DEPOSIT AND CONSULTATION OF DISSERTATION

One copy of your dissertation will be deposited in ORA (Oxford University Research Archive), where it is intended to be freely available online. In order to facilitate this, you are requested to complete and sign the form below.

Please use block capitals

Surname DING
First names (in full) YUQING
Faculty board EDUCATION
Degree name and pathway MSc Child Development and Education
Title of dissertation The relationship between children's narrative retelling and other aspects of language skills N.B. The title stated here must be precisely the same as that stated on the title page of the thesis submitted. **A candidate wishing to amend the title previously approved by the faculty must apply to the faculty board for permission to do so.**
Supervisor Charles Hulme
Subject keywords *Enter your own keywords or phrases to describe your work. This information helps us describe your work in ORA* Narrative retelling, oral language skills, standardized language tests
Research methods used *This information helps us describe your work on SOLO for future students e.g. quantitative, interviews, vocabulary test, systematic review, etc.* exploratory factor analysis, correlation, hierarchical linear regression

Declaration by the candidate as author of the dissertation

1. I understand that I am the owner of this dissertation and that the copyright rests with me unless I specifically transfer it to another person.
2. I understand that the Department requires that I shall deposit one copy of my dissertation in the Oxford University Research Archive ('ORA') where it shall be freely available online for use in accordance with ORA's Terms and Conditions of Use [https://ora.ox.ac.uk/terms_of_use].
3. I understand that this dissertation should not contain material that can be used to personally identify individuals or specific groups of individuals, and that such material should be removed before this dissertation is deposited in ORA.
4. I agree to be bound by the terms of the ORA Grant of Non-exclusive Licence [www.bodleian.ox.ac.uk/ora/deposit-in-ora/deposit-licence] and I warrant that to the best of my knowledge, making my thesis available on the internet will not infringe copyright or any other rights of any other person or party, nor contain defamatory material.
5. I agree that my dissertation shall be available for download in ORA in accordance with paragraphs 2, 3 and 4 above.

Signed [an electronic signature is sufficient]: YUQING DING
Date: 14/09/2021

The relationship between children's narrative retelling and other aspects of language skills

YUQING DING

Brasenose College

Candidate Number: 1057084

Word count: 17527

This dissertation is submitted to the Department of Education, University of Oxford, in partial fulfillment of the requirements for the degree of Master of Science in Child Development and Education

Acknowledgements

First and foremost, I would like to express my gratitude towards my supervisor, Charles Hulme for providing me with his dataset, without his kind support this project would not have happened. I also want to thank my pathway leader and college advisor Sonali Nag who encouraged me during this study. Without your support, I could not have completed my dissertation in a more orderly manner.

Secondly, I must acknowledge the Department of Education and all the library staff for offering an easier book renewal service online and organising departmental picnics to release our pressure when the world is on the verge of a pandemic. Furthermore, thank you for providing all the clear guidelines that supported me to complete this project. I was also blessed with all charming CDE course converters and instructors. I felt very lucky to study with such specialised researchers in this field and they have taught me wonderful courses.

Thirdly, I am indebted to my CDE cohort and accompanying friends Jiahao Amber, Xing Xu and Jiaxuan Rose. Thank you for all of your generosity and your encouragement helped me get through some of the most difficult periods this year.

Fourthly, I am most grateful towards my second rater Katherine Furness-Reed who was kindly involved in the study. I really appreciate your dedication and enthusiasm.

Lastly, I would like to express my heartfelt gratitude to my beloved parents and family. Thank you for your endless patience, love, encouragement, and faith in my abilities as an Oxford graduate student. Thanks again for always being my strongest backing for me on my way to pursue higher degrees. I am extremely appreciative of this journey, and thank you to everyone who is alongside made it possible.

Abstract

Background & Aim: Oral language skills are a critical foundation for educational success. Oral language skills are typically assessed with standardised tests involving diverse measures with highly constrained responses (e.g. CELF receptive vocabulary and sentence repetition subsets). Another measure of language skill that is typically absent from standardised tests is narrative retelling, where a child hears a story and is required to retell it. The aim of the current study is to investigate the correlation between children's narrative retelling and other aspects of oral language skills.

Methods & Procedures: The study involves secondary data analysis based on a sample of 600 children aged between 37 and 58 months. Exploratory factor analyses were performed to examine the factor structure of measures of oral language and narrative retelling ability. Correlations and hierarchical linear regression analyses were used to assess the relationship between variables.

Results: There was a strong correlation (r =.681, p <.01) between narrative retelling and other aspects of oral language skills. Oral language skills are a strong predictor of children's narrative retelling after controlling for demographic measures ($\Delta R^2 = 0.344$). Moreover, age, gender, and English as an additional language (EAL) made further contributions to the prediction of narrative retelling ($p <$ 0.05).

Conclusions: Language is a unitary factor. Standardised measures of oral language skill appear to reflect a unitary trait which is strongly correlated with a measure of oral narrative retelling. It appears that oral narrative retelling is potentially a useful measure of children's oral language skills.

Keywords: oral language skills, narrative retelling, language ability, exploratory factor analysis, correlation, hierarchical linear regression

Table of Contents

List of Abbreviations

ATLAS: Automated Tested of Language AbilitieS

CC: Complete Case

DW: Durbin-Watson

EAL: English as Additional Language

EV: Expressive Vocabulary

GDPR: General Data Protection Regulation

HLR: Hierarchical Linear Regression

INC: the Index of Narrative Complexity

KMO: Kaiser-Meyer-Olkin

LAD: Language Acquisition Device

LC: Listening Comprehension

LD: Listwise Deletion

MAR: Missing At Random

MCAR: Missing Completely At Random

MLR: Mean Length of Response

MLU: Mean Length of Utterance

MLU(m): Mean Length of Utterance in morphemes

MLU(w): Mean Length of Utterance in words

MNAR: Missing Not At Random

NDW: Number of Different Words

NELI: Nuffield Early Language Intervention

NSS: Narrative Scoring Scheme

ONR: Oral Narrative Retelling

PAF(PFA): Principal Axis Factoring

RV: Receptive Vocabulary

SchoolNo: School Number

TNW: Total Number of the Word

VIF: Variance Inflation Factors

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Overview

Oral language skills are a necessary basis for social engagement and academic achievement (e.g., classroom learning) (Hulme, et al., 2020). The development of oral language skills is one of the most impressive and natural achievements in children (Genishi, 1988). Almost all children acquire the norms of their language through use at a young age, as humans are born able to understand the rules of using language, according to Genishi (1988). Their oral language starts to become more complex, such as sentence selection, by age four or four and a half (ibid.). Standardised assessments (e.g., ATLAS, CELF) including a variety of measures with limited responses are commonly used to assess oral language skills. Previous studies have demonstrated that language is a complicated system, yet differences in individual's language skills may be represented by a unitary construct, as showed in Hulme et al. (2020), Klem et al. (2015), and Tomblin and Zhang (2006). Narrative retelling, in which a child listens to a narration and is expected to retell it in order, is another measure of language skills that is often ignored from standardised tests. Narrative ability is one of the best predictors of children's language and academic achievement and is crucial in early elementary education (Fazio et al., 1996; Pakulski & Kaderavek, 2001; Andreou & Lemoni, 2020). However, there is a lack of research on whether narrative retelling might become another effective measure of oral language skills. The present research builds on previous studies by exploring the relationship between narrative retelling and standardised measures of children's language skill.

1.2 Rationale for the Research

To date, studies that explore the relationship between children's narrative retelling and general language-related qualities are limited. The previous research only emphasised retelling or oral language as two independent variables that both predict children's language development, such as reading comprehension (Reed & Vaughn, 2012; Spencer & Slocum, 2010). Further analysis and research findings will be shown in Chapter 2. In addition, narrative retelling, as a subgroup of oral language skills, has been mostly studied as a useful tool to be applied in classes to assess children's reading abilities. None of the research further explored whether narrative retelling affects and even improves children's language skills. Moreover, the narrative retelling is commonly assessed based on the standardised Renfrew bus story (Haccoun, 2001), whereas the novelty of this paper also lies in the non-standard testing of children's narrative retelling with secondary data. This will be addressed in detail in Section 3.6.3. Overall, the motivation in this study is to bring new insights into the field of oral language development and measurements among younger children.

1.3 Aim of the Study

The present paper aims to fill a gap in research by investigating the relationship between children's narrative retelling and other aspects of language skills. In particular, current analysis has examined whether or not children's oral language skills were beyond the prediction explaining age, gender, EAL, and school number (SchoolNo.) demographics. In order to reduce bias in assessing the effect of the performance in narrative retelling and offer further empirical information, demographic measurements were incorporated. Microstructural metrics of productivity, lexical diversity, syntactic complexity, and narrative cohesion were employed as essential measurements to analyse the data gained through narrative retelling. The study's findings can help class teachers, educational policymakers, and schools make educated decisions concerning children's oral language development.

1.4 Significance of the Study

This study is significant because it divided an overall oral language into two parts: narrative retelling and other aspects of oral language skills. It contributes to the research in detail on the correlational relationship between the two in pre-school children. In particular, it has influenced the fledgling research on pre-schoolers' narrative retelling and its measuring approaches in UK schools. The current study relied on a large sample size of secondary data employing different measuring instruments. It will bring more reliable and convincing findings and insights for future school teachers or policymakers to amend or re-design their course plan to improve children's language development thoroughly.

1.5 Structure of the Dissertation

This dissertation is divided into the following five chapters:

Chapter 1 is the Introduction section outlining the background, the rationale, the aim, and the significance of carrying out the study.

Chapter 2 provides a literature overview on both oral language skills and narrative retelling, emphasising the distinction between macrostructure and microstructure and providing previous research on these topics. It discusses the potential demographic factors such as age, gender, and EAL that might influence the performance of narrative retelling. Research questions and hypotheses are also presented at the end of the chapter based on the existing studies.

Chapter 3 gives a detailed description of the methods used in the study, including the research design, sample data, analytical approaches, the selection and implementation of instruments, and reliability and validity of these instruments and measurements.

Chapter 4 provides preliminary findings of the statistical analyses towards each research question, including a detailed description of the statistics, analysis of missing data and the correlation coefficients indicating the correlation between two variables.

Chapter 5 concludes the chapter by further discussing the analytical findings, connecting the results with previous research, drawing concise conclusions, specifying limitations of the current study, and exploring some research implications for future research.

Chapter 2: Literature Review

2.1 Overview of the Chapter

This chapter provides the theoretical framework that underpins the research. It begins by considering language development, relying on three prime theorists: Piaget (1959), Brown (1973), and Chomsky (1986). Next, oral language skills are briefly introduced with a definition and an examination of prior studies. It then continues with an introduction to (narrative) retelling, including two-level microstructural and macrostructural measurements of retelling in academic functioning. The mean length of utterance will be explained as the main measurement for evaluating narrative retelling. Following this, this chapter discusses studies on the key characteristics of retelling, such as retelling as an assessment tool for reading comprehension. Finally, the research questions and hypotheses will be presented following the previous literature review and other study findings.

2.2 Language Development

Words are an essential component of linguistic knowledge (Ranti, 2015). Without them, we are unable to express our thoughts. Someone can acquire thousands of words yet does not know how to communicate in that language. Language acquisition is the process through which people learn to perceive and comprehend language, as well as to generate and utilise words and phrases for communication. Language acquisition is one of the core human qualities as non-humans do not communicate through language (Friederici, 2011:92). The term ‘language acquisition’ generally refers to the process through which infants acquire and learn their native language (Kosslyn & Osherson, 1995:34).

Learning is a multifaceted process that occurs through the interaction of children’s personal experiences with the outside world (Elkind, 1986; Montessori, 1964; Piaget, 1950). Similarly, language learning in children requires them to be exposed to specific language use (NAEYC, 1986). The environment of acquiring a language is an important component that should be carefully noted to assist children in developing communication skills. It includes a range of situations in which language is practised by listening and speaking in a variety of ways. A number of language development theories have been suggested over the last several decades. Piaget (1959) stated that children's language development is due to the interaction with the environment. He also asserted a cognitive theory that children must first comprehend a concept before learning the language which represents the notion. Brown (1973), as one of the well-known children's language development theorists, believed that there is a 'storage bin' inside each child where they store the words, phrases, and sentences they have heard. Later, they reach into the storage box and construct a proper statement using the words they had previously heard. Nevertheless, Brown (1973) argued that such exposure to language is not enough for children to acquire the language systematically. They also need a thorough comprehension of grammatical rules. He then proposed five phases of language development for children based on their language progress (Brown, 1973). He calculated the number of morphemes children may generate using the mean length of utterance (MLU).[1] To characterise each stage, the MLU rises with age. The data is presented in Table 1 below. In Stage I, children use 'telegraphic speech' using just content words like 'ball go' (Padilla & Liebman, 1975:35).[2] In Stage II, they began to employ additional phases to clarify the meaning, such as 'that, without' (ibid.). Stage III includes the use of the negative, interrogative, and imperative phrases (e.g., what, not) (ibid.). At Stage IV, children's simple expressions mirror those of adults. And Stage V incorporates the utilisation of coordination with conjunction words such as 'but' and, 'and'.

Stage	MLU	Age (month)	Expression
I	1.0 – 2.0	12 –26	▪ Telegraphic speech (two- or three-morpheme utterances)
II	2.0 – 2.5	27 – 30	▪ Use of grammatical morphemes
III	2.5 – 3.0	31 – 34	▪ Utterance length increases ▪ Use of interrogative, negation
IV	3.0 – 3.75	35 – 40	▪ Use of indirect of wh- questions
V	3.75 – 4.5	41 – 46	▪ Generate coordinate sentences ▪ Use of 'and' and 'but' as main conjunctions

Table 1: Compiled stage with matching MLU from Roger Brown's *A First Language: The Early Steps* (1973)

[1] Utterance: In line with Bloomfield (1926:154), utterance is frequently defined as 'an act of speech' (Crookes, 1990)
[2] Telegraphic speech: used by children who are at a two- or three-words stage and only use content words instead of function words such as articles (e.g., a, the), prepositions (e.g., with, in), modal verbs (e.g., can, is), and tense morphemes (e.g., -s, -ing) (Brown, 1973; Gabig, 2013; Eilers, 1975).

Chomsky's (1986) theory challenges Brown's, claiming that all human languages are based on certain underlying, innate universal principles (Lightbown & Spada, 2013:42). Children are born with an aptitude to learn all languages, which allows them to construct new sentences whenever they speak. He also believed that all languages have similar rules and structures (referred to as a 'Universal Grammar') and these grammatical principles were naturally developed in each child. Regardless of their surrounding learning environment, children's innate capacity allows them to become proficient language users. Chomsky termed 'language acquisition device' (LAD) to describe such development of innate knowledge (Brown, 2000:24). The LAD incorporates universal principles of all world languages and aids in keeping the child on track, not getting bewildered by the language's complicated rules.

In brief, all children require certain language exposure. The communication environment is a powerful predictor of early language development (Roulstone et al., 2011). Children's performance is directly influenced by their language environment (ibid.). In line with NAEYC (1986), becoming a skilled language user involves more than simply constructing semantic and syntax into sentences, but the skill of communicating with people.

2.3 Oral Language Skills

Oral language has been defined by several linguists as a sophisticated system that connects sounds to meaning and is usually composed of three key components: phonological, semantic, and syntactic (Lindfors, 1987). It has a significant influence on children's language development. It provides a foundation for children's reading and writing abilities when they begin school (Simmons & Kameenui, 1998; Keene & Zimmermann, 2007:40). Oral language is a complex system that needs both receptive (listening) and expressive (speaking) skills (Hulme et al., 2020). As mentioned before, language is a product of interactive processes involving phonological, semantic, and grammatical components (ibid.). Tomblin and Zhang (2006) found that a unidimensional language factor among six-year-old children was the source of several assessments measuring lexical and grammatical skills. West et al. (2021) presented more evidence to confirm that language is a unitary factor. Oral language skills include children's expressive and receptive vocabulary, sentence repetition and recalling, listening comprehension, and narrative retelling. However, narrative retelling is separated from other oral language skills to investigate whether it can be a valid metric to assess oral language proficiency.

Nevertheless, there is a lack of research examining the correlation between narrative retelling and oral language skill, and its potential impact on children's language development.

Several studies addressed the importance of developing speaking and reading skills as they are correlated with success in school. Regarding the study of Hill (2009), oral language (i.e., phonology) and reading are highly correlated, and they are predictive of children's success in reading achievement, thus, they are equally important in children's language development. An increasing amount of research indicates the need for oral involvement in classrooms to improve learning (Wilson et al., 2016). Cowell and Entwisle (1971) also noticed that the link between oral language development and reading has become more obvious. It has also been highlighted that oral language is a requirement for learning to read, as well as the basis for subsequent reading growth and development. Few studies examined the relationship between oral language and reading achievement as well, despite the fact that there is substantial heterogeneity in oral language between social and ethnic groups (ibid.:132).

Research has also found that language is a crucial component of literacy, particularly in the development of reading skills (Snow et al., 1995; Reese et al., 2010). According to Gambrell, Koskinen, and Kapinus (1991), oral language and reading proficiency are significantly correlated. These findings have been confirmed by Roth, Speece, and Cooper (2002). As expected, children's phonological skills predict single word reading in the first and second grades. The results that semantic skills predicted reading comprehension also show that the significance of various oral language abilities to early reading changes depending on language domain (ibid). Gambrell (2004) also researched the correlation between the two. The study found that other factors, aside from phonological awareness, for example, lexical knowledge, are also significant predictors of beginning reading for reading comprehension. She suggests that, in addition to phonemic awareness, early literacy education should include vocabulary development. Thus, reading comprehension relies heavily on oral language and vocabulary.

2.4 Narration and Narrative Retelling

2.4.1 Overview of Narrative

One of the leading and direct approaches applied in the classroom to train and improve children's oral language skills is narration. To better understand it, the definition of the term 'narration' should be investigated first. Narration refers to a story describing actual or imagined events that are made up of words concerning topics, settings, characters, activities, emotions, and outcomes in temporal order (Gillam & Pearson, 2004; Spencer & Slocum, 2010:179). It is also considered a mode of communication, a verbal account of previous events in chronological order (Hoff, 2013). Other researchers defined narration as a description of past events preserved in memory, produced by cognition, selected by the teller before delivering to the audiences, and organised in an anticipated structure (Graesser, Golding & Long, 1996). It is clear from these definitions that narrative has both linguistic and cognitive characteristics. Narratives are representations of past events in sequential order, whose production and interpretation are firmly based on human cognitive processes.

Narrative analysis is a valid and highly effective therapeutic and research technique since oral narratives provide a source of information about children's lexical knowledge, grammatical frameworks, and narrative structure (Botting, 2002:3; Cleave et al., 2010). What is more, oral narratives are typically measured within individual utterances (Heilmann et al., 2010:154), such as microstructural metrics, for example, analysing children's morphological and syntactic skills with MLU (Parker & Brorson, 2005:365; Millier, 1981; Brown, 1973), sentence complexity (e.g., Petersen, Gillam & Gillam, 2008; Nippold et al., 2005), and lexical diversity (Owen & Leonard, 2002; Klee, 1992). As a result, narrative analysis is an economical and sufficient technique for researchers and therapists. It is recognised that narratives require the integration of linguistic, cognitive, and social abilities (Kunnari et al., 2016).

