An exploration of the relationship between the EGPS Test and writing outcomes.

James Canniford (2021)
UCL – Institute of Education
Doctor of Education

Declaration

I, James Canniford, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed:

Word Count: 44,808

(Exclusive of abstract, impact statement, appendices and list of references)

Abstract

Background. The English Grammar, Punctuation and Spelling (EGPS) Test is a statutory assessment of writing convention skills in Year 6 in England. The relationship between the EGPS Test and children's writing outcomes has not been examined empirically. This thesis aimed to establish whether writing convention skills assessed by the EGPS Test demonstrate a relationship with writing quality, accuracy, productivity, and complexity, and whether the EGPS Test can be used to identify struggling writers.

Method. A sample of N = 210 was drawn from English primary schools. Participants completed the EGPS 2018 Test, alongside tasks that assessed writing quality at word, sentence, and paragraph levels. Writing at the microstructural level was assessed by measures of productivity, accuracy, and complexity. In addition, pupils completed a measure of spelling ability, handwriting fluency, and reading comprehension. Assessments were completed as whole class task and administered by the teacher to adhere with COVID regulations at the time.

Results. The EGPS Test correlated strongly with writing quality. Domains of the EGPS Paper 1 loaded on a single factor, theorised as EGPS Writing Conventions (EGPS-WC). EGPS-WC correlated strongly with the composite Written Expression Subtest (WES) and paragraph-level writing, and moderately with sentence-level writing. Sentence-level writing was predicted by EGPS-WC performance; paragraph-level and overall quality (WES) of writing were predicted by EGPS-WC and spelling ability. Path analyses demonstrated that EGPS-WC contributed both directly and indirectly to overall writing quality. Multiple regression analyses identified that spelling and EGPS-WC predicted writing accuracy. Finally, The EGPS 2018 Test offered excellent predictive validity for struggling writers.

Conclusions. EGPS is theorised as a distal (indirect) measure of writing. A theoretical rationale for the role of EGPS-WC as part of the text generation process is proposed, emphasising the importance of freeing cognitive resources through the development of writing convention knowledge.

[Word Count – 296]

Impact Statement

As a requirement of the Doctor of Education programme, an impact statement is now offered. The aim of this statement is to explore how this research may be beneficial beyond the course of study in both an academic and professional context. A summary of professional impact is made from the conclusions of the final chapter of the thesis, and issues of academic contribution and wider dissemination are presented.

Professional Impact

When embarking on the Doctor of Education course, I was keen to explore the wider implications of the teaching of grammar in the primary classroom. My proposal for the research was rooted in the challenge from wider professionals and the media to remove the English Grammar Punctuation and Spelling Test from English primary schools; I was unsure of the evidence base to support its removal. It was through the structure of this doctorate that I was able to rigorously examine the EGPS test.

As a result of this research, several procedural and policy changes have taken place within my own institutions around writing and assessment. These are as follows:

- Exploring and directly teaching genre in writing lessons much more rigorously. *As a result of the literature review in this study, we have re-considered the role that knowledge of genre and academic register play in children's writing. In turn, this thesis has directly impacted the daily teaching of writing for thousands of children in London.*

- The EGPS is used as an early warning for struggling writers. *We are now giving much greater weighting to EGPS scores as a means of early warning for children who need writing interventions.*

- Importance of developing a declarative body of knowledge of language convention skills. *We now as an institution have raised our prioritisation of ensuring that children have secure body of writing convention knowledge to better facilitate the text generation process.*

Potential Academic Contribution

I consider that this thesis contributes to knowledge in several ways:

- Data from this study demonstrate that the EGPS test is a reliable and valid way of measuring writing abilities in Key Stage 2
- Data from this study indicate that the EGPS test may have merit as an early warning for struggling writers
- Data from this study demonstrate that the EGPS Paper 2 offers convergent validity with a standardised spelling test
- Data from this study indicate that writing convention skills may constrain working memory in text generation processes
- Data from this study suggest that 'EGPS Writing Convention Skills' are a distal factor in writing

Dissemination

The findings of this thesis are to be shared with the institutions where data was collected through a summary letter to schools. Findings from this study will also be disseminated in the following ways:

