goat sat on the horse that ate the grass.	OS3
	OO	The goat sat on the grass that the horse ate.	OO3
		The goat sat on the grass the horse ate.	OO3
	Control	The goat ate the grass.	Control3
4	SS	The pig that rolled in the mud sprayed the mouse.	SS4

5	SO	The pig that the mouse sprayed rolled in the mud.	SO4
		The pig the mouse sprayed rolled in the mud.	SO4
	OS	The mouse sprayed the pig that rolled in the mud.	OS4
	OO	The pig rolled in the mud that the mouse jumped in.	OO4
		The pig rolled in the mud the mouse jumped in.	OO4
	Control	The pig rolled in the mud.	Control4
	SS	The chicken that flapped over the cat sang to the rabbit.	SS5
	SO	The chicken that the rabbit sang to flapped over the cat.	SO5
		The chicken the rabbit sang to flapped over the cat.	SO5
	OS	The rabbit sang to the chicken that flapped over the cat.	OS5
6	OO	The chicken flapped over the cat that the rabbit sang to.	OO5
		The chicken flapped over the cat the rabbit sang to.	OO5
	Control	The chicken sang to the rabbit.	Control5
	SS	The donkey that wore the scarf jumped over the dog.	SS6
	SO	The dog that the donkey jumped over wore the scarf.	SO6
		The dog the donkey jumped over wore the scarf.	SO6
	OS	The donkey jumped over the dog that wore the scarf.	OS6
	OO	The donkey jumped over the scarf that the dog wore.	OO6
		The donkey jumped over the scarf the dog wore.	OO6
	Control	The donkey wore the scarf.	Control6
7	SS	The frog that baked the cake taught the fish.	SS7
	SO	The fish that the frog taught baked the cake.	SO7
		The fish the frog taught baked the cake.	SO7
	OS	The frog taught the fish that baked the cake.	OS7
	OO	The frog taught the cake that the fish baked.	OO7
		The frog taught the cake the fish baked.	OO7
	Control	The frog taught the fish.	Control7
8	SS	The snake that built the tower hissed at the fox.	SS8
	SO	The fox that the snake hissed at built the tower.	SO8
		The fox the snake hissed at built the tower.	SO8
	OS	The snake hissed at the fox that built the tower.	OS8
	OO	The snake hissed at the tower that the fox built.	OO8
		The snake hissed at the tower the fox built.	OO8
	Control	The snake built the tower.	Control8

About the Author

Rebecca Hunt graduated with a first-class degree in English Language and Linguistics from the University of Sussex in July 2022. This article is developed from her undergraduate dissertation, which was prompted by her keen interest in both structural linguistics and child language development. She then went on to complete her MA in Applied Linguistics and TESOL, graduating in January 2024. Her interests have since shifted slightly, and she is now in the first year of her PhD at Sussex, focusing on grammaticalisation in Cameroon Pidgin English.

Acknowledgements

I would like to thank all of the English Language and Linguistics faculty at the University of Sussex for their continued support throughout my time with them. In particular, thank you to my undergraduate dissertation supervisor, Professor Lynne Murphy, for her patience, encouragement, and advice whilst I was undertaking this project. I also thank the anonymous reviewers for their thoughtful comments and feedback, and the wonderful team at JoULAB for this opportunity.