2.4.2 The Importance of Narrative

The capacity to tell a narrative is essential for school-aged children; it has been related to improved social (Davidson et al., 2017) and academic results (Griffin et al., 2004). The contributions of narratives in both the social and academic worlds of childhood have been widely established. More narrative-related everyday conversations are initiated by children (McArthur, Adamson & Deckner, 2005; Preece, 1987). Narratives also help to maintain friendships and fit in with peer groups

(Davidson et al., 2017). The narrative is treated as a form of transformation from non-verbal to verbal. Nelson (2006) illustrated the use of this approach in a study of a two-year-old's bedtime dialogues and monologues. Furthermore, narrative competence has been shown to predict academic outcomes like reading comprehension (Fuchs et al., 2001; Griffin et al., 2004), vocabulary, and participation in academic remediation up to seven years later (Fazio et al., 1996).

2.4.3 Introduction of Narrative Retelling

A narrative is usually elicited in two ways: either with a verbal model (narrative retelling) or without (narrative generation) (Tsimpli et al., 2014); in either case, the narration is typically elicited with non-linguistic support such as a picture sequence. Both of approaches develop children's oral language skills, however, this chapter will mainly focus on the analysis and discussion of the former elicitation way for the purpose of the current study. Oral narrative retelling is a sophisticated linguistic and cognitive language assessment that requires children to comprehend a narrative that has been read aloud and to organise a coherent and comprehensive retelling in their own words (Lucero, 2018; Hipfner-Boucher et al., 2015; Gutiérrez-Clellen, 2002; Morrow, 1986). It differs from a narrative generation or story summary as details in the narrative, such as events, characters, and specific phrases need to be restated in order. Thus, narrative retelling as a clinical device helps children to organise, synthesise, and process the information they have just heard (Beers, 2003). What is more, Morrow (1986:135) illustrated that narrative retelling provides an opportunity for children to be actively involved in a reading experience delivered by oral language. Thus, narrative retelling can be an effective educational approach for improving children's dictation of original narrative as well as oral language complexity.

An examination of children's oral narratives gives insight into how children simultaneously integrate multiple language systems (Miller et al., 2006:31). When compared to spontaneous narrative exercises, narrative retelling imposes additional demands on language processing, such as working memory, the temporary storage of information, and short-term memory (Pauls & Archibald, 2021:2; Gutiérrez-Clellen, 2002). Narrative retelling can provide young children with a decent measure of language competency regarding the processing needs of memory recall and attention to details (ibid:186).

2.4.4 Microstructure and Macrostructure

Typically, narrative retelling is examined at two levels: microstructure and macrostructure (Gagarina et al.,2016; Lucero, 2018:249). Microstructural measurements focus on the linguistic components of a retelling, such as productivity (Hao et al., 2018), lexical diversity (ibid.), syntactic complexity (Miller, 1981; Nippold et al., 2005; Scott & Stokes, 1995), and clarity of cohesion (Liles, 1985; Pauls & Archibald, 2021). Moreover, the type-token ratio is usually applied to measure children's lexical diversity, provided that each child speaks the same number of utterances (Templin, 1957). Regarding microstructure, productivity describes the length and diversity of the words or sentences, which is often assessed by the total number of words (TNW), and the total number of different words (NDW) (Hao et al., 2018:346; Muñoz et al., 2003). TNW is calculated by counting how many words are used in the proposed sample, whereas NDW is found by counting the overall number of novel words used in the sample. Coherence and cohesion are usually measured by conjunctions, including both coordinating conjunctions (e.g., 'and', 'but'), and subordinating conjunctions (e.g., 'unless', 'if'). Thus, a child's narrative cohesion can be assessed by calculating the total number of conjunctions and the number of the novel ones (Heilmann et al., 2010).

Another important aspect of narrative microstructure is syntactic complexity, which refers to the complexity of the sentence structure in the narrative (Liles et al., 1995; Nippold et al., 2005). Syntactic complexity is generally measured by the MLU. MLU is a quantitative measure of morphosyntactic complexity that has been extensively researched and used to predict variations in language skills (Ranti, 2015; Miller et al., 1980). It is regarded as one of the most robust indices of children's language development (Rice et al., 2010). It displays the number of morphemes or words used by each child per utterance. The MLU(w) is calculated by dividing the total number of words by the number of utterances, whereas MLU(m) is the ratio of the total number of morphemes to utterances (Parker & Brorson, 2005). Nice (1925) demonstrated the predictable patterns of children's language development with the term 'mean length of response' (MLR), which is the term later replaced by Brown (1973) with MLU(w) (Parker & Brorson, 2005:366). Thus, MLU(w) has long been used to measure narrative length. According to the findings in Oosthuizen and Southwood (2009) and Ranti (2015), MLU(w) and MLU(m) are almost perfectly correlated. In additional to English, MLU(w) is highly correlated with MLU(m) in other languages, such as Irish (Hickey, 1991) and Dutch (Arlman-Rupp et al., 1976). As a result, MLU(w) has been widely incorporated and can be a reliable alternative to MLU(m) several issues about calculating morphemes were found by previous researchers (Ezeizabarrena & Garcia Fernandez, 2018). For instance, different languages vary in inflexion, and there are arbitrary judgements about morpheme production (Hickey, 1991; Arlman-Rupp, 1976). Overall, MLU(w) can be used as efficiently as MLU(m) to assess children's narrative abilities and gross language development (Parker & Brorson, 2005:372). It makes the measurement faster, simpler, more reliable, and less arbitrary (ibid.).

With respect to the calculation of MLU, longer utterances should indicate higher linguistic skills. This finding was also proved in Kless et al (1989) where the authors investigated the rate of MLU improvement relying on age range in three different studies. Surprisingly, the results were quite similar: MLU grows linearly up to 36 months of age and then continues to rise until 48 months of age. Children at 24 months have an MLU score of 2, while children at 36 months achieved a score of 3, and children at 48 months reached up to 4. Thus, MLU varies among different ages. In addition, many studies have also investigated gender differences in MLU as males and females frequently react differently to linguistic stimuli, as is widely documented (Fagot & Kavanagh, 1993). However, there were no gender differences discovered. The MLU values obtained by males and females did not vary substantially. Brown (1973) even recognised that MLU is not a perfect metric even though it is important to understand that as utterances lengthen, so do MLU (Ranti, 2015). However, it is not the only measurement of a child's narrative retelling ability as a child can repeat the same word multiple times in one utterance. (S)he achieved a high MLU but with less complexity and no diversity.

The second level of narrative retelling is macrostructure, called 'story grammar' in Pails and Archibald (2021), which refers to the overall organisation of the narrative and how the narrative is structured (van Dijk & Kintsch, 1983). It is also assessed with two main perspectives: macrostructural productivity (i.e., story grammar) and macrostructural complexity (i.e., episodic complexity) (Gillam et al., 2012). Story grammar refers to the completeness of the narrative, including the total number of the settings, topics, scenarios, characters, internal states, and problem resolutions the child has mentioned (Méndez et al., 2018; Diez-Itz et al., 2018). Episodic complexity measures the sequential order of the scenes, episodes, events, characters (Diez-Itz et al., 2018). As a narrative can contain more than one episode, some researchers have applied episodic complexity as a marker to indicate its narrative macrostructure. Gazelle and Stockman (2003) also discovered that when prompts were provided, children's episodic structures became richer; however, the narrative details were less developed with less productivity, lexical diversity, and syntactic complexity.

Children's narrative performance tends to be measured at both the lexical and the grammatical level (i.e., microstructural) and the story level (i.e., macrostructural) (Heilmann et al., 2010). But the focus and choice of these two levels still depend on the research needs. There are some studies focusing on the effect and relationship of these two levels of narratives. For example, Heilmann et al. (2010) used hierarchical regression analysis to explore the correlation between microstructural and macrostructural measurements of narratives among five- and seven-year-old typically developing children. They found that there is a significant relationship between narrative macrostructure and lexical diversity. They infer from the data that, until they are skilled in employing sophisticated syntax, children frequently rely on their lexical abilities to organise their narratives (ibid.:161).

2.5 The Characteristics of (Narrative) Retelling

2.5.1 An Assessment Tool for Reading Comprehension

Retelling, as an assessment strategy, indicates the extent of children's understanding and memory of the story. It shows their ability to organise and synthesise, and the growth of lexical and oral language development (Morrow, 1996). In addition, retelling can also indicate children's understanding of narrative structure and their ability to organise and integrate the information they have received.

As illustrated in previous studies, retelling and reading comprehension are correlated moderately (Reed & Vaughn, 2012; Morrow, 1984, 1985, 1986; Koskinen et al., 1988). The results in Gambrell, Koskinen, and Kapinus (1991) also demonstrated that retelling significantly improves student's free- and cued-recall reading performances. As a result, some researchers suggest using retelling as an assessment technique in schools to predict children's reading skills (Spencer & Slocum, 2010). Furthermore, there are no age restrictions in retelling assessment. It is developmentally suitable for every age group, even those who are just acquiring a language. Regarding Reese et al (2010), retelling is the most accurate and reliable device to predict a child's or person's reading comprehension compared with other oral language assessments. Nevertheless, Gambrell, Koskinen and Kapinus (1991) addressed that retelling can be problematic without a prior instruction from the class teacher. As it measures both oral and understanding skills at the same time, retelling needs to be instructed properly by teachers for assessment, and it is suggested to provide children with instructed procedures before the assessment. Overall, all findings implied that retelling can be used as a diagnostic and reading comprehension evaluation technique. Teachers can evaluate a child's ability to remember details, structure sentences, and sequence scenarios through delivering a retelling test.

2.5.2 A Strategy for Social Interaction

Retelling provides an interactive channel of communication for students (John, Lui & Tannock, 2003). Compared with traditional teacher-led reading sessions, children combined listening, speaking, reading, and even writing, and transferred them into their own words, when they were asked to retell a story. Thus, retelling is more participatory, interactive, and engaging. Moreover, throughout the retelling phase, children's communication channels were opened, so they could interact socially and collaboratively with their peers and teachers. This interaction directly boosts the development of reading comprehension and oral language skills.

2.6 Overview of Retelling Research

Retelling, as an effective and time-efficient strategy to boost both oral language and reading comprehension skills, has been explored in many studies. Retelling requires a child to listen to or read a story, and then organise the information to produce a personal rendition of it. Retelling is moderately correlated with reading comprehension, as children tend to gain a better sense of narrative structure and higher oral language proficiency after retelling sessions (Reed & Vaughn, 2012; Morrow, 1984, 1985, 1986). Therefore, retelling can be a useful alternative to the traditional comprehension practises in the class, particularly for those who are having trouble with silent reading (Koskinen et al., 1988). In addition, retelling entails a process of reorganisation and reconstruction. Students must read or listen to the story and then transfer it into their own words. Thus, retelling denotes an absorption or assimilation by the students, and it further demonstrates their comprehension, which has been termed as 'integration of information' by Morrow et al. (1986).

As discussed in previous sections, a child is actively involved in retelling tasks using their oral language, memory recall, and reading or listening comprehension skills. It is a valid strategy that helps children become more aware of the substance of a narrative and displays how deep their understandings are towards it. Therefore, it is a basic but efficient technique for assessing children's comprehension of a narrative once a teacher has appropriately modelled and instructed it for them.

2.7 Significance of (Narrative) Retelling

According to previous research, retelling, narrative retelling in particular, substantially enhances children's narrative comprehension, memory recall, oral language complexity, and concepts of story structure (Gambrell, Koskinen & Kapinus, 1991; Morrow, 1984, 1985, 1986). These improvements might be caused by the speaker's active participation and interaction with the listener (Amati & Ziegler, 1973). With regards to Brown's (1975) study, children's narrative comprehension improves significantly when they are actively involved in the reconstruction of a story.[3] Retelling is also treated as an assessment tool for reading comprehension and a strategy for social interaction. It requires the organisation of information and holistic reconstruction of knowledge. In short, it is an effective assessment metric that indicates a person's ability to summarise information they received and transfer it into an intelligible text in order.

2.8 Measurements of Narrative Retelling

There are three main approaches to scoring narrative retellings: the narrative scoring scheme (NSS) (Heilmann, Miller, & Dunaway, 2010), the Strong Narrative Assessment Procedure (SNAP) (Strong, 1998), and multilingual assessment instrument for narratives (MAIN) (Gagarina et al., 2012). NSS attempts to provide a simple scoring system. It combines several components of the narrative retelling process into a single grading rubric and offers a general assessment of the child's narrative aptitude. There are seven categories in the scoring scheme, including introduction, character development, mental states, referencing, conflict resolution, cohesion, and conclusion (Heilmann, Miller &

[3] Reconstruction is a process where children recall each narrative event and organise pictures of the story in sequential order.

Dunaway, 2010:165-166). Each characteristic will be awarded a scaled score varying from 0 to 5 (0 = no response or undefined points; 5 = perfectly addressed). SNAP is an individually administered criterion-based approach for assessing multiple narrative abilities (Han, 2018). It is appropriate for children from kindergarten to eighth grade (ibid.). SNAP assessment contains four audiotaped stories in which children recollect key elements of narrative, summarise the stories, and answer questions that measure their comprehension. The purpose of this assessment is to provide a therapeutic approach that can assist children with social interactions. It is scored by scaling the complexity of settings, characters, actions, temporal markers based on the Index of Narrative Complexity (INC) scoring from 0 to 3 (0 = no information provided; 3 = include more than one event or character with a specific name) (Petersen, Gillam, & Gillam, 2008). In contrast with NSS, SNAP is time-consuming, and it does not always focus on the most important aspects of narratives, nor does it provide a composite score for clinical treatment (ibid.). The final measurement, MAIN, is widely used by researchers as it has been translated into dozens of languages for analysing different narrative elicitation modes: generation, retelling, and telling after a model sample (Gagarina et al., 2010). It is quite complex to score the MAIN test. It includes a maximum of 17 points for Section 1 in production and a maximum of 10 points in Section 2 for comprehension (ibid.:66). Overall, the selection of the scoring metrics depends on the aim of the specific research.

2.9 Demographics and Narrative Retelling

Demographic factors (e.g.. age and gender) sometimes play a role in predicting certain variables. Despite a dearth of empirical literature specifying gender differences in narrative retelling, past research illustrated that narrative retelling differs by gender in certain categories. Crow et al. (1998) indicated that girls are more talkative than boys, so girls are expected to produce longer narratives using more words (i.e., TNW is high). Moreover, since girls are more emotional than boys (Lips, 1993), so girls would focus more on the details of the emotional components of the narrative compared with boys. Apart from gender, age has been identified as a major demographic predictor of narrative retelling. For example, it is found that older children tend to recall more events when they were asked to retell a story they have just heard compared with younger ones (Griffith, Ripich & Dastoli, 1986). They also employ a more sophisticated self-monitoring mechanism than younger children for self-correction in narrative retelling tasks (ibid,:554). With regards to the study of John, Lui, and Tannock (2003), both age and gender differences affect the results of children's narrative retelling and comprehension. It shows that girls outperformed boys in comprehension, and older children receive higher scores than younger ones.

2.10 Gaps in the Existing Literature

As previously indicated, existing research has deciphered the impact of both oral language and narrative retelling skills on children's language development. Nevertheless, none of the studies investigates the correlation between narrative retelling and oral language skills, and whether demographic factors predict the performance of narrative retelling. The foremost motivation for conducting the current study was a lack of adequate attention in this area. This research bridges the gap and contributes to helping children thoroughly develop their oral language and comprehension.

2.11 Research Questions and Hypotheses

This study covers the following research questions to assess the effectiveness of oral language skills in enhancing children's narrative retelling.

1) What is the relationship between children's narrative retelling and the other aspects of oral language skills?
2) To what extent do oral language skills predict children's narrative retelling, after controlling for demographic measures? What demographic measures contribute to the prediction of children's performance in the narrative retelling?

Research Question 1: What is the relationship between children's narrative retelling and the other aspects of oral language skills?

In the current study, I intend to test the relationship between narrative retelling (a subgroup of skills belonging to retelling) and other aspects of oral language skills (multiple oral language skills). It is hypothesised that these two variables are significantly correlated. The higher the oral language skills, the better their narrative retelling. The rationale behind this hypothesis was supported by previous research on retelling and language development. It suggests that retelling is an accurate device to measure children's reading comprehension and oral language skills (Reed & Vaughn, 2012; Gambrell, Koskinen & Kapinus, 1991; Morrow, 1984, 1985, 1986).

Research Question 2: To what extent do oral language skills predict children's narrative retelling, after controlling for demographic measures? What demographic measures contribute to the prediction of children's performance in the narrative retelling?

The second research question, as a follow-up question, focuses on the predictions of demographic factors selected in the sample (e.g., age, gender, EAL, and SchoolNo.). After a review and evaluation of the existing research, it can be hypothesised that other aspects of oral language skills largely predict children's narrative retelling after controlling for demographic factors. Moreover, age, gender and EAL all play a role in predicting the performance of narrative retelling. The rationale for hypothesis 2 was supported by the research in Crow et al. (1998), Lips (1993), and John, Lui, and Tannock (2003) as presented in Section 2.9.

2.12 Summary

This chapter examined the current research regarding language development, oral language skills, mean length of utterance (MLU(w) and MLU(m)), and (narrative) retelling. Many studies found that both oral language and retelling are correlated with reading comprehension. Other literature argued that both skills could be applied in designing therapeutic treatments for children with reading difficulties. Gaps in the studies were then exposed after reviewing the available literature. Finally, the research questions and hypotheses were presented to bridge the gap relying on the previous research implications and limitations. The sample and methodological approaches will be described in the next chapter.

Chapter 3: Methodology

3.1 Introduction

This chapter will discuss the methodological approaches to answer the two main research questions introduced in Section 2.11. The choice of the research design will be explained in Section 3.2. Next, Section 3.3 depicts the selected existing dataset, including the description of the dataset, the sample, and the new usage of the data. Then, the definition of secondary data analysis, as well as the characteristics of this approach will be introduced in Section 3.4. Following that, ethical considerations are discussed in Section 3.5, followed by an introduction of the overall measuring instruments in the present study with some background on an app named ATLAS (Automated Test of Language AbilitieS), other language tests, and the narrative retelling of bus story tests in Section 3.6. Section 3.7 presents a detailed description of the research procedure: the scoring system selection and data analysis approach. There is a critical discussion on the validity and reliability of the choices on selecting these measuring instruments in Section 3.8. Section 3.9 includes the introduction of the selected analysing methods. Finally, Section 3.10 shows a concise summary of the analytical method used in the current study.

3.2 Research Design

The current research is a fully quantitative-based secondary data analysis aiming to explore the association between narrative retelling and other aspects of language skills among pre-schoolers. Quantitative data were adopted from an existing NELI dataset (see Section 3.3.2 for sample descriptions). The rationale of the choice of secondary data analysis is mainly due to the outbreak of COVID-19, since conducting fieldwork in nursery schools was not possible. Exploratory factor analysis was carried out to extract the independent and dependent variables from several observed variables. Before delving into the instrument and data analysis, the following part summarises the data collection, as well as the sample population.