- As a summary paper for the Southwest London English Hubs network
- In conversations with leaders and practitioners considering the future of the EGPS Tests within their settings
- In teacher training programmes for writing in primary schools
- Conferences and papers exploring the role of writing convention skills in the writing process

[500 Words]

Contents

Figures

Tables

Candidate Statement

This reflective statement is an overview of my journey through my Doctor of Education (EdD) programme, from pre-doctoral studies in the Post Graduate Diploma in Social Science Research Methods, through the taught modules of the EdD, and the final Thesis. This statement begins with an overview of my course of study, followed by a discussion around my learning within the taught modules and the institution-focused study (IFS). Finally, this statement reflects on my personal and professional journey throughout the thesis stage and summarises my achievements in line with the aims of the EdD programme handbook.

Programme overview

This section offers a summary of the entire EdD programme. The following table offers the modules that I have engaged in throughout the course of study, offering the titles and grades of each of the projects I have engaged in.

EdD Module	Title	Grade
SSRM Module 1 Developing Research Questions	What effect does grammatical pedagogical content knowledge have on primary school teacher's beliefs about the teaching of grammar?	A
SSRM Module 2 Methods of Investigation	The Poor Writers Research Project: Research Placement	A
SSRM Module 3 Designing a Research Study	Belief is Knowledge: A comparative study of teachers' grammar knowledge and beliefs about the teaching of grammar.	A
SSRM Module 4 Developing a Research Proposal	The SPAG Test – Purposeful and Enjoyable? An exploration of the purpose, validity and enjoyability of the KS2 SPAG Test.	A
Foundations of Professionalism (FoP)	Shaping a professional identity: W*hat is the role of a 'Director' in a newly formed multi-academy trust?*	A
Institution Focused Study (IFS)	The EGPS Test: Purposeful and valid? An exploration of the purpose and validity of the KS2 EGPS Test.	A

Thesis	The EGPS Test: An exploration of the relationship between the EGPS Test and writing outcomes.	

The UCL IOE Research Community

One of the greatest advantages of engaging in this doctoral programme has been the membership of an active and supportive academic research community within UCL IOE. A fellow student-colleague of mine surmised his experience of the EdD programme in a way that perfectly encapsulates the EdD experience. Quoting what is widely attributed to C.S. Lewis, he states:

Isn't it funny that day by day nothing changes, but when you look back,
everything is different. (source unknown)

I now look back in this regard, critically reflecting on my experience of the EdD programme and its impact upon my academic and professional development.

From absentee to academic (pre-EdD)

As someone who was identified at a young age as being intelligent, yet never 'applying themself', I had no expectation that I would ever make it through to higher education, let alone post-graduate and doctoral study. As a disillusioned learner who understood the content of courses yet who could never seem to recall it in exams, I began to skip classes in favour of earning money. After complete failure in my A-level exams, a series of unexpected events which included stints as a nature ranger in New York, a 'dinner-lady' (as described by the children), and an early-years teaching assistant, led to me being offered a place as a trainee teacher on a trial undergraduate work-based scheme – to the surprise of my friends and family.

Whilst on this initial teacher training course, the care and supervision of my Mentor, Wendy Hoggett, led to the diagnosis of my severe working memory difficulties which had plagued my exams throughout my life. I finished with first class honours and an academic award for the highest grade of the cohort.

After spending several years in the primary classroom and engaging in further successful post graduate study domestically and internationally, I had the pleasure of working with a colleague-mentor, David Porritt, who was a student on the EdD programme at IOE. With

a passion at the time for coaching, I was keen to explore this further through research. However, without recent research methods training, I felt ill-prepared to apply for doctoral study. After discussion with Dr Sue Taylor, the Social Science Research Methods course was the agreed first-step towards an Education Doctorate.

PGDip Social Science Research Methods (SSRM)

This course, a pre-doctoral preparation which could forsake the first modules of the taught doctorate, was one of the most formative experiences of the entire EdD journey. The blended learning approach challenged me and my fellow learners to take ownership of our learning around research methods. It was at this point that I began to realise the distinction between pedagogy and andragogy and its implications. I grew to thoroughly enjoy hosting sessions on quantitative methods, despite never having taken a statistics course in my life. It taught me the importance of self-discipline and self-study when approaching entire new fields of academia, as well as the value of sharing and discussing as a means of consolidating learning.