3.3 Sampling

3.3.1 Dataset

The data applied in the current research were mainly collected through the Nuffield Early Language Intervention (NELI) programme. NELI is a 20-week oral language intervention that has been tested in research funded by the Education Endowment Foundation and the Nuffield Foundation. It is aimed towards children aged three to five years old in Reception (the first year of formal education in the UK) who have inadequate spoken language abilities (Fricke et al., 2018). The programme targets children's listening comprehension, vocabulary, and narrative skills, with extra targeted support and assistance for individuals with limited language skills. It mixes small groups with one-on-one sessions provided by qualified teachers or teaching assistants, intending to enrich the oral language of all children in nursery, particularly those who have poorer language skills (Gillian et al., 2021). However, the intervention was stopped due to the outbreak of COVID-19 in 2020, so it became a 10-week programme without an intervention. The datasets were barely collected within the 10 weeks, including a series of vocabulary and narrative tests.

The dataset included 1534 children from the age of 25 months to 58 months. Sixty-two schools were enrolled in the programme, encompassing around 70 classrooms. All children completed four oral language tests (further introduced in Section 3.6) through ATLAS app. The schools themselves carried out all tests, and scores from these tests were combined to provide an estimate of a child's overall language ability. The individual in-depth assessments were only undertaken by a selected group of about 600 children (missing data will be tested and analysed in Section 4.2 with a statistic output). The chosen group was composed of two types of performers: the last six per class who achieved the lowest scores and the additional four to six children who were randomly selected to represent the whole class enrichment.

3.3.2 Description of Sample

In the current study, I only focused on children's narrative retelling transcripts. This task taps into children's story retelling ability. The sample consisted of 600 children (280 female, 320 male) from

62 different schools (see Appendix B specifying the number of children in each school), after deleting some missing data (see Section 4.3 for missing data treatment). Their ages ranged from 37 months to 58 months (Mean = 42.78; SD = 3.606). The average age was 42.80 (SD = 3.63) for male children and 42.76 (SD = 3.58) for female children. Thus, 600 transcripts of children's retelling were analysed, excluding those who did not attend narrative retelling assessment or other language tests (see Table 2).

Variable	**Value**	**N**	**SD**	**Min**	**Max**
Gender	M	320 (53.3%)	3.63	37	57
	F	280 (46.7%)	3.58	37	58
EAL	EAL	280 (46.7%)	3.49	37	57
	Non-EAL	320 (53.3%)	3.71	37	58
All		600	3.61	37	58

Table 2: Demographics of the children selected in the sample

3.3.3 New Use of the Data

This analysis uses a range of novel ways of using NELI data. First, exploratory factor analysis was performed to extract the independent and dependent variables: oral language skills and narrative retelling. The procedures of the factor extraction will be discussed in Section 3.9.1, and the extracted results will be presented in Section 4.2. Second, the researcher non-standardised scored the measurements of narrative retelling mainly following the rules of thumb of Brown (1973), which previous researchers in NELI studies did not use. Besides, the major goal is to investigate the relationship between two extracted variables ('narrative retelling' and 'other aspects of oral language skills'). Most of the research focuses on narrative retelling and comprehension where narrative retelling is an independent variable predicting children's comprehension skills. However, there is a lack of research focusing on: exploring the correlation between narrative retelling and the other aspects of language skills, as well as the prediction from demographic variables. Thus, the research will provide direct insight into the role of narrative retelling skills for children in nursery schools in the UK and help school teachers and policymakers re-design the course structure of narrative retelling from vocabulary-based exercises. Theoretically, they emphasise the importance of language as a basis for narrative retelling within the NELI programme.

3.4 Analytic Approach: Secondary Data Analysis

There are a variety of research methodologies on the topic of children's narrative (re)telling. The employment of diverse methods and procedures in a field of study can improve research quality by investigating similar topics from different perspectives with distinct empirical approaches (Boeren, 2018; Robson, 2011). Apart from the primary method of collecting data in person, researchers could use existing datasets undertaking quantitative research, which procedure is technically referred to as "secondary data analysis" (Boeren, 2018:72-73). As previously stated, this paper aims to tackle new research topics using data obtained from an existing dataset. The features, advantages, and disadvantages of secondary data analysis will be explained briefly in this section.

3.4.1 Definition of Secondary Data Analysis

The term "secondary data analysis" refers to the examination of data that has already been gathered by others (Donnellan & Lucas, 2013). It is a social science research method that applies "creative analytical techniques to data that have been amassed by others" (Kiecolt & Nathan, 1985:10). Unlike primary data analysis, in which researchers construct a study to answer their own research questions, secondary data analysis allows researchers to leverage existing resources to answer new questions (Tripathy, 2013; Szabo & Strang, 1997). Compared with primary data sets, secondary data is typically collected through a larger sample of people representing the population (Fielding & Fielding, 2003). While the majority of the datasets mentioned are longitudinal, there are also numerous cross-sectional datasets (Vartanian, 2010). However, no research method is perfect. There are still many advantages and disadvantages when taking a secondary data analysis.

3.4.2 Advantages of Secondary Data Analysis

The major advantages are that secondary data analysis is less time-consuming and less costly than primary data analysis (Donnellan & Lucas, 2013). The data in the current paper was designed and collected beforehand and sponsored by the university research team. Therefore, it saves time and resources to use the data, especially during the pandemic. Second, it allows academicians to examine research topics with large-scale high-quality datasets, with potentially better quality than the researcher is capable of administering (Donnellan & Lucas, 2013; Smith, 2008). It eliminates research repetition by conducting a thorough examination of the existing dataset. Transparency and diverse findings are two additional benefits of secondary data analysis. To be more explicit, researchers are compelled to disclose their results transparently as the database is publicly accessible.

3.4.3 Disadvantages of Secondary Data Analysis

On the other hand, using secondary data presents some struggles and limitations. The biggest drawback is certain ethical issues. It centres on the possible harm to individual participants and the issue of return for consent (Tripathy, 2013). If the data is adequately coded, the researcher may not have access to the codes. Thus, the data needs to be fully anonymous before reusing. Another disadvantage is the lack of hands-on engagement in data gathering (Smith, 2008). It does not allow the secondary researcher to get first-hand knowledge into the lives of their subjects. Thus, using secondary data introduces a lack of control as the investigator was not involved in the research design and on-site data collection, which may cause some issues, such as mismatching concepts and variables (Vartanian, 2010).

3.5 Ethical Considerations

The ethical approval was obtained in the original NELI programme as secondary analysis research. The data had previously been anonymised using unique index codes and numbers for each child. A collection of codes that matched the names and identities of children were safely kept by previous researchers. All information was kept privately and solely used for specific research purposes. In addition, no school, child, or private information could be recognised under the General Data Protection Regulation (GDPR). The data will be retained for the duration of the MSc programme and then deleted. Before granting access to data to both the researcher and the secondary scorer for the current analysis, a declaration of the appropriate use of data was made.

3.6 Measuring Instruments

The current study employed the following measures, and the data was collected through the NELI Individual In-depth Assessment Record Form (see Appendix A).

3.6.1 ATLAS (= LanguageScreen)

ATLAS, also known as LanguageScreen test, is an assessment digitally measuring a child's speaking and listening language abilities. It is a short but reliable oral language assessment developed by a research team at the University of Oxford. As a tablet-based app, it is meant to be administered by school teachers or teaching assistants. It is composed of four subtests: expressive vocabulary (EV), receptive vocabulary (RV), sentence repetition (SR), and listening comprehension (LC). The whole test may take less than 10 minutes, and it is automatically scored through the website. The scores from these subtests were combined to give an estimate of a child's overall language ability. During the test, children can only see the screen when conducting EV and RV tests. For the other two tasks, the examiner needs to ensure the child does not see the tablet. Once the PLAY button is pressed, the app will read the instructions to the child. When the child is ready to begin the test, the START ACTIVITY button is pressed. Each item equals 1 point. The overall score of the four tests is 77 with 1 point for each question.

EV tests a child's ability to name pictures of objects. There are 24 items (= 24 points). The pictures will be presented to the child accordingly, and the child should name them in order. If they encounter any problems, the examiner can give prompts by asking "What is it? or What is this called?". Then the examiner scores the child's answer as either correct or incorrect by clicking the buttons on the top of the screen. LC assesses children's ability to understand short stories and their ability to make inferences from what they have heard. There are three stories within the LC test and each story can only be played once. Children listen to three short stories, each of which is followed by four to six comprehension questions. The first story includes four questions and the other two stories have six questions each. The sum of the score for each story will constitute the child's final score (= 16 points).

RV assesses a child's ability to understand spoken words. There are 23 items (= 23 points) with an increased difficulty throughout the test. SR assesses the child's ability to remember and repeat sentences which tests their grammatical skills. There are 14 items (= 14 points), which become progressively more difficult.

3.6.2 Other Language Tests

Apart from ATLAS tests, additional language measurements were taken by children: CELF Preschool II^{UK} *Expressive Vocabulary* (picture naming) and *Recalling Sentences* (SR), and APT tests (Wiig, Secord & Semel, 2004; Fricke et al., 2013). The Clinical Evaluation of Language Fundamentals (CELF) is a comprehensive language assessment designed for children ages 3:0 to 6:0 to measure children's language development in vocabulary and grammar. Two CELF tests were included in the current dataset: CELF expressive vocabulary (CELF_ev) and CELF recalling sentences (CELF_rs). CELF_ev refers to the task of naming the pictures of people, objects, and actions. There are 20 items in total with 1 point for each question. Regarding CELF_rs, researchers need to score each item by comparing the response to the stimulus sentence by counting the number of errors in the response. The score varies from 0 to 2 (0 = OK response, 1 = 1 error, and 2 = two and more than two errors). There are 13 items in total with a score range from 0 to 2 for each question.

In contrast, the Renfrew Action Picture Test (RAPT) (Renfew, 1997) is a standardised assessment to test children's spoken language skills, usually scored for both information and grammatical parts (Bolderson et al., 2011). Ten culturally and linguistically relevant action pictures are utilised, a significant figure for retaining children's attention and motivation (Afreen et al., 2014). There are 10 pictures inside the kit, accompanied by 10 questions at the back of each (Appendix C attached with 10 APT cards). The test provides information (verbal formulation) and grammar (function words and word endings) scores. Both information and grammar scores have no exact maximum score as the scores vary in different pictures. However, the half-point is allowed for information scores.

Table 3 (below) summaries the scoring statistics in ATLAS, CELF, and RAPT.

Test		Number of questions	Maximum score for each question	Total score
ATLAS	EV	24	1	24
	TV	23	1	23
	SR	14	1	14
	LC	16	1	16
	Total	77	1	77
CELF	EV	20	1	20
	RS	13	2 (varying 0-2 points)	26
RAPT	Grammar	No exact maximum score, varying in different questions (*RAPT_info includes half-point)		
	Information			

Table 3: Summary of scoring statistics for variables in ATLAS, CELF, and RAPT

3.6.3 Narrative Retelling of the Bus Story

The narrative retelling of the bus story used in the current study is a type of non-standardised test, and has been previously used to test children aged between two and five.[4] The task is composed of a short story and three accompanying coloured pictures (see Appendix D.1). Children in the selected group were shown each of the three pictures while the examiner told the story. The examiner reads a scripted story to the child and then asks them to retell it to a little toy called NELI, using pictures as a guide without prompting words (see Appendix D.2 for original story transcripts). The task requires them to retell the story in as many details as possible to the original relying on the pictures as prompts (Botting, 2002). There are no practice trials, and all responses contribute to their final performance. Indirect prompts except for recalling parts can be provided with general questions to

[4] A non-standard assessment is one that does not rely on a standardised test, typically because the student being evaluated does not fit the normative sample for the test (Spear-Swerling, 2013).

encourage children to proceed in the narration, such as "what happened next?" (see Appendix D.3). The examiner audio- and hand-recorded the child's story verbatim during the 10-week NELI programme, even including grammatical mistakes they had made (e.g. "She goed" rather than "went"). No labelling or directions to the child's story was needed when transcribing (e.g. do not transcribe "Picture 1..."). The secondary data only include each child's retelling transcripts. In order to evaluate their narrative retelling ability, I selected five elements to give five scores to each child regarding narrative scoring rubrics from previous research. Moreover, a second scorer was recruited in the study for verifying inter-rater reliability (Section 3.8.2).

3.7 Procedure

3.7.1 Self-scoring

The scoring for bus stories is complicated, and errors commonly appear (Dockrell & Marshall, 2015; Haccoun, 2001). According to previous research, five scores were derived to indicate children's oral narrative retelling abilities in the current paper.[5] Children's narrative retelling was coded for productivity and complexity by five microstructural metrics:

- Productivity, measured by the total number of the words (TNW),

- Lexical diversity, measured by the total number of the different words (NDW),

- Syntactic complexity, measured by the mean length of utterance in words (MLU(w)), and

- Narrative cohesion, measured by the total number of the conjunction words, and

- The total number of the unique conjunction words.

The first three measurements were applied based on Fricke et al (2013:283) to measure children's narrative retelling skills. In light of Méndez et al. (2018), narrative microstructure entails using language at a words and sentence level. Therefore, it conveys meaning through the application of linguistic knowledge and is usually described by metrics like lexical variety (i.e. TNW, NDW, and MLU) (Castilla-Earls et al., 2015; Hipfner-Boucher et al., 2015; Méndez et al., 2018). Mean length

[5] All figures were accurate to two decimal places

of utterance (MLU) was a useful approach to measure children's syntactic development (Nippold et al, 2005). However, the calculation is debatable (Parker & Brorson, 2005). First, how utterances are segmented is essential to MLU calculation (Eisenberg et al., 2001), although there is no consensus on what defines an utterance. Second, a number of non-linguistic variables, including memory difficulties, external distractions, and tiredness, might also affect the production of an utterance (Oosthuizen & Southwood, 2009). Notwithstanding, MLU is one of the most suitable approaches to measure children's narrative performance despite several limitations. MLU(w) was selected in this study as it is much faster and simpler to implement in 600 samples. It is also as reliable as MLU(m) based on Parker and Brorson (2005:372) argument from Chapter 2. MLU(w) was then calculated by totalling the number of words for each utterance, adding them up, and dividing by the number of utterances (Parker & Brorson, 2005: 372). Traditional methods of calculating MLU based on Brown (Brown, 1973:54) were applied in scoring children's narrative retelling skills (see Appendix E for a list of rules). The researcher and a second scorer counted the remaining four measurements based on the same set of narrative transcripts. Responses were coded and calculated following the same scoring criteria by the second scorer allowing for analysis of errors. Table 4 presents a summary of the variable statistics in narrative retelling.

Test	Measurements	Score
Narrative retelling	TNW	Count the total number of the words
	NDW	Count the total number of the novel words
	MLU(w)	$\frac{\text{TNW}}{\text{Total\_utterances}}$
	Total_conjunction	Count the total number of the conjunctions
	Total_different conjunctions	Count the total number of the different conjunctions

Table 4: Summary of scoring statistics for variables in narrative retelling

3.7.2 Data Analysis

To address the research questions, two major analyses were carried out in SPSS software (version 27.0). Pearson's correlations were applied to look into the correlations between narrative retelling and other oral language skills. Following the correlations, a hierarchical linear regression analysis was used to answer the second research question by determining the relative effect of the other predictors (i.e. age, gender, EAL, and school number) on language abilities.

Researchers often examine the strength of relationships between variables using Pearson correlation or regression analysis to determine the degree to which they are related. Two analytical approaches were chosen to answer two research questions in the current study.

Answer to Research Question 1: A simple Pearson correlation analysis of children's narrative retelling and other aspects of language skills answered to this question.

Answer to Research Question 2: This question was answered by applying hierarchical linear modelling analysis using two models. The factors studies in the initial model included demographic parameters such as age, gender, EAL and SchoolNo. The key predictor, oral language skills, was included in the second model.

3.8 Validity and Reliability

Validity and reliability are essential psychometric properties in assessments, which impact the accuracy and generalisability of results. The validity and reliability notions are closely linked. However, they indicate different characteristics of the instrument. Based on the applicability of the SPSS software programme, its internal reliability, inter-rater reliability and construct validity were examined for the validity and reliability of the individual language assessments (ATLAS, CELF, APT).

3.8.1 Internal Reliability

Internal reliability (aka internal consistency) is a representative statistical metric that investigates whether multiple items meant to assess the same broad concept would inter-correlate with one another (Netemeyer et al., 2003). Cronbach's alpha (α) is used to measure the strength of the coefficient (Pisani, et al., 2015:9). It generally ranges from 0 to 1, with the closer the alpha is to 1, the stronger the internal reliability of items in the instrument being evaluated (George and Mallery, 2005:223). The coefficient of reliability will increase once the correlations between two items measuring that construct also increase. According to George and Mallery's (2005:231) rule of thumb, Cronbach's alpha (α) range and its strength of association are presented in Table 5.

[redacted]

Table 5: Cronbach alpha α coefficient range (Source: George and Mallery, 2005:231)

3.8.2 Inter-rater Reliability

Inter-rater reliability refers to the measurement of the same concept by two or more raters, which was recommended as one of the effective measures of reliability (Robson, 2011; McCauley and Swisher, 1984). It indicates the agreement in the ratings supplied by different raters and works to improve the scale.

All children's narrative-related scores were verified for inter-rater reliability to confirm that the researcher had consistently implemented the scoring system. After filling out the MOU research agreement with the researcher (see Appendix F), the second rater was asked to access the identical narrative transcripts and was aligned with the first rater's core guide based on Broan (1973).[6] Scores from the first and second rater were compared separately. Both raters' scores on TNW, NDW, MLU(w), Total_conj, Total_diff conj were entered into SPSS. The inter-rater reliability between the two raters will be measured with the Pearson correlation coefficient *r* (see Section 4.4 for data analysis).

3.9 Analysing Approaches

3.9.1 Exploratory Factor Analysis

Exploratory factor analysis (EFA) was utilised in the present study. It was a fundamental tool for summarising and interpreting the underlying data correlation (Young & Pearce, 2013). It is a set of techniques for examining how underlying components impact responses to a variety of circumstances (DeCoster, 1998; Matsunaga, 2010). Factor analysis is classified into two types: exploratory and confirmatory (Kim & Mueller, 1978). EFA aims to identify a collection of unobserved (i.e. latent) variables that rebuild the complexity of observed data in a useful manner. Confirmatory factor analysis (CFA) determines if a particular set of constructs influences responses in an anticipated manner (Mvududu & Sink, 2013:79; Thompson, 2004 & DeCoster, 1998). The pattern of relationships between the observed measurements is examined in factor analysis.

The focus of the current study was on EFA as the researcher has no assumptions regarding the number of common factors or which measured variables would be influenced by the same common factors (Fabrigar & Wegener, 2012). The major goal of an EFA is to reduce variables and ascertain the total number of common factors that influence a set of measurements (Yong & Pearce, 2013; DeCoster, 1998). Pallant (2010) identified two key considerations when evaluating whether a certain dataset for a sample is suitable for conducting factor analysis: the sample size and the strength of the association among the variable items. There are various perspectives on a fixed number of sample sizes; and numerous rules were addressed by Williams, Onsman and Brown (2010). Hair et al. (1995) recommended that sample sizes be greater than 100. According to Tabachnick and Fidell's (2007) rule of thumb, a factor analysis requires at least 300 samples. Comrey and Lee (2013) considered a sample size of 500 as a very good number to carry out a factor analysis. In the present study, a sample

[6] The second rater is a native English speaker studying at University of Oxford majoring in Classics

size of 600 is appropriate for factor analysis. The strength of the relationship among the variables will be demonstrated in the extraction method subsection.

Factor analysis in the present study was conducted following three main steps: reliability consistency analysis, assumptions underpinning principal axis factoring (PAF), and rotation choices. The degree to which a measure is unbiased in guaranteeing consistent measurement is determined by its reliability. The Cronbach's (1951) alpha (α for continuous variables) coefficient test was performed to determine the reliability of the instrument employed in the study. I computed Cronbach's (1951) alpha (α for continuous variables) coefficient for all the observed variables to evaluate its internal consistency. The results will be shown in Section 4.2. In SPSS, many different estimation metrics can be used with EFA models to extract the number of factors, with PAF being one of the most widely used methods (McNeish, 2017). PAF is particularly appropriate to determine probable latent constructs in the dataset and enables a more precise analysis of the item correlations. Scree plot (Cattell, 1966b) is the most intuitive and helpful way to determine the number of factors to retain in factor analysis (Field, 2018:801; D'agostino Sr. & Russell, 2005). The eigenvalue is another alternative to measure how many variations were explained by a factor (Field, 2018:790). Researchers generally follow Kaiser's criterion (eigenvalues > 1) when the variable count is fewer than 30; additionally, if the sample size exceeds 250, it will be even more accurate (ibid.:790). Factors are often rotated for better interpretation (Field, 2018:803; Yong & Pearce, 2013:84). Rotation helps to achieve an ideal basic structure in which each variable load is distributed over a few factors as feasible while maximising the number of high loadings on each variable (Rummel, 1988:145). There are two types of rotations: orthogonal and oblique (Corner, 2009). Regarding Gorsuch (1983:203-204), the factors in the analysis are assumed to be uncorrelated in orthogonal rotation (varimax, quartimax). In contrast, oblimin rotation (direct oblimin) assumes the expected correlations between variables (Field, 2013:793)**.** An unrotated extraction method that attempts to explain the most variation can best capture the nature of the data (Brody, 2018). Two separate EFAs will be conducted in the current dataset to extract two factorised variables ('narrative retelling' and 'oral language skills') for further analysis.

3.9.2 Correlation Analysis

In order to answer research question 1, the researcher computed correlational analysis between the derived narrative retelling scores from EFA and the oral language skills scores. The assumption checks (i.e. variable types, normality) were carried out in Section 4.7.1 before presenting the correlational findings. Pearson's correlation coefficient (*r*) was then analysed to determine the relationship between children's narrative and oral skills. Pearson's correlation is often known as a statistical method used to assess the strength of the correlation among the variables (Puth, Neuhäuser & Ruxton, 2014). The relationship strength can be measured through coefficient values, as can be seen in Table 6. However, it is well-known that these cut-off points are arbitrary and inconsistent in different research (Schober, Boer & Schwarte, 2018). A majority of researchers agree that a coefficient less than 0.1 suggests an insignificant correlation, while a coefficient larger than 0.9 shows a nearly perfect association. Values in between are quite debatable (ibid.).

[redacted]

Table 6: The magnitude of the effect of correlation relationship strength (Sabilla, Sarno & Triyana, 2019:82)

The effect size indicates the strength of the correlation and shows how strong/weak the relationship is (Sabilla, Sarno & Triyana, 2019). In Chapter 4, the direction of the correlation (positive or negative) and the size of the effect (See Table 7) will be specified.

[redacted]

Table 7: Effect size measure (Field, 2018:340)

3.9.3 Hierarchical Linear Regression

Regression analysis is a statistical approach to model and analyse a number of variables to investigate the correlation between independent and dependent variables. It explains how the dependent variable value varies when the independent variables are statistically changed (David, 2005). To assess the predictive values of the variables (i.e. demographic predictors, oral language skills) set inside the model, hierarchical linear regression (HLR) will be applied to answer the second research question. HLR is a statistical technique for investigating the relationship between a single outcome or criterion variable and a number of predictor variables (To & Mandracchia, 2019:2). More variables will be introduced to the model in "blocks" rather than all at once. It is typically used to measure if one or more variable(s) significantly improves a model's ability while statistically controlling for other variables (i.e. whether one variable can predict and affect the relationship between two other variables) (ibid:3). As there are a few demographic variables (such as age, gender, EAL, etc), HLR was considered to measure these potential predictive variables in this case. A two-stage HLR was carried out to examine the extent to which control variables can predict variation in children's narrative retelling. In the first block, age, gender, EAL, and school number were simultaneously added to the prediction model. The derived factor for oral language skill was then entered into the second block. Assumptions and results of the HLR will be presented in Chapter 4.

3.10 Summary

This chapter started with the description of the NELI dataset and demographic samples utilised for the investigation. Next, the analysis, secondary data analysis, and ethical issues were shown in this chapter. The measuring instruments for the analysis were reported, including both oral language tests and narrative retelling tests. Then, the reliability and validity of the instruments were discussed regarding the narrative scoring rubrics. Finally, we examined the statistical method, correlation, and regression modelling. In the next chapter, the results will be presented from the statistical analysis.

Chapter 4: Results and Analysis

4.1 Overview of the Chapter

The primary objective of this study was to explore the relationship between a non-standardised measurement of story retelling and oral language skills. In the present study, oral language skills were measured in two different ways: both a novel app-based ATLAS and traditional paper-based CELF and RAPT. The second goal was to investigate whether demographic factors contribute to the prediction of children's narrative retelling performance.

This chapter will address and present the extracted results from SPSS for the analyses of the selected dataset in the current study. These analyses were intended to help answer the specified research questions laid out in Section 2.11 and the aims articulated in Section 1.3. The preliminary analysis phase will comprise a detailed description of the statistic, missing data, and inter-rater reliability check. Normality tests were then presented to confirm that the sample was normally distributed, followed by an account of the KMO and Bartlett's Test for the narrative retelling test and other oral language tests. Next, two major research questions will be discussed in turn with an application of Pearson correlation and hierarchical linear regression analyses. Finally, there is a summary of the whole chapter.

4.2 Descriptive Statistics

Table 8 displays descriptive statistics for all measurement scores on narrative retelling (total number of the words, total number of the unique words, MLU in word, total number of the conjunctions, and total number of the different conjunctions) and oral language skills with four ATLAS subtests (EV, RV, SR, and LC), two CELF subtests (EV, RS), and APT (information, grammar). The descriptive statistic table includes the means, standard deviations, skewness, kurtosis, and Cronbach's alpha value. All variables satisfied the requirements for univariate normality and were acceptably normal for skewness and kurtosis values (Kline, 2015; Curran et al., 1996). Sekaran (2003) indicated that if the skewness and kurtosis values fall within a range of [-2, +2] and [-3, +3], respectively, the sample is acceptable and normally distributed.

	N valid	Mean	Std. Deviation	Skewness	Kurtosis	Cronbach's alpha
TNW	600	19.96	15.848	0.808	0.326	0.742
NDW	600	13.23	9.048	0.308	-0.720	
MLU(w)	600	4.06	2.611	0.635	0.406	
NT_Conj	600	1.55	2.192	1.630	2.813	
NT_DiffConj	600	0.82	1.051	1.261	0.965	
EVtotal	600	6.07	4.553	0.467	-0.599	0.926
RVtotal	600	10.06	4.146	0.114	-0.480	
SRtotal	600	3.72	3.584	0.826	-0.328	
LCtotal	600	2.33	3.158	1.530	1.748	
CELFev	600	8.49	6.265	0.627	-0.305	
CELFrs	600	5.84	6.800	1.478	1.988	
APT_info	600	14.982	8.719	-0.059	-0.974	
APT_gram	600	9.733	7.720	0.528	-0.501	

Table 8: Descriptive statistics of all measurements for narrative retelling and other aspects of oral language skills (n = 600)

Reliability Analysis

Table 9 shows Cronbach's alpha value for each variable.

Variables	N of items	Cronbach's Alpha
Other oral language skills	8	.926
Narrative retelling	5	.742

Table 9: Cronbach's alpha value of variables

Nunnally (1978), Field (2009), and Pallant (2007) defined high reliability as a minimum level of 0.70 for the scale of variables. Table 9 shows that the Cronbach's alpha value of the variables was greater than 0.70, indicating that out scale has a good level of internal consistency with this particular sample.

All variables selected in the study were included for a correlational check. Table 10 (below) presents the correlation matrix among all the variables. It shows that there seems to have been a correlation among EV, RV, SR, LC, CELF_ev, CELF_rs, APT_info, and APT_gram, similarly to TNW, NDW, MLU_w, TNC, and NDC. Thus, EFA was applied to condense the information in the correlation matrix.

		EVtotal	RVtotal	SRtotal	LCtotal	CELFev	CELFrs	APTinfo	APTgram	TNW	NDW	MLU(w)	TNC	NDC
EVtotal	Pearson correlation	1	.747**	.650**	.657**	.789**	.609**	.742**	.693**	.491**	.548**	.542**	.403**	.451**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	600	600	600	600	600	600	600	600	600	600	600	600	600

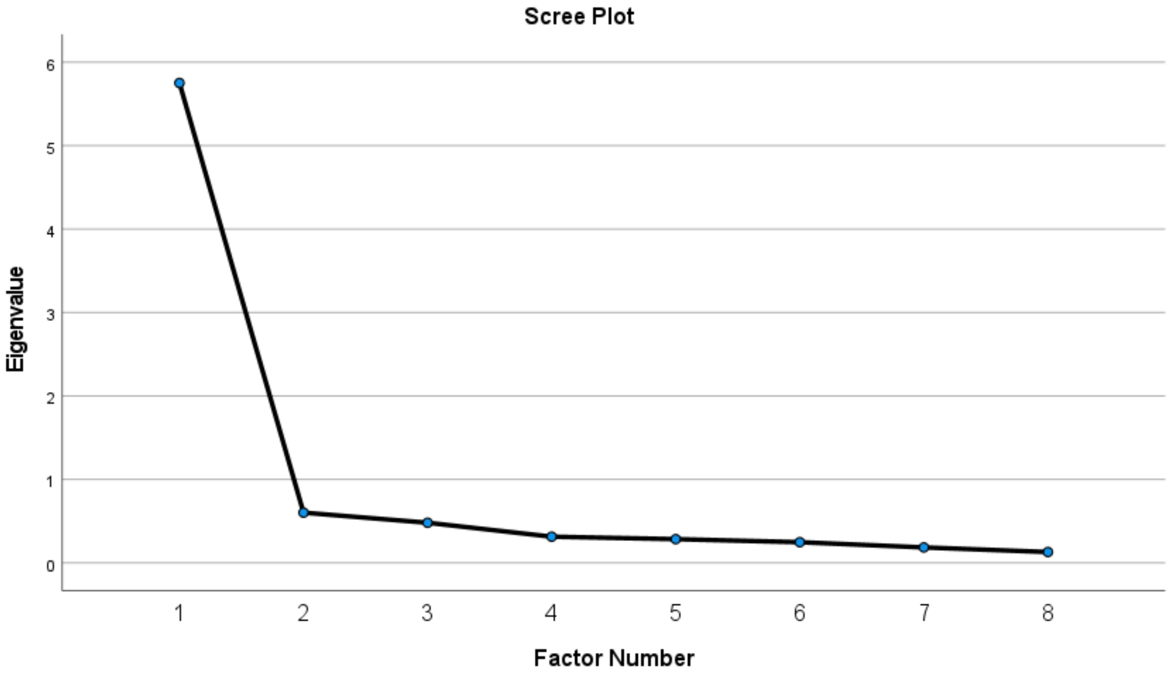
** Correlation is significant at the 0.01 level (2-tailed)

Table 10: Correlation matrix of all oral language and narrative-related variables (TNC: total number of conjunctions; NDC: total number of different conjunctions)

EFA Extraction Methods

Thus, the current study verified both the scree plot and its eigenvalue to decide the number of factors (see Figure 1 for scree plot and extraction output).

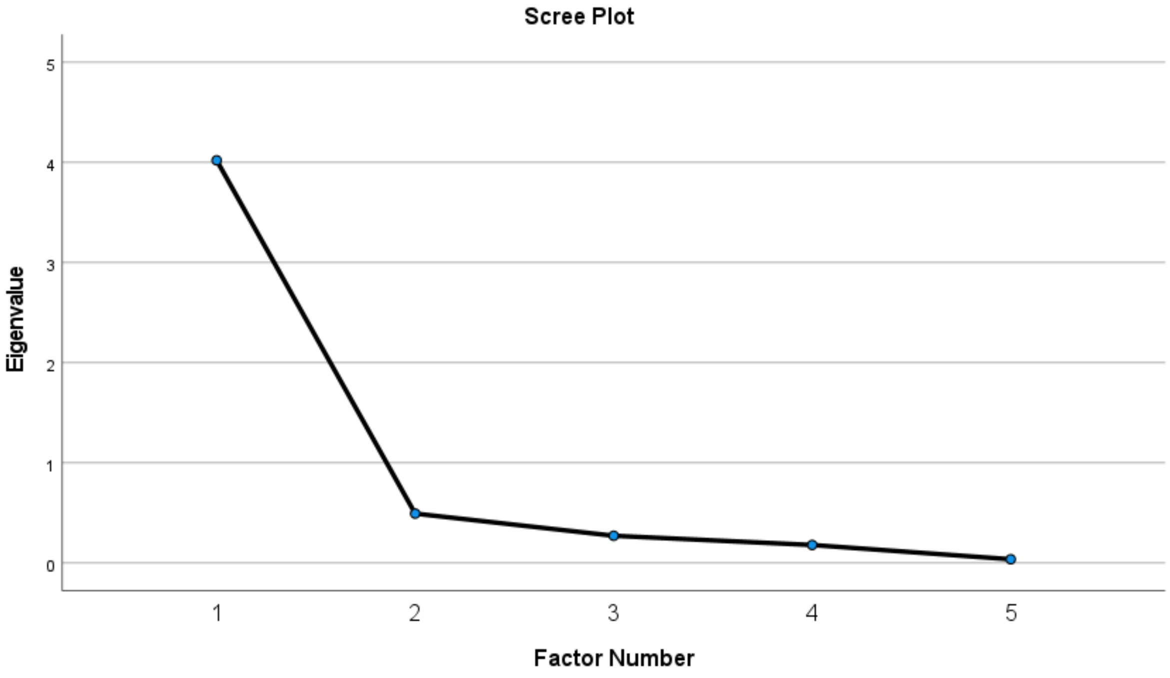
Scree Plot and factor matrix for variable 'Oral Language Skills'



	Factor
	1
EVtotal	.852
RVtotal	.767
SRtotal	.803
LCtotal	.779
CELFev	.867
CELFrs	.792
APT_info	.866
APT_gram	.861

Extraction Method: Principal Axis Factoring

Scree plot and factor matrix for ‘Narrative Retelling’



	Factor
	1
NT_TotWord	.918
NT_Tot Unique Word	.929
MLU_w	.844
NT_Conj	.837
NT_DiffConj	.815
Extraction Method: Principal Axis Factoring	

Figure 1: Scree plot and extraction output of oral language skills and narrative retelling

It suggests that there was only one factor each being extracted in narrative retelling and oral language skills. There is only one eigenvalue greater than 1, and the scree plot also levels out after one component with the majority being above .8. The data seems to suit the one-factor structure properly.

Factor Rotation

The results showed that only one factor was extracted with an eigenvalue greater than 1, indicating that that all the selected items fit one single theoretical construct, so no rotation is needed.

Specific variables were selected to properly answer each research question. The first research question explored the relationship between children's narrative retelling and other aspects of language skills using a simple bivariate correlational analysis. As presented earlier in this section, EFA was carried out twice. The first EFA examined the number of the factors under five observed variables. Only one factor was extracted with eigenvalue greater than 1 for narrative retelling. Similarly, the second EFA was conducted for language skills, and one factor was also extracted from eight observed variables after applying the data reduction technique.

The second research question examined the extent to which children's oral language skills predict their narrative retelling after controlling for demographic measures, and whether any demographic factors contribute to the prediction of narrative performance. Following the correlations, this question was addressed using a two-block hierarchical linear regression to identify the relative impact of the demographic inputs (i.e. age, gender, EAL, and school number).

4.3 Missing Data

Missing data is a prevalent issue in all types of studies with large datasets. Incomplete responses (e.g., dropout) are frequently unavoidable due to the nature of data gathering. When employing the NELI datasets with a sample size of above 600, missing data across sweeps and within sweeps is unavoidable. As a result, it is critical to consider why data is missing and how to cope with it in each study. Even if the results may not be biased, the missing data can reduce sample size, jeopardising statistical power (Langkamp, Lehman & Lemeshow, 2010; Madley-Dowd, et al., 2019). Nevertheless, power is not an issue with a large enough sample (n = 600) after missingness (Kang, 2013).

Missing data can be treated in a variety of ways, including listwise/pairwise deletion and single (SI) or multiple imputations (MI) (Ender, 2010). Imputation is a widely used flexible approach to handle missingness (Little and Rubin, 2019), although there are a few possible challenges to the principle of imputation (Dempster and Rubin, 1983).

The most common approach for missing data by far is to simply keep the remaining, which is called a listwise deletion (LD), as well as "complete case" (CC) analysis (Kang, 2013). It is a form of statistical analysis where participants with missing data will be excluded in the analysis (Jakobsen, et al, 2017:3; Garson, 2015:10). One of the most important limitations of LD is that I am essentially shifting the generalisation and statistical inference. Many researchers argue that when the missing cases are small (typically missingness less than 5% in larger samples, although no agreed-upon consensus on what constitutes a 'large sample' is provided), listwise deletion has been treated as the default measurement (Garson, 2015:10; Madley-Dowd, 2019; Schafer, 1999). According to the current dataset, 613 children completed both narrative retelling tests and other language tests. However, there are 13 missing cases where children's test scores were not reported. It was merely 2.2% of missing data. Thus, LD seems unlikely to impact and change the conclusion.

Moreover, it is very dependent on the nature of the missingness in comparison to the whole data. Little and Rubin (2019) have identified three standard scenarios for classification of how missing data are produced: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR) (Jakobsen et al., 2017; Dziura et al., 2013; Little et al., 2012). MCAR refers to any observed or unobserved variables that have no bearing on the likelihood of data loss (Dziura et al., 2013). It is usually caused by research design where the samples are missing in transit or because of the termination of a follow-up study (Kang, 2013:402-403; Dziura et al., 2013:345). MAR denotes how much missingness could be explained by observed variables (Garson, 2015:15). The likelihood of missing values has no bearing on the probability of missingness in MAR (ibid.). The third type of missing data MNAR occurs when missingness is dependent on unobserved variables (Dziura et al., 2013).

Little's MCAR test (Little, 1988) is the most often used test for instances that are missing completely at random. It is claimed that if the *p*-value is not significant for Little's MCAR tests, it may be considered that the data is MCAR. Thus, LD should be applied to all missing values (Garson, 2015:12). Table 11 presents the output of Little's MCAR test.