A consistent theme of this foundation programme was research questions and the importance of their accuracy, clarity, and operationalisation. While I knew I was interested in coaching trainee teachers, I did not have the breadth of knowledge in the field to offer any particularly insightful contribution. For ease, I decided to complete my first module on another interest of mine: grammar teaching. This less established area of study began to invite some interesting challenges.

The first module of the SSRM course required a literature review. This provided an opportunity to explore this increasingly prominent area of teaching and assessment within the primary curriculum and which also highlighted the dearth of empirical research around the study of grammar in English primary schools. Seeing gaps in the body of knowledge empowered me to be more critical and reflective on what my next steps as a researcher should be.

A small-scale research project was the next module of the SSRM course. I decided to begin to explore this gap in knowledge with a research project around *teacher's knowledge of and beliefs about grammar*. As my first piece of quantitative research, I was unexpectedly enamoured with the importance that simple numbers could bring and the implications of these for the profession.

Alongside this module, I was required to undertake a research placement. As part of this, I was tasked with a literature review related to children's writing with Professor Julie Dockrell – now my principal supervisor. After discussing Julie's field of interest, struggling writers, and my field of interest, teacher's beliefs about grammar, we both agreed on the need for someone to empirically test the relationship between the EGPS Test and writing. This subsequently led to my final project – a research proposal based on an exploration of the EGPS Test.

Foundations of Professionalism.

This module infuriated me more than any study I have ever undertaken. At first this was because the content seemed so utterly academic and irrelevant. However, by the end of the course, this module was so entirely relevant that I wished I had undertaken studies on professionalism much earlier in my career. Furthermore, I am now of the opinion that my fellow teachers and I are only quasi-professionals – an opinion I would have found offensive in previous years and a framing which accounts for such discord around expectations upon teachers and workload. This opportunity for professional growth has been formative in my recent career development. In particular, the importance of reflexivity has proven beneficial in every aspect of my role as a leader in school – particularly when honestly and openly exploring impact of my actions.

Institute Focused Study (IFS)

The IFS was a formative stage in my research journey as it forced me to consider methodology more rigorously than ever before. Until this point, I had maintained surface-level engagement with research methods, seeing it as a theoretical hoop to jump through, rather than a practical operationalisation of theory. However, under the pupillage of Prof. Julie Dockrell, I have come to understand the value of methodology and methods that are informed by the literature. I also came to realise at this stage that I could contribute more than just 'playing at research', as my study began to explore issues that had not previously been examined. I was being asked to present my research at conferences and professional development courses as it was, to my surprise, interesting and novel.

My newly developed understanding of paradigm and its related concepts has also profoundly affected my approach to strategic and organisational change within school leadership. In understanding that not all individuals view the world or 'truth' in the same way,

it has deeply impacted the approaches I use when presenting any new initiatives. Offering theoretical frameworks behind any strategic change enables colleagues to understand where decision making has come from, even if they hold vastly different world views.

Thesis Stage

My thesis was written in the context of one of the most challenging periods of my life. Already situated in the context of a special measures school, the COVID-19 pandemic made my research almost impossible. However, with the support of my supervisors, I was able to push forward, making decisions informed by the literature on how to make the study work in this brave new world of lockdown. While not as comprehensive a study as I had initially anticipated, I was able to design a study which offered rigour and a contribution to knowledge within the confines of school COVID-19 bubbles.

During the upgrade process, I was complemented on the fact that this was one of the most sophisticated EdD proposals that the examiner had ever read. However, when engaging in the data analysis, the hard reality of this statement became apparent. A difficult and rapid learning journey took place around the analytical methods that I had chosen. This emphasised to me the importance of the Dunning-Kruger Effect, whereby my lack of knowledge led to an overinflated level of confidence in my choice in methods. However, after a huge amount of personal study, and challenge from my supervisors, I overcame these hurdles and was able to engage in rigorous analysis of the data. What became apparent by the end was that I had clear justification on why I had made decisions around methods and analysis, as opposed to my previous position of stating that 'this is what I saw another researcher do'. This is one of the greatest lessons that I have learnt, and for this I am truly grateful to my supervisors.

The Future

I have been stretched and developed far beyond I would ever have expected through engagement in this professional doctorate. During this course of study, I have progressed from middle leader to Headteacher, and from a relatively uninformed teacher to a reflective practitioner and leader. I have no doubt that I will continue to engage in research around writing, exploring the importance of writing assessment. I have had the pleasure of presenting my research at an international conference, and I now look forward to disseminating my research to a wider audience.

Overall Summary

The UCL Institute of Education EdD Handbook 2020-21 states:

> The Doctor in Education (EdD) aims, through a combination of taught courses and supervised personal research, to develop research skills and knowledge, which together with reflection on professional experience enable participants to interpret, evaluate, conduct, and disseminate research that is relevant to, and has potential impact on, their professional development and workplace practice, as well as meeting the requirements of rigour and originality expected of a doctorate. (UCL, 2020)

In engaging in this Doctor of Education programme, I am wholeheartedly convinced that the course of study has enabled me to fulfil all the aims set out in the handbook. I have far exceeded my own expectations around conducting and disseminating my research, and I have found the wider professional learning essential to my workplace practice and development as a young school leader. I have no doubt that this course of study has changed me for the better, and it will continue to frame my future development in school leadership, policy development, and ongoing research into grammar and writing.

[1986 Words]

Acknowledgements

To my dear grandmother, Phyllis.

For correcting every hand-written text
that I ever dared leave around the house
for grammar, punctuation, and spelling.

CHAPTER 1 : General Introduction

1.1 Personal and Institutional Context

The personal and institutional scope of this thesis is rooted in my own personal interest in grammar. In 2013, I first oversaw the introduction of the English Grammar, Punctuation, and Spelling Test (EGPS Test) within my classroom. As an early career teacher, it became apparent that I was ill-equipped to ensure my students' success in this new statutory assessment of writing conventions (grammar, punctuation, and spelling). Subsequent research regarding teachers' responses to the new curriculum identified that I was not in an isolated struggle (Bell, 2016). On reflection, my own deficit of writing convention knowledge was highlighted when I spent several years living and working in a variety of countries as an L2 language speaker where I was deficient of the knowledge of basic grammatical structures being presented to me in language classes, even when these were explained in my native language. The problem was not the meta-language, but my own lack of declarative understanding around grammar.

Upon returning to the UK, I began to make use of my increasing knowledge of writing conventions inform my pedagogy. Anecdotally, I found that when children were taught knowledge about language in context (see Myhill et al., 2010), their writing outcomes improved at a greater rate than in previous cohorts. Children who had not previously understood the role and functions of a *fronted adverbial* – just one element of the new grammar curriculum – were able to use a variety of adverbial details with increasing awareness and control once they had understood the variety of potential semantic functions. Additionally, owing to my more developed understanding of semantic functions in sentences, my subsequent understanding of punctuation and spelling was drastically improved. As I now understood how parts of speech should correctly be combined, I subsequently understood where punctuation such as commas and semi-colons needed to be placed. This increased understanding of grammar also impacted my spelling knowledge, with decisions around words such as advise or advice now becoming a simple decision, as I understood the distinction between a verb and a noun. This understanding enabled me to teach punctuation and spelling with more confidence and authority than ever before. Anecdotally, the impact was noticeable in children's successive scores in EGPS Tests, which rose in my cohorts year on year within my Year 6 classes. Yet, I was not able to confidently attribute the improvement of EGPS scores to increasing knowledge about language beyond a range of other factors. It is quite possible that general improvement

and understanding in how to take the EGPS Test had developed over time. Yet during this period, my interest around writing convention skills, writing assessment, and the additional factors influencing children's writing products began to develop.