Chi-square	Degrees of freedom	P-value
24.460	46	.996

Table 11: Output tables of Little's MCAR test (See Appendix G for further details)

The test findings indicated that the missing data were regarded as MCAR (p = .996) since it is not statistically significant at the .996 level. Therefore, the assumption of MCAR is satisfied with a big sample size in the current dataset, the missing data were assumed not to affect the analysis, so LD had been taken into consideration to produce unbiased results, and only cases with no missing values would be analysed.

4.4 Inter-rater Reliability

In the present study, a second rater was recruited for inter-rater reliability check. The Kappa value is not considered to assess inter-rater reliability since that is designed for categorical data (Park and Kim, 2015; Morrison and Ferrari, 2009). As the narrative scores (e.g. MLU(w)) are measured on a continuous scale in this case, it was decided that the Pearson correlation coefficient r would be applied to indicate inter-rater reliability. This confirmed that the researcher implemented a narrative scoring system consistently and in accordance with the scoring guidance described in Brown (1973). The five correlation coefficients (Pearson's r) between the two scorers will be presented in Table 12 below:

Measuring items	Pearson correlation *r*	Sig. (2-tailed)
Total number of words	1	.000
Total number of unique words	1	.000
MLU(w)	.760	.000
Total number of conjunctions	.946	.000
Total number of unique conjunctions	.895	.000

Table 12: Correlation coefficient of MLU(w), the total number of words, the total number of unique words, the total number of conjunctions and the total number of unique conjunctions between rater 1 and rater 2

As demonstrated above, the inter-rater reliability was highly statistically significant ($p < .05$). The strong correlations between the two raters' narrative scores demonstrate a pattern of agreement.

4.5 Normality Test

The standard normal distribution was defined by Mishra et al., (2019:69) as follows: "the most important continuous probability distribution has a bell-shaped density curve". Normality tests vary under different criteria (Mishra et al., 2019:70). In this study, as the sample size is exceeded 300, normality was determined by histograms and absolute skewness and kurtosis values without taking additional standards (e.g. z-values) into account. Therefore, an absolute skewness less than or equal to 2 (≤ 2) and an absolute kurtosis ≤ 4 could be applied to consider its normality (ibid.:70; Kim, 2013). The normality check relied on two extracted factors: narrative retelling and oral language skills. With regard to the results in the present research, no extreme outliers were identified and, all data were acceptable once the normality tests were performed. Thereby, n = 600 remained in the final sample size. Table 13 (below) shows the standardised findings of the skewness and kurtosis values (See Appendix H for histograms).

Item	Skewness	Kurtosis
FAC_narrative retelling	0.590	-0.478
FAC_oral language skills	0.521	-0.513

Table 13: Standardised skewness and kurtosis values of FAC_narrative retelling and FAC_oral language skills

4.6 KMO and Bartlett's Test

The researcher computed Bartlett's Test of Sphericity and a Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy in verifying the strength and adequacy of data for EFA (Bartlett, 1954; Kaiser, 1970, 1974; Field, 2013). If the sphericity value ($p < .05$) is statistically significant, the correlation matrix can be factored, and is suitable for factor analyses (Mvududu & Sink, 2013:83; Pallant, 2010; Field, 2000). Although the threshold of KMO varies among different researchers, the default value of KMO above 0.5 indicates a sufficient and adequate sample (Hadi, Abdullah & Sentosa, 2016). This section will demonstrate the results of the KMO and Bartlett's Test for extracted independent variable oral language skills and dependent variable narrative retelling.

<u>Independent variable 'other aspects of oral language skills'</u>

According to Table 14, the KMO of the independent variable is 0.921, which exceeds the default value of 0.5 (Hadi, Abdullah & Sentosa, 2016). The result of Bartlett's Test of Sphericity is also statistically significant ($p < 0.05$), indicating that the correlation matrix is factorable (Mvududu & Sink, 2013:83).

Kaier-Meyer-Olkin Measure of Sampling Adequacy		.921
Bartlett's Test of Sphericity	Approx. Chi-Square	4175.921
	df.	28
	Sig.	.000

Table 14: KMO and Barlett's Test of Sphericity for independent variable 'other aspects of oral language skills'

Dependent variable ‘narrative retelling’

Similarly, Table 15 displays the KMO and Bartlett’s Test findings for the dependent variable. The KMO value is 0.785, which is also above the minimum value of 0.50. The factorability of the correlation matrix was confirmed by Bartlett’s Test of Sphericity, which was statistically significant ($p < 0.05$).

Kaier-Meyer-Olkin Measure of Sampling Adequacy		.758
Bartlett’s Test of Sphericity	Approx. Chi-Square	3369.907
	df.	10
	Sig.	.000

Table 15: KMO and Barlett’s Test of Sphericity for dependent variable ‘narrative retelling’

4.7 Research Question 1

What is the relationship between children’s narrative retelling and other aspects of oral language skills?

4.7.1 The Assumptions of Pearson’s Correlation Coefficients

The correlation coefficient between children’s narrative retelling and other aspects of language skills was evaluated in the current chapter. The analysis was based on the Pearson correlation coefficient and began with an evaluation of assumptions. Both variables satisfied the assumptions of this statistical test based on the following criteria (Schober, Boer & Schwarte, 2018; Field, 2019):

- Variables need to be measured at either an interval or ratio level
- Information is drawn from a random and representative sample
- Distribution must be normal and without significant outliers

In the current study, both variables are continuous and the normality of the sample size (n = 600) was believed to be achievable by analysing correlation coefficients when the sample exceeded 30 (Field, 2018).

4.7.2 The Results of Pearson's Correlation Coefficients

This question focused on the investigation of the relationship between children's narrative retelling and other oral language skills. The computed Pearson correlation coefficient demonstrated a statistically significant correlation between narrative retelling and language skills. With regards to Section 3.7.4 Tables 6 and 7 (Sabilla, Sarno & Triyana, 2019:82), these two continuous variables showed a strong, positive correlation, r (598) = .681 , p < .001 in a large effect size. Despite the argument advanced by Schober, Boer, and Schwarte (2018) that cut-off points varied among different studies, the coefficient of correlation of 0.681 found in this study could be interpreted by the rule of thumb as a 'strong' correlation. Therefore, the output indicates that children's narrative retelling was significantly correlated with their language skills (see Table 16 for statistical values).

		Narrative retelling
Oral language skills	Pearson coefficient	.681**
	Sig. (2-tailed)	.000
	95%confidence interval lower bound	.622
	95%confidence interval upper bound	.740

Table 16: Pearson correlation coefficient for children's narrative retelling (NR) and other aspects of language skills (OLS)

An overview of the findings is summarised in a scatter plot (see Figure 2). The x- and y-axis represent the extracted factor of oral language skills and narrative retelling, respectively. This scatterplot implies a positive linear correlation: the higher the level of language skills, the better the child's narrative retelling.

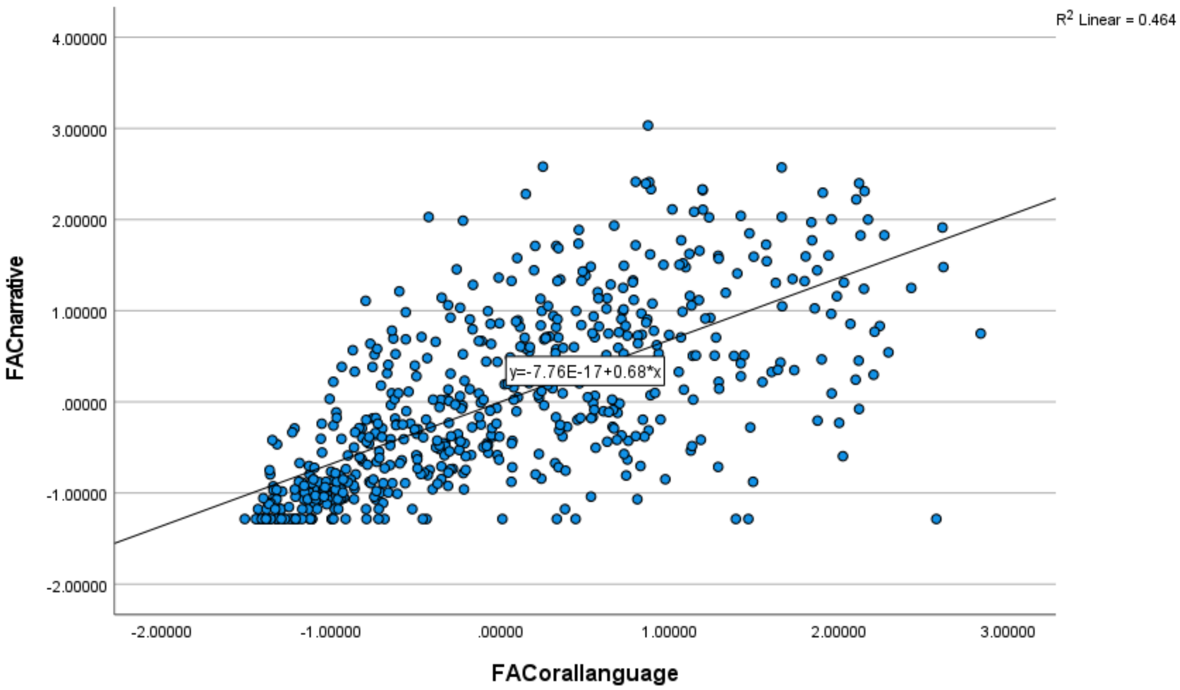


Figure 2: Factor_oral language skills and FAC_narrative retelling scatter plot

4.8 Research Question 2

To what extent do oral language skills predict children's narrative retelling, after controlling for demographic measures? What demographic measures contribute to the prediction of children's performance in narrative retelling?

The second research question focused on whether other demographic factors are expected to impact children's performance in narrative retelling. A two-model hierarchical linear regression analysis was performed in this subsection to identify the predictor and how it contributed to these criteria. The first model examined the effects of demographic factors (i.e. age, gender, EAL, and school number). The second model included the impact of the main predictor, the other aspects of oral language skills, after controlling for these potential demographic inputs.

4.8.1 The assumptions of hierarchical linear regression

Hierarchical linear regression aims to assess the prediction from a set of demographic variables after controlling for one or more of either variable. The assumptions of hierarchical linear regression need to be strictly followed in order to assure the adequacy of the regression analysis outputs. The main assumptions in this study will be considered as follows (To & Mandracchia, 2019:5):

- The residuals are distributed normally
- The variation of the residuals is similar at each point in the model
- There is no multicollinearity in the selected dataset
- The residual values are independent, and not autocorrelated

The Normal Probability Plot (P-P plot), scatterplot, and other analysing measurements for the model were applied for regression analysis.

In this case, the normality of the distribution could be verified using a histogram of the residuals in our model. Most of the points were positioned on the fit line, as presented in Appendix I Figure 4.1. Despite some points deviating sharply from the straight diagonal line, there was no evidence of a significant deviation from normality. As a consequence, the normality check was acceptable, and the residuals were normally distributed.

Furthermore, homoscedasticity was computed via a scatter plot in Appendix I Figure 4.2. The scatter plot showed that the majority of the residuals were randomly distributed with an even spread of data points on each predicted value. Thus, homoscedasticity under the selected dataset is guaranteed. Even though two probable outliers (marked red circles) were discovered, most of the residual points were distributed fairly, thus, it is still acceptable.

Multicollinearity is another important assumption in regression analysis. Belsley, Kuh, & Welsh (2005) described multicollinearity as a statistical phenomenon whereby two or more of the predicted variables in the regression model have an exact linear relationship (aka shared variance).

Multicollinearity was measured mainly based on two collinearity statistics: tolerance and variance inflation factors (VIF). The VIF is a technique for measuring the variation and quantifying the inflation of the variance (Daoud, 2017:4). The rules presented in Table 17 (below) are used to interpret the value of VIF (ibid.:4). Multiple relationships among variables exist if the VIF value is larger than 10 (Uyanık & Güler, 2013:238).

VIF	Interpretation
VIF = 1	No correlation
1 < VIF ≤ 5	Moderate correlation
VIF > 5	High correlation

Table 17: VIF values for interpretation

Multicollinearity can also be found by analysing the tolerance for each independent variable. It is usually defined as "the amount of variability in an independent variable that cannot be explained by other independent variables" (Daoud, 2017:5; Uyanık & Güler, 2013:238). Collinearity is indicated by a tolerance value smaller than 0.10 (ibid.).

As demonstrated in Appendix I Table 4.1, all predictors' VIF values were less than 10, and they all had tolerance values greater than 0.10. Thus, it is evident that there was no multicollinearity among the selected variables.

The fourth assumption was detected regarding the Durbin-Watson (DW) statistic. The DW statistic ranges from 0 to 4, with values closer to 2 showing that the assumption is met and the models do not have autocorrelation issues (Hassan et al., 2019). As reported by Appendix I Table 4.2, the DW value in the current study was 1.954, extremely near to 2, indicating that there was no autocorrelation in the residuals.

4.8.2 The results of hierarchical linear regression

Hierarchical linear regression was carried out in SPSS to examine the hypothesis that other aspects of oral language skills would statistically significantly predict children's narrative retelling after controlling for demographic variables (age, gender, EAL, SchoolNo). All the statistics for demographic variables were entered into the first block of the predictor variables. The extracted oral language skills score was put into the second block in rapid sequence alongside the scores attached in the first block. Table 18 presents a summary of the results from this analysis by presenting the effect size measures R^2, the adjusted R^2, their changes in R^2, and significant *p* values of the entire model:

Model	R	R Square	Adjusted R Square	R Square Change	Sig.
1	.360	.129	.123	.129	.000
2	.688	.473	.469	.344	.000

*Model 1: Narrative retelling = demographic variables (R^2 = 0.129)

*Model 2: Narrative retelling = demographic variables + oral language skills (R^2 = 0.473, ΔR^2 = 0.344)

Table 18: Results in predicting narrative retelling from a hierarchical linear regression

Table 18 includes the R and R^2 values of the model. In hierarchical linear regression, each block is an independent model. Field (2013, 2018) states that predictability improvements are assessed based on the increase in variance in the variables (aka R^2) between each hierarchical model. For this data, block 1 contains four demographic variables (i.e. age, gender, EAL, SchoolNo.) as predictors ($p < 0.05$), demonstrating quite a weak relationship ($R = 0.360$). Thus, it would appear that these variables were relatively less predictive of the outcome. As Table 18 displays, Model 1 (i.e., demographic factors) accounted for 12.9% of the variance in the narrative retelling. The table indicates that 87.1% of the variance in the narrative retelling remained unaccounted for, so other variables may have played a major role.

When other aspects of oral language skills were entered into the second block, together with previously inserted four demographic predictors accounted for 46.9% of the variation in narrative retelling. The added variable turned the model into a successful predictor of the result by showing a strong relationship (R = 0.688). As predicted, the difference in R^2 between models 1 and 2 was statistically significant, indicating that the second model with the addition of the variable oral language skill accounted for 34.4% of the variance in children's narrative retelling ($\Delta R^2 = 0.344$).

The results in ANOVA demonstrated that both models 1 and 2 were significant predictors of narrative retelling, with $F(4, 595) = 22.089$, $p < 0.05$, and $F(5,594) = 106.626$, $p < 0.05$, respectively (Appendix I Table 4.3). The coefficient table in Appendix I Table 4.4 shows the contribution of each individual predictor variable to the model. In Model 1, apart from the predictor SchoolNo ($p = 0.834 > 0.05$), the other variables (EAL, gender, and age) were all positively and significantly correlated to the narrative retelling with coefficients $r = .309$, $r = .111$, and $r = .148$, respectively ($p < 0.05$) (see Appendix I Table 4.5). Thus, Model 1 indicated that these three predictors were associated with children's narrative retelling; however, both gender and age demonstrated a weak correlation compared with EAL. As a consequence, we concluded that the regression model shows a good fit of the data, and that results under this model can predict children's narrative retelling significantly.

4.9 Summary

In this chapter, all of the performed analyses were thoroughly addressed to examine two main overarching questions. The sample was normally distributed according to the normality tests. The validity and reliability of the instruments used in this study were confirmed by exploratory factor analysis and Cronbach's alpha coefficient test. Correlational analysis was applied to construct the correlation between the independent variable, 'other aspects of oral language skills', and the dependent variable, 'narrative retelling'. Hierarchical linear regression was also designed to investigate the contribution of demographic predictors on children's narrative retelling. Chapter 5 will explain these findings in greater depth.

Chapter 5: Discussion and Conclusions

5.1 Overview of the Chapter

The preceding chapter described the analysis and results of the two main research questions that guided the current study. The major goal of this chapter is to combine the material from the results and literature review chapters in order to further interpret the findings in connection to each research question. Following this, I will discuss the methodological constraints of the current study that may restrict the applicability and generalisability of the findings. Next, future research suggestions will be presented based on the potential limitations. The contributions and implications of the current study for pedagogical research will then be discussed before drawing a concise conclusion to the whole paper.

5.2 Interpretation of Findings

The primary goal of the current study was to examine the relationship between children's narrative retelling and other aspects of language skills. Based on the first research aim, the second goal focused on the predictions of demographic factors on children's narrative retelling performance. First, a bivariate correlation analysis was carried out on two extracted variables. Second, demographic inputs were explored with a hierarchical linear regression.

A significant highly positive correlation coefficient was found between scores on narrative retelling and other aspects of language skills. It was therefore considered appropriate to combine scores on these items to create factor scores. Oral language skills were found to directly impact the narrative retelling performance: the better a child's oral language, the higher their retelling score. While prior research has found that retelling can be used to assess children's language development (Reed & Vaughn, 2012; Morrow, 1984, 1985, 1986), the present study also reveals that retelling can also be evaluated by measuring oral language skills. Thus, the correlation works both ways, and both narrative retelling and other aspects of oral language skills can be categorised as effective tools for assessing language developmental skills. In terms of the regression conducted for demographic variables (i.e. age, gender, EAL, SchoolNo.), all four of the inputs accounted for only 12.9% of the variance in the narrative retelling, oral language skill, alone accounted for 34.4% of the variation. Therefore, oral language can largely predict children's narrative performance after controlling for demographic measures. Within the four demographic inputs, only age, gender, and EAL significantly contributed to the prediction of narrative even though with fewer contributions compared with the main factor (i.e. other aspects of oral language skills). The results demonstrated that narrative retelling was well-delivered in older, female and predominantly native English-speaker groups. The findings also consolidated previous research by Crow et al. (1998), Lips (1993), Friffith, Ripich, and Dastoli (1986), and John, Lui, and Tannock (2003) on age and gender impacts on retelling.

5.3 Limitations

Although this study can be considered as a contribution to research on the relationship between children's narrative retelling and other aspects of oral language skills, it is important to highlight several methodological limitations of the current study that may have influenced drawing conclusive inferences from the findings, and thus less likely to obtain generalised results.

First and foremost, the study was limited by access to the original voice recordings. I was unable to gain access to the audio-recorded files without breaching the General Data Protection Regulation (GDPR) signed between the former researcher and schools. This implicitly increased the margin of error in interpreting where children paused or stopped, and their states or moods of narrative retelling (i.e. happy, laboured) as this information had to be gleaned second-hand from the accounts of former transcribers. In addition, these transcribers may have wrongly interpreted the speech of children who were non-native English speakers. They may also have been unused to children's speech styles and natural speech impediments. As a result, I would speculate, for example, that in one of the children's retelling transcriptions the child may have said "got a bit wet" but been unable to pronounce the word 'bit' with the precision of an adult speaker. However, the transcription marked 'dip' in brackets. This type of uncertainty arose in many of the transcriptions: without access to the audio files it was difficult to verify whether the mistake was made by the child or the transcriber. In addition, punctuation and capitalisation were not used consistently (possibly owing to different transcribers). This influenced the marking of where the utterances ended and began, in turn misleading the researcher and second-rater in defining utterance segmentation. Furthermore, frequent word or phrase repetition in the transcriptions increased, the difficulty of counting the TNW, NDW, and MLU(w) as neither the second scorer nor I could tell whether this repetition was caused by emphasising or stuttering without directly accessing the original recordings. Finally, ellipsis in the transcription further confused the scorers as there was no way of knowing whether this denoted the children saying something unintelligible or children stuttering to repeat the word.

An additional limitation of the current study is the scoring of children's narrative retelling, particularly in the calculation of MLU(w). In line with previous research, 50 utterances are usually considered an adequate and representative sample of children's language, with MLU demonstrating acceptable reliability (Miller & Chapman, 1981; Pezold, Imgrund & Storkel, 2020; Casby, 2011). In Guo and Eisenberg's study (2014), they discovered that samples of 100 utterances were more diagnostically accurate and trustworthy than samples of 50 utterances. In this case, none of the children produced 50 utterances, the average number of utterances being around five or six, meaning that the data obtained in the current study would not meet the criteria for reliability based on volume of language sample outlined above (Parker & Brorson, 2005:370). However, Brorson and Dewey (2005) found that there was no significant difference in MLU(w) scores regardless of the length of the transcripts. According to these findings, the calculation of the number of utterances is fairly arbitrary. Without a fixed standard for segmenting and calculating children's utterances, some error and extreme values are inevitable. For example, some of the children acquired higher scores on MLU(w) by repeating the same idea with different words, and well-inserted conjunctions to connect sentences with similar meanings together. Conversely, those who retold the story with a certain number of segmented utterances better addressed the key points but obtained much lower scores. In short, the results were inexorably influenced by the constraints and biases inherent in secondary data analysis.

5.4 Future Research

The first recommendation for further research would be to thoroughly assess the validity and reliability of the narrative scoring systems. None of the scoring systems are without flaws. It will be more productive for researchers to verify convergent validity by using multiple scoring methods on the same set of data, as well as examining whether the findings are influenced considerably by the selection of different scoring systems. Furthermore, it is recommended to measure both micro- and macro-structural levels as they evaluate a person's narrative performances in both lexical/grammatical and coherent/organisational areas. There are other ways to measure children's narrative retelling, such as SNAP (Strong, 1998), MAIN (Gagarina, et al., 2012), or NSS (Heilmann, Miller & Dunaway, 2010) for analysing narrative macrostructure, and it is necessary for researchers to compare these different metrics to ascertain which on will best address their research questions.

The sample size of the current study is satisfactory. The second recommendation is to further investigate the current dataset, considering, for example, the socioeconomic background of each child, their home language environment, and their bedtime reading materials. Although the present research found oral language skills to be a main predictor of narrative retelling performance (i.e. accounting for 34.4% of the variance in children's retelling) after controlling for provided demographic inputs, a large number of variances remained unaccounted for. Hence, further research into this dataset can explore other variables significantly influencing narrative retelling.

5.5 Implications for and Contributions to Pedagogical Research

The findings explored here on oral language and narrative retelling have massive implications for both theory and educational policy. Theoretically, they underline the strong connection between the structural components of language, it also highlight that although language is a complex system, it reflect a unitary trait. As shown in previous literature, it is noted that oral language and reading are significantly correlated (Hill, 2009); similarly, Reed and Vaughn (2012) found that retelling is highly correlated with reading development. Our study concluded that oral language skills can promote positive outcomes in children's narrative retelling performance. Thus, a key implication of this research is that oral exercises should be encouraged in the classroom for their positive effect on narrative fluency. School teachers need to instruct their students to perform more narrative retellings in the classroom. Oral-based activities usually provide pupils with opportunities to develop their oral language skills, and are helpful for those in the early stages of language acquisition. Practising with oral language, particularly narrative retelling, can not only benefit children's academic achievement, but also promote their ability to logically organise and reconstruct texts in the future. Morrow (1985) also recommended that teachers should positively expose children to the narrative environment and interact with them in storytelling. In this vein, both retelling and oral language have powerful implications for language assessment and testing. Although the benefits of oral language and narrative retelling extend far beyond increasing children's academic test scores, they can be an efficient tool for teachers to boost students' language abilities and thus their performance in written examinations.

5.6 Conclusion

Based on a novel app-based test (ATLAS) measuring oral language skills and a non-standardised measurement for narrative retelling, the present study investigated the association between narrative retelling and other oral language skills. The results indicated that children's narrative retelling is highly correlated with oral language skills. My work contributes further evidence to substantiate the view that language skills can all be treated as one general factor (as covered in Hulme, et al., 2020). Increased awareness of the relationship between narrative retelling and oral language skills can aid in educational development. Furthermore, we discovered that age, gender, and EAL all have a minor impact on children's narrative retelling. This observation should guide school teachers in arranging the classroom environment to meet specific needs. For example, boys should be encouraged to exhibit greater loquacity and attention to detail in narrative retelling tasks. Teachers of younger children can be encouraged to devote greater attention to building vocabulary, which will effectively boost children's retelling skills.

Additionally, narrative retelling and oral language can both be considered beneficial for children of all ages. Both can be used as a diagnostic tool to assess children's reading and listening comprehension skills. Retelling may also assist teachers in shifting their perspective from reading as a set of isolated skills to reading as a process for converting and recreating meaning. These far-reaching implications for both theory and educational policy can direct future research. Overall, without an application of traditional assessments, this study sheds new light on the existing research.

Language is one unitary factor. Narrative and oral language become another useful and valid way of measuring children’s language development at schools.

References

Afreen, S., Kamal Butt, A., & Malik, A. N. (2014). Development of an Action Picture Information Test (APIT): JRCRS. 2014; 2 (2): 33-37. *Journal Riphah College of Rehabilitation Sciences, 2*(2), 33-37.

Andreou, G., & Lemoni, G. (2020). Narrative Skills of Monolingual and Bilingual Pre-School and Primary School Children with Developmental Language Disorder (DLD): A Systematic Review. *Open Journal of Modern Linguistics*, *10*(05), 429. https://doi.org/10.4236/ojml.2020.105026

Arlman-Rupp, A. J., Van Niekirk-de Hahn, D. & Van de Sandt-Koenderman, M. (1976). Brown's early stages: Some evidence from Dutch. Journal of Child Language, 3, 267–274. https://doi.org/10.1017/S0305000900001483

Bartlett, M. S. (1954). A note on the multiplying factors for various χ 2 approximations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 296-298. https://doi.org/10.1111/j.2517-6161.1954.tb00174.x

Beers, K. (2003). *When Kids Can't Read: What teacher's can do, 20*(3). Portsmouth, NH: Heinemann.

Bloomfield, L. (1926). A set of postulates for the science of language. *Language*, *2*(3), 153-164. https://doi.org/10.2307/408741

Boeren, E. (2018). The methodological underdog: A review of quantitative research in the key adult education journals. *Adult Education Quarterly*, *68*(1), 63-79. https://doi.org/10.1177/0741713617739347

Bolderson, S., Dosanjh, C., Milligan, C., Pring, T., & Chiat, S. (2011). Colourful semantics: A clinical investigation. *Child Language Teaching and Therapy*, *27*(3), 344-353. https://doi.org/10.1177/0741713617739347

Botting, N. (2002). Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child language teaching and therapy*, *18*(1), 1-21. https://doi.org/10.1191/0265659002ct224oa

Brody, N. (2018). Factor analysis: Rotated matrix. In: Allen, M. (Ed.) *The SAGE encyclopedia of communication research methods*, pp. 526-531. Thousand Oaks: SAGE Publications. https://dx.doi.org/10.4135/9781483381411.n190

Brorson, K. & Dewey, C. (2005). Effect of language sample size on MLUw. Hearsay, 17, 46–56

Brown, A. (1975). Recognition, reconstruction and recall of narrative sequences of preoperational children. Child Development, 46, 155-156. https://doi.org/10.2307/1128844

Brown, H. D. (2000). *Principles of language learning and teaching* (Vol. 4). New York: Longman. https://doi.org/10.2307/414380

Brown, R. (1973). A First Language, Cambridge, MA: Harvard University Press.

Casby, M. W. (2011). An examination of the relationship of sample size and mean length of utterance for children with developmental language impairment. *Child Language Teaching and Therapy*, *27*(3), 286-293. https://doi.org/10.1177/0265659010394387

Castilla-Earls, A., Petersen, D., Spencer, T., & Hammer, K. (2015). Narrative development in monolingual Spanish-speaking preschool children. *Early Education and Development*, *26*(8), 1166-1186. https://doi.org/10.1080/10409289.2015.1027623

Cattell, R. B. (1966b). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*(2), 245-276. https://doi.org/10.1207/s15327906mbr0102_10

Chomsky, N. (1986). Knowledge of language: Its nature, origin, and use. New York: Praeger.

Cleave, P. L., Girolametto, L. E., Chen, X., & Johnson, C. J. (2010). Narrative abilities in monolingual and dual language learning children with specific language impairment. Journal of Communication Disorders, 43, 511–522. https://doi.org/10.1016/j.jcomdis.2010.05.005

Comrey, A. L., & Lee, H. B. (2013). *A first course in factor analysis*. Psychology press.

Corner, S. (2009). Choosing the right type of rotation in PCA and EFA. *JALT testing & evaluation SIG newsletter*, *13*(3), 20-25.

Cowell, M. D., & Entwistle, N. J. (1971). The relationships between personality, study attitudes and academic performance in a technical college. *British Journal of Educational Psychology*, *41*(1), 85-90. https://doi.org/10.1111/j.2044-8279.1971.tb00661.x

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *psychometrika*, *16*(3), 297-334. https://doi.org/10.1007/BF02310555

Crookes, G. (1990). The utterance, and other basic units for second language discourse analysis. *Applied linguistics*, *11*(2), 183-199. https://doi.org/10.1093/applin/11.2.183

Crow, T.J., Crow, I..R., Done, D.J., & Leask, S. (1998). Relative hand skills predicts academic ability: Global deficits at the point of hemispheric indecision. Neuropsychologia, 36(12), 1275-1282. https://doi.org/10.1016/s0028-3932(98)00039-6

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological methods*, *1*(1), 16. https://doi.org/10.1037/1082-989X.1.1.16

D'agostino Sr, R. B., & Russell, H. K. (2005). Scree test. *Encyclopedia of biostatistics*, *7*. https://doi.org/10.1002/0470011815.b2a10082

Davidson, A. J., Walton, M. D., Kansal, B., & Cohen, R. (2017). Narrative skills predict peer adjustment across elementary school years. *Social Development*, *26*(4), https://doi.org/891-906. 10.1111/sode.12219

Daoud, J. I. (2017). Multicollinearity and regression analysis. In *Journal of Physics: Conference Series*. *949*(1), 1-7. IOP Publishing. https://doi.org/10.1088/1742-6596/949/1/012009

DeCoster, J. (1998). Overview of factor analysis.

Dempster, A. P., & Rubin, D. B. (1983). Overview. Incomplete Data in Sample Surveys, Vol II: Theory and Bibliographies. Academic, New York, 3–10. In: Madow, W. G., Olkin I., & Rubin, D. B. (eds). https://doi.org/10.1016/0198-9715(90)90059-3

Diez-Itza, E., Martínez, V., Pérez, V., & Fernández-Urquiza, M. (2018). Explicit oral narrative intervention for students with Williams syndrome. *Frontiers in psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.02337

Dockrell, J. E., & Marshall, C. R. (2015). Measurement issues: assessing language skills in young children. *Child and Adolescent Mental Health*, *20*(2), 116-125. https://doi.org/10.1111/camh.12072

Donnellan, M., & Lucas, R. (2013). Secondary Data Analysis. *The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2, Statistical Analysis*. https://doi.org/10.1093/oxfordhb/9780199934898.013.0028

Dziura, J. D., Post, L. A., Zhao, Q., Fu, Z., & Peduzzi, P. (2013). Strategies for dealing with missing data in clinical trials: from design to analysis. *The Yale journal of biology and medicine*, *86*(3), 343.

Eilers, R. E. (1975). Suprasegemental and grammatical control over telegraphic speech in young children. *Journal of Psycholinguistic Research*, *4*(3), 227-239. https://doi.org/10.1007/BF01066928

Eisenberg, S. L., Fersko, T. M., & Lundgren, C. (2001). The use of MLU for identifying language impairment in preschool children. https://doi.org/10.1044/1058-0360(2001/028)

Elkind, D. (1986). Formal Education and Early Childhood Education: An Essential Difference. *Phi Delta Kappan, 67*(9), 631-636.

Ezeizabarrena, M. J., & Garcia Fernandez, J. I. (2018). Length of utterance, in morphemes or in words?: MLU3-w, a reliable measure of language development in early basque. *Frontiers in psychology*, *8*, 2265. https://doi.org/10.3389/fpsyg.2017.02265

Fabrigar, L. R., & Wegener, D. T. (2012). Exploratory factor analysis. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199734177.001.0001

Fagot, B. I., & Kavanagh, K. (1993). Parenting during the second year: effects of children's age, sex, and attachment classification. Child Development, 64, 258-271. https://doi.org/10.1111/j.1467-8624.1993.tb02908.x

Fazio, B. B., Naremore, R. C., & Connell, P. J. (1996). Tracking children from poverty at risk for specific language impairment: A 3-year longitudinal study. *Journal of Speech, Language, and Hearing Research, 39*(3), 611-624. https://doi.org/10.1044/jshr.3903.611

Field, A. (2009). *Discovering statistics using SPSS: (And sex and drugs and rock "n" roll)* (3rd ed.). Los Angeles, CA: Sage.

Field, A. (2013). Discovering statistics using SPSS (4th Edition). Thousand Oaks, CA: Sage.

Field, A. (2018). Discovering statistics using SPSS (5th ed). London, Engladn: Sage.

Fielding, N. G & Fielding, J. L (2003). Resistance and adaptation to criminal identity: Using secondary analysis to evaluate classic studies of crime and deviance. *Sociology*, 34(4): 671–689. https://doi.org/10.1177/S0038038500000419

Fricke, S., Bowyer-Crane, C., Haley, A. J., Hulme, C., & Snowling, M. J. (2013). Efficacy of language intervention in the early years. *Journal of Child Psychology and Psychiatry*, *54*(3), 280-290. https://doi.org/10.1111/jcpp.12010

Fricke, S., Bowyer-Crane, C., Snowling, M. J., & Hulme, C. (2018). The Nuffield early language intervention. Oxford: Oxford University Press.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific studies of reading*, *5*(3), 239-256. https://doi.org/10.1207/S1532799XSSR0503_3

Gabig, C. S. (2013) Telegraphic Speech. In: Volkmar F.R. (eds) Encyclopedia of Autism Spectrum Disorders. Springer, New York, NY. https://doi.org/10.1007/978-1-4419-1698-3_1123

Gagarina, N. V., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., ... & Walters, J. (2012). MAIN: Multilingual assessment instrument for narratives. *ZAS papers in linguistics*, *56*, 155-155.

Gagarina, N., Klop, D., Tsimpli, I. M., & Walters, J. (2016). Narrative abilities in bilingual children. Applied Psycholinguistics, 37, 91–122. https://doi.org/10.1017/S01427164150000399

Gambrell, L. B. (2004). Exploring the connection between oral language and early reading. *The Reading Teacher*, *57*(5), 490.

Gambrell, L. B., Koskinen, P. S., & Kapinus, B. A. (1991). Retelling and the reading comprehension of proficient and less-proficient readers. *The Journal of Educational Research*, *84*(6), 356-362. https://doi.org/10.1080/00220671.1991.9941816

Garson, G. D. (2015). Missing values analysis and data imputation. Asheboro, NC: Statistical Association Publishers, 1-26

Genishi, C. (1988). Young Children's Oral Language Development. ERIC Digest.

George, D., & Mallery, P. (2005). *SPSS for Windows step by step : A simple guide and reference, 12.0 update* (5th ed.). Boston ; London: Pearson.

Gillam, R. B., & Pearson, N. (2004). Test of narrative language. Austin, TX: PRO-ED

Gillam, S. L., Gillam, R. B., and Reece, K. (2012). Language outcomes of contextualized and decontextualized language intervention: results of an early efficacy study. Lang. Speech Hear. Serv. Sch. 43, 276–291. doi: 10.1044/0161- 1461(2011/11-0022)

Gorsuch, R. L. (1983). Factor analysis (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates

Graesser, A., Golding, J. M., & Long, D. L. (1996). Narrative presentation and comprehension. In: Barr, R., Kamil, M. L., Mosenthal, P. B., & Pearson, P. D. (Eds.), Handbook of reading research, II. 191–205. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Griffin, T. M., Hemphill, L., Camp, L., & Wolf, D. P. (2004). Oral discourse in the preschool years and later literacy skills. First Language, 24, 123–147. https://doi.org/10.1177/0142723704042369

Griffith, P. L., Ripich, D. N., & Dastoli, S. L. (1986). Story structure, cohesion, and propositions in story recalls by learning-disabled and nondisabled children. *Journal of Psycholinguistic Research*, *15*(6), 539-555. https://doi/org/10.1007/BF01067635

Guo, L. Y., & Eisenberg, S. (2014). The diagnostic accuracy of two tense measures for identifying 3-year-olds with language impairment. *American Journal of Speech-Language Pathology*, *23*(2), 203-212. https://doi/org/10.1044/2013_AJSLP-13-0007

Gutiérrez-Clellen, V. F. (2002). Narratives in two languages: Assessing performance of bilingual children. *Linguistics and Education*, *13*(2), 175-197. https://doi.org/10.1016/S0898-5898(01)00061-4

Haccoun, R. R. (2001) 'Test review of the Renfrew Bus Story'. In: B. S. Plake and J. C. Impara (eds), The 14th Mental Measurements yearbook.

Hadi, N. U., Abdullah, N., & Sentosa, I. (2016). An easy approach to exploratory factor analysis: Marketing perspective. *Journal of Educational and Social Research*, *6*(1), 215-215. https://doi.org/10.5901/jesr.2016.v6n1p215

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). Multivariate data analysis new york. *NY: Macmillan*.

Han, Y. (2018). Strong Narrative Assessment Procedure (SNAP), 2nd Edition. In F. R. Volkmar (Ed.), Encyclopedia of Autism Spectrum Disorders. Yale Child STudy Centre, New Haven, USA. https://doi.org/10.1007/978-1-4614-6435-8_102255-1

Hao, Y., Sheng, L., Zhang, Y., Jiang, F., de Villiers, J., Lee, W., & Liu, X. L. (2018). A narrative evaluation of mandarin-speaking children with language impairment. *Journal of Speech, Language, and Hearing Research*, *61*(2), 345-359. https://doi.org/10.1044/2017_JSLHR-L-16-0367

Hassan, S. H., Aziz, H. A., Zakariah, Z., Noor, S. M., Majid, M. R. A., & Karim, N. A. (2019). Development of linear regression model for brick waste generation in Malaysian construction industry. In *Journal of Physics: Conference Series*. *1349* (1), 1-8. IOP Publishing. https://doi.org/10.1088/1742-6596/1349/1/012112

Heilmann, J., Miller, J. F., Nockerts, A., & Dunaway, C. (2010). Properties of the narrative scoring scheme using narrative retells in young school-age children. https://doi.org/10.1044/1058-0360(2009/08-0024)

Hickey, T. (1991). Mean length of utterance and the acquisition of Irish. Journal of Child Language, 3, 553–569. https://doi.org/10.1017/S0305000900011247

Hill, S. (2009). Oral language and beginning reading: exploring connections and disconnections. *Forum on Public Policy*, 1-13.

Hipfner-Boucher, K., Milburn, T., Weitzman, E., Greenberg, J., Pelletier, J., & Girolametto, L. (2015). Narrative abilities in subgroups of English language learners and monolingual peers. *International Journal of Bilingualism*, *19*(6), 677-692. https://doi.org/10.1177/1367006914534330

Hoff, E. (2013). *Language development*. Cengage Learning.

Hulme, C., Snowling, M. J., West, G., Lervåg, A., & Melby-Lervåg, M. (2020). Children's language skills can be improved: Lessons from psychological science for educational policy. *Current Directions in Psychological Science*, *29*(4), 372-377. https://doi.org/10.1177/0963721420923684

Jakobsen, J. C., Gluud, C., Wetterslev, J., & Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials–a practical guide with flowcharts. *BMC medical research methodology*, *17*(1), 1-10. https://doi.org/10.1186/s12874-017-0442-1

John, S. F., Lui, M., & Tannock, R. (2003). Children's story retelling and comprehension using a new narrative resource. *Canadian Journal of School Psychology*, *18*(1-2), 91-113. https://doi.org/10.1177/082957350301800105

Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, *35*(4), 401-415. https://doi.org/10.1007/BF02291817

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*(1), 31-36. https://doi.org/10.1007/BF02291575

Kang, H. (2013). The prevention and handling of the missing data. *Korean journal of anesthesiology*, *64*(5), 402. https://doi.org/10.4097/kjae.2013.64.5.402

Keene, E. O., & Zimmermann, S. (2007). Mosaic of thought: The power of comprehension strategy instruction. (2nd ed.). Portsmouth, NH: Heinemann.

Kiecolt, K., & Nathan, L. (1985). *Secondary analysis of survey data* (Quantitative applications in the social sciences; No. 07-053). Beverly Hills, Calif.: Sage Publications.

Kim, H. Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restorative dentistry & endodontics*, *38*(1), 52-54. https://doi.org/ 10.5395/rde.2013.38.1.52

Kim, J. O., Ahtola, O., Spector, P. E., & Mueller, C. W. (1978). *Introduction to factor analysis: What it is and how to do it* (No. 13). Sage. https://dx.doi.org/10.4135/9781412984652

Klee, T. (1992). Developmental and diagnostic characteristics of quantitative measures of children’s language production. Topics in Language Disorders, 12, 28–40. https://doi.org/10.1097/00011363-199202000-00005

Klem, M., Melby-Lervåg, M., Hagtvet, B., Lyster, S. A. H., Gustafsson, J. E., & Hulme, C. (2015). Sentence repetition is a measure of children's language skills rather than working memory limitations. *Developmental science*, *18*(1), 146-154. https://doi.org/10.1111/desc.12202

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.

Koskinen, P. S., Gambrell, L. B., Kapinus, B. A., & Heathington, B. S. (1988). Retelling: A strategy for enhancing students' reading comprehension. *The Reading Teacher*, *41*(9), 892-896.

Kunnari, S., Välimaa, T., & Laukkanen-Nevala, P. (2016). Macrostructure in the narratives of monolingual Finnish and bilingual Finnish–Swedish children. Applied Psycholinguistics, 37(1), 123-144. https://doi.org/10.1017/S0142716415000442

Langkamp, D. L., Lehman, A., & Lemeshow, S. (2010). Techniques for handling missing data in secondary analyses of large surveys. *Academic pediatrics*, *10*(3), 205-210. https://doi.org/10.1016/j.acap.2010.01.005

Lightbown, P., & Spada, N. (2013). *How languages are learned* (Fourth ed., Oxford handbooks for language teachers). Oxford.

Liles, B. Z. (1985). Cohesion in the narratives of normal and language-disordered children. *Journal of Speech, Language, and Hearing Research*, *28*(1), 123-133. https://doi.org/10.1044/jshr.2801.123

Liles, B. A., Duffy, R. J., Merritt, D. D., & Purcell, S. L. (1995). Measurement of narrative discourse ability in children with language disorders. Journal of Speech and Hearing Research, 38, 415–425. https://doi.org/10.1044/jshr.3802.415

Lindfors, J. W. (1987). Children’s language and learning. (2nd ed.). Englewood Cliffs: Prentice-Hall.

Lips, H. M. (1993). Masculinity and femininity: Myths and stereotypes. In Sex and gender: An introduction (pp. 1-36). Mountain View, CA: Mayfield Publishing Company.

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404), 1198-1202.

Little, R. J., D'Agostino, R., Cohen, M. L., Dickersin, K., Emerson, S. S., Farrar, J. T., Frangakis, C., Hogan, J. W., Molenberghs, G., Murphy, S. A & Stern, H. (2012). The prevention and treatment of missing data in clinical trials. *New England Journal of Medicine*, *367*(14), 1355-1360. https://doi.org/10.1056/NEJMsr1203730

Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data*, *793*. John Wiley & Sons.

Lucero, A. (2018). Oral narrative retelling among emergent bilinguals in a dual language immersion program. *International Journal of Bilingual Education and Bilingualism*, *21*(2), 248-264. https://doi.org/10.1080/13670050.2016.1165181

Madley-Dowd, P., Hughes, R., Tilling, K., & Heron, J. (2019). The proportion of missing data should not be used to guide decisions on multiple imputation. *Journal of clinical epidemiology*, *110*, 63-73. https://doi.org/10.1016/j.jclinepi.2019.02.016

Matsunaga, M. (2010). How to Factor-Analyze Your Data Right: Do's, Don'ts, and How-To's. *International journal of psychological research*, *3*(1), 97-110. https://doi.org/10.21500/20112084.854

McArthur, D., Adamson, L. B., & Deckner, D. F. (2005). As stories become familiar: Mother-child conversations during shared reading. *Merrill-Palmer Quarterly (1982-)*, 389-411. https://doi.org/10.1353/mpq.2005.0025

McCauley, R. J., & Swisher, L. (1984). Psychometric review of language and articulation tests for preschool children. Journal of Speech and Hearing Disorders, *49*, 34-42. https://doi.org/10.1044/jshd.4901.34

McNeish, D. (2017). Exploratory factor analysis with small samples and missing data. *Journal of personality assessment*, *99*(6), 637-652. https://doi.org/10.1080/00223891.2016.1252382

Méndez, L. I., Perry, J., Holt, Y., Bian, H., & Fafulas, S. (2018). Same or different: Narrative retells in bilingual Latino kindergarten children. *Bilingual Research Journal*, *41*(2), 150-166. https://doi.org/10.1080/15235882.2018.1456984

Miller, J. F., Chapman, R. S., Branston, M. B., & Reichle, J . (1980). Language comprehension in sensorimotor stages V and VI. Journal of Speech and Hearing Research, 23, 284-311. https://doi.org/10.1044/jshr.2302.284

Miller, J. F. (1981). Assessing language production in children. Baltimore, MD: University Park Press

Miller, J. F., & Chapman, R. S. (1981). The relation between age and mean length of utterance in morphemes. *Journal of Speech, Language, and Hearing Research*, *24*(2), 154-161. https://doi.org/10.1044/jshr.2402.154

Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. *Annals of cardiac anaesthesia*, *22*(1), 67-72. https://doi.org/10.4103/aca.ACA_157_18

Montessori, M. (1964). The Montessori method. Cambridge, MA: Robert Bentley.

Morrison, S. C., & Ferrari, J. (2009). Inter-rater reliability of the Foot Posture Index (FPI-6) in the assessment of the paediatric foot. *Journal of foot and ankle research*, *2*(1), 1-5. https://doi.org/10.1186/1757-1146-2-26

Morrow, L. (1984). Effects of story retelling on young children’s comprehension and sense of story structure. In J. A. Niles & L. A. Harris (Eds.), Changing perspecfives on research in reading/language pmming and hfrucrion. Thirty-third yearbook of the National Reading Conference. 95-100. Rochester, NY: National Reading Conference.

Morrow, L. M. (1985). Retelling stories: A strategy for improving young children's comprehension, concept of story structure, and oral language complexity. *The Elementary School Journal*, *85*(5), 647-661. https://doi.org/10.1086/461427

Morrow, L. M. (1986). Effects of structural guidance in story retelling on children's dictation of original stories. *Journal of Reading Behavior, 18*(2), 135-152. https://doi.org/10.1080/10862968609547561

Morrow, L. M. (1996). Story Retelling: A discussion strategy to develop and assess comprehension. In Gambrell, L. B. & Almasi, 1. F. (Eds.), Lively Discussions! Fostering engaged reading (pp. 265-285). Newark, DE: International Reading Association.

Muñoz, M. L., Gillam, R. B., Peña, E. D., & Gulley-Faehnle, A. (2003). Measures of language development in fictional narratives of Latino children. Language, Speech, and Hearing Services in Schools, 34(4), 332–342. https://doi.org/10.1044/0161-1461(2003/027)

Mvududu, N. H., & Sink, C. (2013). Factor analysis in counseling research and practice. Counseling Outcome Research and Evaluation, 4(2), 75-98. https://doi.org/10.1177/2150137813494766

National Association for the Education of Young Children Position Statement on Developmentally Appropriate Practice in Programs for 4- and 5-Year-Olds. (1986). *Young Children, 41*(6), 20-30.

Nelson, K. (Ed.). (2006). *Narratives from the crib*. Harvard University Press.

Nice, M. (1925). Length of sentences as a criterion of a child's progress in speech. Journal of Educational Psychology, 16, 370–379. https://doi.org/10.1037/h0073259

Nippold, M. A., Hesketh, L. J., Duthie, J. K., & Mansfield, T. C. (2005). Conversational versus expository discourse: A study of syntactic development in children, adolescents, and adults. Journal of Speech, Language, and Hearing Research, 48, 1048–1064. https://doi.org/10.1044/1092-4388(2005/073)

Nunnally, J. C. (1978). Psychometric Theory (2nd ed.). New York: McGraw-Hill.

Oosthuizen, H., & Southwood, F. (2009). Methodological issues in the calculation of mean length of utterance (MLU). *South African Journal of Communication Disorders*, *56*(1), 76-87.

Owen, A. J., & Leonard, L. B. (2002). Lexical diversity in the spontaneous speech of children with specific language impairment. https://doi.org/10.1044/1092-4388(2002/075)

Padilla, A. M., & Liebman, E. (1975). Language acquisition in the bilingual child. *Bilingual Review/La Revista Bilingüe*, *2*(1/2), 34-55.

Pakulski, L. A., & Kaderavek, J. N. (2001). Narrative Production by Children Who Are Deaf or Hard of Hearing: The Effect of Role-Play. *Volta Review*, *103*(3).

Pallant, J. (2007). SPSS survival manual, 3rd. *Edition. McGrath Hill*, *15*, 361-371.

Pallant, J. (2010). SPSS survival manual: a step by step guide to data analysis using SPSS.

Parker, M. D., & Brorson, K. (2005). A comparative study between mean length of utterance in morphemes (MLUm) and mean length of utterance in words (MLUw). *First language*, *25*(3), 365-376. https://doi.org/10.1177/0142723705059114

Park, C. U., & Kim, H. J. (2015). Measurement of inter-rater reliability in systematic review. *Hanyang Medical Reviews*, *35*(1), 44-49.

Pauls, L. J., & Archibald, L. M. (2021). Cognitive and linguistic effects of narrative-based language intervention in children with Developmental Language Disorder. *Autism & Developmental Language Impairments*, *6*, 1-19. https://doi.org/10.1177/23969415211015867

Petersen, D. B., Gillam, S. L., & Gillam, R. B. (2008). Emerging procedures in narrative assessment: The index of narrative complexity. *Topics in language disorders*, *28*(2), 115-130. https://doi.org/10.1097/01.TLD.0000318933.46925.86

Pezold, M. J., Imgrund, C. M., & Storkel, H. L. (2020). Using computer programs for language sample analysis. *Language, Speech, and Hearing Services in Schools*, *51*(1), 103-114. https://doi.org/10.1044/2019_LSHSS-18-0148

Piaget, J. (1950). The psychology of intelligence. London: Routledge & Kegan Paul

Piaget, J. (1959) . Language and thought. New York: Humanities Press.

Pisani, L., Borisova, I., & Dowd, A. J. (2015). International development and early learning assessment technical working paper. *Save the Children*.

Preece, A. (1987). The range of narrative forms conversationally produced by young children. Journal of Child Language, 14, 353–373. https://doi.org/10.1017/S0305000900012976

Puth, M. T., Neuhäuser, M., & Ruxton, G. D. (2014). Effective use of Pearson's product–moment correlation coefficient. *Animal behaviour*, *93*, 183-189. https://doi.org/10.1016/j.anbehav.2014.05.003

Ranti, A. (2015). Mean length utterance of children morphological development. In *Proceedings of the 1st National Conference on English Language Teaching*, 96-109.

Reed, D. K., & Vaughn, S. (2012). Retell as an indicator of reading comprehension. *Scientific studies of reading*, *16*(3), 187-217. https://doi.org/10.1080/10888438.2010.538780

Reese, E., Suggate, S., Long, J., & Schaughency, E. (2010). Children’s oral narrative and reading skills in the first 3 years of reading instruction. Reading and Writing, 23(6), 627-644. https://doi.org/10.1007/s11145-009-9175-9

Renfrew C (1997) The Renfrew language scales: Action picture test. Winslow Press.

Rice, M. L., Smolik, F., Perpich, D., Thompson, T., Rytting, N., & Blossom, M. (2010). Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments. https://doi.org/10.1044/1092-4388(2009/08-0183)

Robson, C. (2011). *Real World Research: A Resource for Users of Social Research Methods in Applied Settings*. Chichester, England: J. Wiley & Sons.

Roth, F. P., Speece, D. L., & Cooper, D. H. (2002). A longitudinal analysis of the connection between oral language and early reading. *The Journal of Educational Research*, *95*(5), 259-272. https://doi.org/10.1080/00220670209596600

Roulstone, S., Law, J., Rush, R., Clegg, J., & Peters, T. (2011). Investigating the role of language in children's early educational outcomes.

Rummel, R. J. (1988). *Applied factor analysis*. Northwestern University Press.

Sabilla, S. I., Sarno, R., & Triyana, K. (2019). Optimizing threshold using pearson correlation for selecting features of electronic nose signals. *Int. J. Intell. Eng. Syst*, *12*(6), 81-90. https://doi.org/10.22266/ijies2019.1231.08

Schafer, J. L. (1999). Multiple imputation: a primer. *Statistical methods in medical research*, *8*(1), 3-15. https://doi.org/10.1177/096228029900800102

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia*, *126*(5), 1763-1768. https://doi.org/10.1213/ANE.0000000000002864

Scott, C., & Stokes, S. (1995). Measures of syntax in school-age children and adolescents. Language, Speech, and Hearing Services in the Schools, 26, 309–319. https://doi.org/10.1044/0161-1461.2604.309

Sekaran, U. (2003). Research method for business: A skill building approach. 4th edition, John Riley & Sons Inc.

Simmons, D. C., & Kameenui, E. J. (1998). *What reading research tells us about children with diverse learning needs: Bases and basics*. Routledge.

Smith, E. (2008). Pitfalls and promises: The use of secondary data analysis in educational research. *British Journal of Educational Studies*, *56*(3), 323-339. https://doi.org/10.1111/j.1467-8527.2008.00405.x

Snow, C. E., Tabors, P. O., Nicholson, P. A., & Kurland, B. F. (1995). SHELL: Oral language and early literacy skills in kindergarten and first-grade children. *Journal of Research in Childhood education*, *10*(1), 37-48. https://doi.org/10.1080/02568549509594686

Spear-Swerling L. (2013). Nonstandard Assessment. In: Volkmar F.R. (eds) Encyclopedia of Autism Spectrum Disorders. Springer, New York, NY.

Spencer, T. D., & Slocum, T. A. (2010). The effect of narrative intervention on story retelling and personal story generation skills of preschoolers with risk factors and narrative language delays. Journal of Early Intervention, 32(3), 178-199. https://doi.org/10.1177/1053815110379124

Strong, C. J. (1998). The Strong narrative assessuent procedure. Ean Claire, WI: Thinking Publication.

Szabo, V., & Strang, V. R. (1997). Secondary analysis of qualitative data. *Advances in nursing science*, *20*(2), 66-74.

Templin, M. (1957). Certain language skills in children: Their development and interrelationships. Minneapolis: University of Minnesota Press

Thompson, B. (2004). Exploratory and confirmatory factor analysis: Understanding concepts and applications. *Washington, DC*, *10694*. https://doi.org/10.1037/10694-000

To, Y., & Mandracchia, J. (2019). *Learn About Hierarchical Linear Regression in SPSS With Data From Prison Inmates*. SAGE Publications, Limited. https://doi.org/10.4135/9781526494290

Tomblin, J. B., & Zhang, X. (2006). The dimensionality of language ability in school-age children. Journal of Speech, Language, and Hearing Research, 49, 1193–1208. https://doi.org/10.1044/1092-4388(2006/086)

Tripathy, J. P. (2013). Secondary data analysis: Ethical issues and challenges. *Iranian journal of public health*, *42*(12), 1478.

Tsimpli, I. M., Agathopoulou, E., Andreou, M., & Masoura, E. (2014). Narrative production, bilingualism and working memory capacity: A study of Greek-German bilingual children. In *Selected Papers of the 11th International Conference on Greek Linguistics* (No. IKEECONF-2015-841, pp. 1730-1742). Aristotle University of Thessaloniki.

Uyanık, G. K., & Güler, N. (2013). A study on multiple linear regression analysis. *Procedia-Social and Behavioral Sciences*, *106*, 234-240. https://doi.org/10.1016/j.sbspro.2013.12.027

van Dijk, T. A., & Kintsch, W. (1983). Strategies of discourse comprehension. Academic Press. https://doi.org/10.2307/415483

Vartanian, T. (2011). *Secondary data analysis* (Oxford scholarship online). New York; Oxford: Oxford University Press.

West, G., Snowling, M. J., Lervåg, A., Buchanan-Worster, E., Duta, M., Hall, A., ... & Hulme, C. (2021). Early language screening and intervention can be delivered successfully at scale: evidence from a cluster randomized controlled trial. *Journal of Child Psychology and Psychiatry*. https://doi.org/10.1111/jcpp.13415

Wiig, E. H., Secord, W. A., & Semel, E. (2004). Clinical Evaluation of Language Fundamentals – Preschool 2. San Antonio, TX: Harcourt/PsychCorp.

Wilson, J., Fang, C., Rollins, J., Valadez, D. (2016). An urgent challenge enhancing academic speaking opportunities for English learners. Multicultural Education, 52-54

Yong, A.G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on Exploratory Factor Analysis. Tutorials in Quantitative Methods of Psychology, 9(2), 79-94. https://doi.org/10.20982/tqmp.09.2.p079

Appendices

Appendix A: NELI Individual In-depth Assessment Record Form

UNIVERSITY OF OXFORD

NELI-N Trial - Record Form

UPN (Unique Pupil Number):	_ _ _ _ _ _ _ _ _ _ _ _ _
School Name:	
School Postcode:	_ _ _ _ _ _ _
Child's Initials:	_ _
Child's First Name:	
Date of Birth:	_ _ _ _ _ _ _ _ _
Child's Gender:	Boy / Girl / Undisclosed
English as an Additional Language:	Yes / No / Not Known
School Year:	Nursery / Reception / Year 1 / Year 2 / Not Known
Tester Name:	
Date of Testing:	_ _ _ _ _ _ _ _ _

Comments/Observations:

PLEASE RECORD ALL TASKS

TEST	ADMINISTERED	SCORED	ENTERED
CELF Preschool II: Expressive Vocabulary			
CELF Preschool II: Recalling Sentences			
HTKS Task			
Renfrew Action Picture Task			
Narrative Task			
Intervention Target Vocabulary			

University of Oxford NELI-N PROJECT Pre- and Post-test Record Sheet (2019 / 2020)

Appendix B: The number of children in each school

School no.	No. of selected children
1	10
2	8
3	10
4	8
5	9
6	9
7	10
8	9
9	9
10	10
11	7
12	9
13	9
14	9
15	19
16	10
17	10
18	9
19	10
20	6
21	10
22	9
23	10
24	9
25	9
26	9

27	8
28	10
29	9
30	10
31	9
32	8
33	10
34	9
35	10
36	9
37	9
38	9
39	9
40	19
41	9
42	10
43	9
44	9
45	10
46	9
47	10
48	10
49	8
50	10
51	9
52	10
53	8
54	10

55	10
56	8
57	10
58	10
59	20
60	9
61	9
62	9

Appendix C: RAPT 10 picture cards

Appendix D: Oral Narrative Retelling Task

D.1 Bus Story (3 pictures)

D.2 Story Transcripts

It is a sunny day.

The boy is going to visit his friend.

He needs to get on the bus but he is late.

He sees the bus at the bus stop and runs to catch it.

He tries to run as fast as he can.

But the bus driver as away before he can get to the bus stop.

So he misses it and has to wait for the next one.

The boy is very cross.

Suddenly it starts to rain.

The boy forgot to bring his umbrella and he is not even wearing a coat.

Soon the boy is very wet. “I won’t be late again” says the boy.

D.3 Instructions

(1) The examiner puts the picture cards out in order on the table in front of the child
(2) The examiner says “I am going to tell you a story. You need to listen very carefully because when I have finished you are going to tell the story to Little Neli the elephant.”
(3) The examiner tells the story, pointing to the relevant parts of the corresponding pictures as they do so.
(4) When the examiner has finished, it is the child’s turn to retell the story to Neli.
(5) Indirect prompts may be provided, such as “What happened first?”, “What happened next?” or “And then?”
(6) The examiner records the child’s story recall verbatim, preserving any grammatical mistakes (e.g., ‘she goed’ rather than ‘she went’).
(7) The examiner must allow the child to give as much detail as possible before prompting them.
(8) Once the child has finished telling the story, the examiner can ask, “Can you remember anything else about the story?”
(9) If child is reluctant to speak, the examiner should point to the first picture and say, “What happened in this picture?” the examiner should do this for each of the three pictures.

Appendix E: Traditional Methods of Calculating MLU based on Brown (Brown, 1973:54)

Include	Exclude
Stuttering or repetitive words or utterances are included, but count once, in the most complete form produced (Oosthuizen & Southwood, 2009:85; Ezeizabarrena & Garcia Fernandez, 2018; Ranti, 2015:103).	Fillers such as *mm*, *oh* and the equivalents like *um* and *uh* are excluded, but ***no***, ***yeah***, and *hi* are counted (Oosthuizen & Southwood, 2009:85; Ranti, 2015:103)
Only completely transcribed utterances are used; no blanks are included. Parts of utterances are utilised, with parentheses indicating doubtful transcription (Ranti, 2015:103)	Utterances that are completely or partially unintelligible (Oosthuizen & Southwood, 2009:85)
Compound words (two or more free morphemes), proper names, and ritualised reduplications are all counted as single words. For example, "*rackety-boom*", "*choo-choo*", and "*night-night*" (Ranti, 2015:103)	

Appendix F: MOU Research Agreement – Inter-rater Reliability

MEMORANDUM OF UNDERSTANDING BETWEEN [YUQING DING] AT UNIVERSITY OF OXFORD AND [NAME ANOYMISED]

Background

(A) This Memorandum sets out the basis upon which the researcher [YUQING DING] at the University of Oxford (CDE pathway) and [NAME ANOYMISED] have agreed to explore the research topic on "The relationship between children's narrative retelling and other aspects of language skills"

(B) By entering into this Memorandum, [NAME ANOYMISED] has indicated the intention to cooperate with [YUQING DING] completing a series of scorings following the same criteria

(C) [NAME ANOYMISED] agrees with the confidentiality of the research and data. The data will be purely used for inter-rater reliability analysis.

Signature: [NAME ANOYMISED]
Date: 02/07/2021

Appendix G: Output Tables with Information of Little's MCAR Test

Little's MCAR test includes the EM Means, EM Covariances, and EM

EM Means[a]

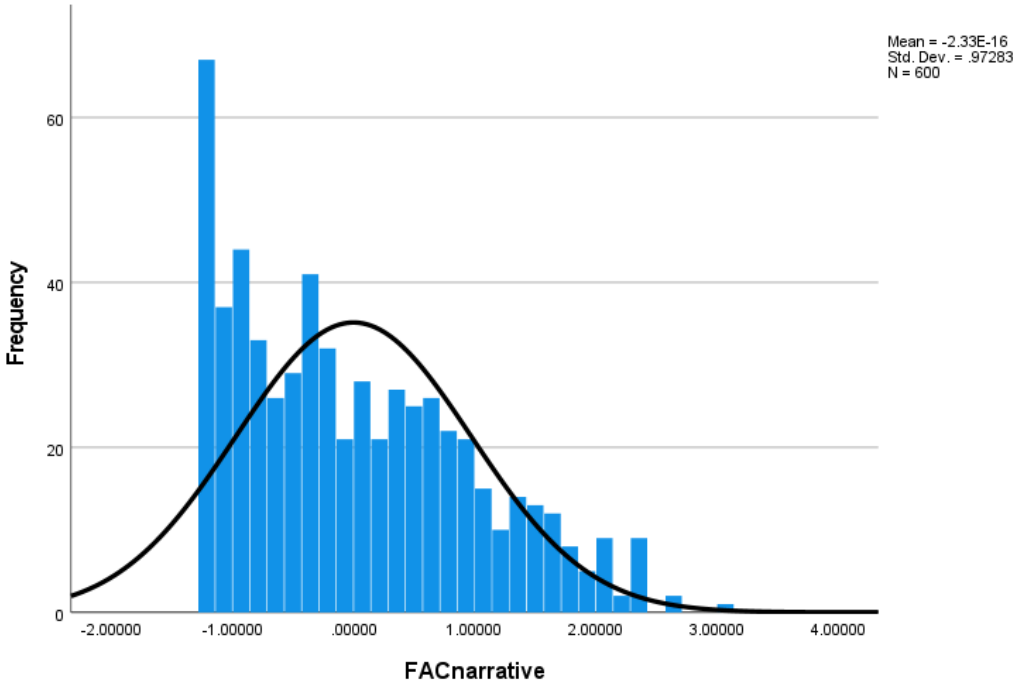
age	SchoolNo	CELFev	CELFrs	APT_info	APT_gram	EVtotal	RVtotal	SRtotal	LCtotal	NT_TotWord	NT_Conjunctions	NT_DiffConjunctions
42.81	31.92	8.46	5.81	14.900	9.654	6.04	10.04	3.70	2.30	19.73	1.52	.70

a. Little's MCAR test: Chi-Square = 24.460, DF = 46, Sig. = .996

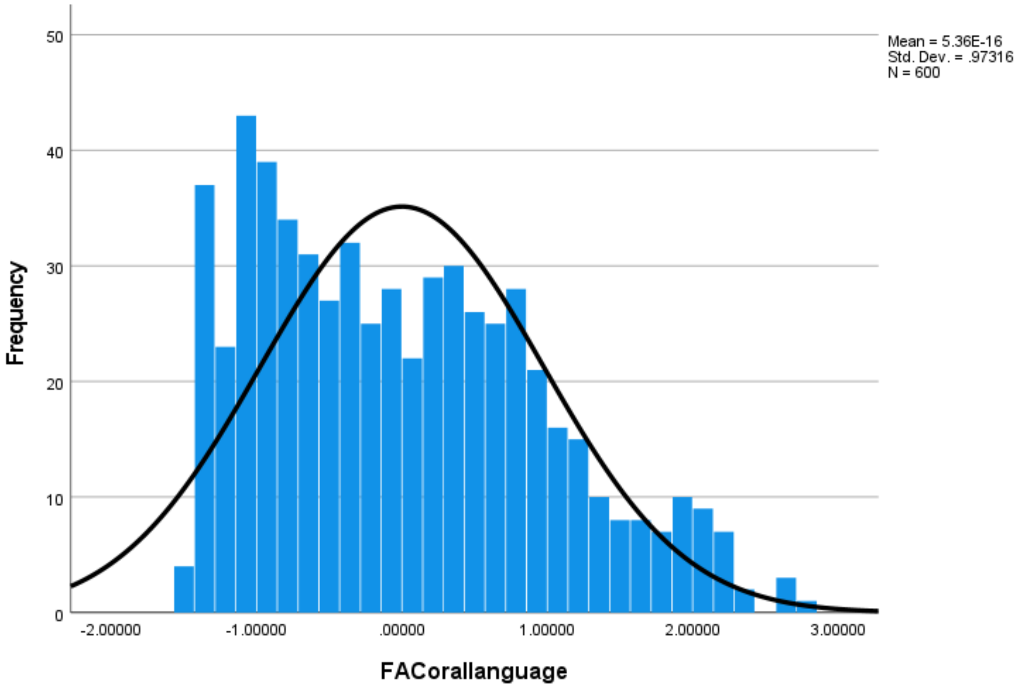
Chi-square	Degrees of freedom	P-value
24.460	46	.996

Appendix H: Histograms for FAC_narrative retelling and FAC_oral language skills

H1. Histogram of FAC_narrative retelling



H2. Histogram for FAC_oral language skills



Appendix I: Hierarchical Linear Regression Assumption Tests

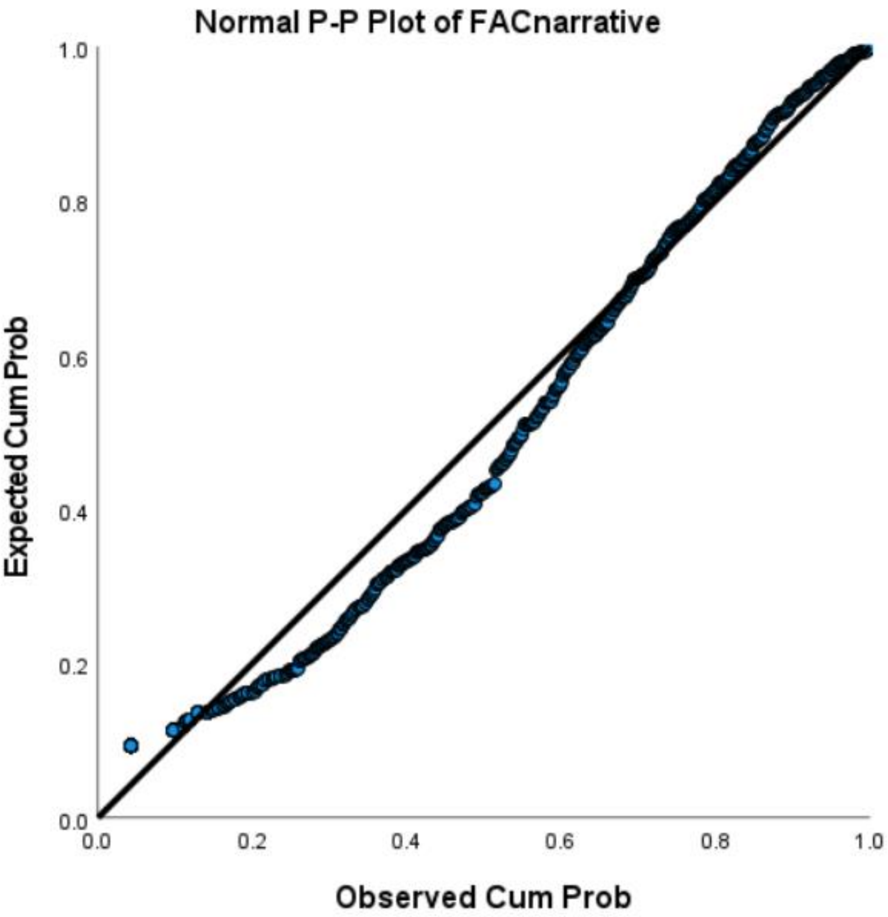


Figure 4.1: Normal P-P plot of regression standardised residual for narrative retelling (DV)

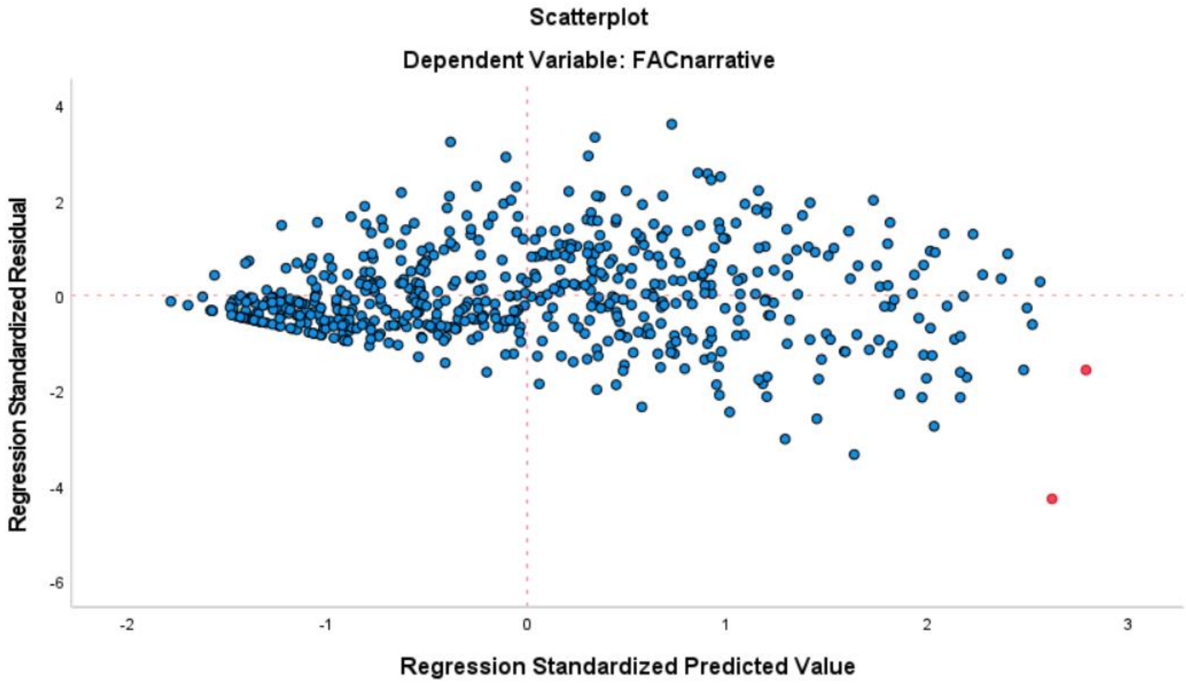


Figure 4.2: The residual scatter plot showing homoscedasticity for the observed and predicted values

Model		Unstandardised coefficients		Standardised coefficients	t	Sig.	Collinearity Statistics	
		B	Std.Error	Beta			Tolerance	VIF
1	(constant)	-2.134	.453		-4.712	.000		
	Gender	.215	.075	.111	2.890	.004	1.000	1.000
	EAL	.600	.075	.308	8.046	.000	1.000	1.000
	Age	.040	.010	.147	3.848	.000	.999	1.001
	SchoolNo.	.000	.002	.008	.209	.834	.999	1.001
2	(constant)	.608	.379		1.603	.110		
	Gender	.131	.058	.067	2.257	.024	.994	1.006
	EAL	-.088	.068	-.045	-1.296	.196	.734	1.362
	Age	-.013	.008	-.050	-1.587	.113	.898	1.114
	SchoolNo.	-.001	.002	-.027	-.898	.370	.996	1.004
	FAC_oral langauge	.714	0.36	.715	19.682	.000	.673	1.486

Table 4.1: Coefficient values

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	*.360*[a]	.129	.123	.91080842	
2	*.688*[b]	.473	.469	.70919259	**1.954**

a. Predictors: (constant), SchoolNo, EAL, Gender, Age

b. Predictors: (constant), SchoolNo, EAL, Gender, Age, FAC_oral langauge

Table 4.2: Model summary

Model		Sum of squares	df	Mean Square	F	Sig.
1	**Regression**	73.299	4	18.325	22.089	$.000^{a}$
	Residual	493.595	595	.830		
	Total	566.894	599			
2	**Regression**	268.140	5	53.628	106.626	$.000^{b}$
	Residual	298.755	594	.503		
	Total	566.894	599			

a. Predictors: (constant), SchoolNo, EAL, Gender, Age
b. b. Predictors: (constant), SchoolNo, EAL, Gender, Age, FAC_oral langauge

F(regression df, residual df) = F-ratio, *p* = Sig.

Table 4.3: ANOVA output

Model		Unstandardised Coefficient		Standardised Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(constant)	-2.134	.453		-4.712	.000
	EAL	.600	.075	.308	8.046	.000
	Gender	.215	.075	.111	2.890	.004
	Age	.040	.010	.147	3.848	.000
	SchoolNo.	.000	.002	.008	.209	.834
2	(constant)	.608	.379		1.603	.110
	EAL	-.088	.068	-.045	-1.296	.196
	Gender	.131	.058	.067	2.257	.024
	Age	-.013	.008	-.050	-1.587	.113
	SchoolNo.	-.001	.002	-.027	-.898	.370
	FAC_orallanguage	.714	.036	.715	19.682	.000

Table 4.4: Coefficients in two-block hierarchical linear regression

	EAL	Gender	Age	SchoolNo	FAC_oral language skill
FAC_narrative retelling	.309	.111	.148	.001	.681

Table 4.5: Correlations between narrative retelling and other demographic predictors