1.2 Rationale for the research

While a wide body of research exists regarding the place of grammar within the English National Curriculum, (R. Andrews et al., 2006; R. Andrews, Torgerson, Beverton, Locke, et al., 2004; Braddock, Lloyd-Jones, & Schoer, 1963; Hillocks Jr & Smith, 1991; Myhill & Watson, 2014; Sheard, Chambers, & Elliott, 2012; Weaver, 1996; Wyse, 2001), there remains an absence of research that explores the relationship between the Key Stage 2 EGPS Test and children's writing outcomes. To date, empirical research has only examined the impact of the test on teachers and teaching (Bell 2016; Safford 2016), recognising teacher anxiety which is rooted in fears around being ill -prepared for the delivery of a grammar-heavy curriculum. In the years following the introduction of the EGPS Test, the introduction of meta-linguistic terminology into the curriculum was met with reported dissatisfaction, arguing that grammar was thought to be outdated and unhelpful (Mansell, 2017; Safford, 2016), with disagreement among teachers regarding the extent to which explicit grammar teaching and testing was having a positive impact upon pupils (Safford, 2016) . A collective feeling had arisen among teachers and researchers that the government of the time moved towards a system of only testing grammar as a proxy for writing assessment (Mansell 2017). Interestingly, similar dissatisfaction has not been evident with either spelling or punctuation. This may suggest that these two elements of the EGPS Test are regarded as essential or were already widely established within the primary classroom in England.

Since the removal of the Writing SATs in 2012 owing to high misclassification rates (William, 2000), the EGPS has been commonly used as a proxy for writing assessment. As the only formally tested element of writing, EGPS was brought to the forefront of English teaching in English primary schools. Yet evidence regarding the validity of the EGPS as a proxy for writing is lacking. To date, only one empirical study exploring the relationship between the EGPS Test and writing exists: my Institution Focused Study.

1.3 Link to Institution Focused Study

Within my Institute Focused Study (IFS), I carried out an initial exploration of the relationship between EGPS and writing using a standardised assessment with a sample of Year

6 children (N = 96) from within my own institution (Canniford, 2019). A review of the literature revealed that there was little evidence that grammar offers any meaningful relationship with writing outcomes. However, the findings from my IFS demonstrated an association between EGPS scores and children's scores on a standardised writing assessment, offering evidence of strong convergent validity. However, this study acknowledged a number of limitations.

The first limitation was that the sample distribution presented a negative skew on both the EGPS Test and standardised writing assessments. This meant that the sample performed comparatively well on both assessments when compared to national norms, calling in to question how representative the sample was against a wider population. Future research should consider a more reparative sample, using a greater sample size that cites OFSTED attainment data to ensure a more representative sample. The next limitation was that the standardised writing assessment assessed only one genre: narrative writing. Within the IFS, it was noted that genre should be a key consideration in the exploration of relationship between the EGPS Test and writing as assessment of a single genre may be likely to fail at capturing the various demands placed on children between differing text types (Scott and Windsor, 2000; Olinghouse and Wilson, 2013). Dockrell et al (2015) argue that differing text types, expository writing for example, can be more demanding than the narrative text that was used in this IFS. Therefore, future research should explore both narrative and expository writing in relation to the EGPS Test. A final limitation of the IFS was that the EGPS Test was explored as a composite assessment, rather than exploring the role of the individual components: grammar, punctuation, and spelling. The role of each of the different skills that underpin writing has been explored to varying degrees. For example, the relationship between spelling and writing is supported by a strong evidence base (see Chapter 3). Therefore, the component skills of the EGPS Test should be explored separately to explore their individual contributions to writing.

This thesis set out to extend that original study, seeking to contribute to the body of knowledge through a more robust and representative study through a larger and more representative sample, by exploring the place of genre in children's writing, and by exploring relationships between the different EGPS Test domains and writing.

The main aim of this thesis was to further explore the relationship between the EGPS Test and children's writing outcomes. Due to the COVID-19 global pandemic, all statutory

assessments at the time of data collection had been halted in English primary schools. This was also set against the already contentious backdrop of the 2017 Education Select Committee on Primary Assessment, which called for the EGPS Test to be made non-statutory in Key Stage 2 (House of Commons, 2017) based upon evidence presented by Myhill, an original author of the statutory grammar curriculum, that the EGPS Test could not be used as a proxy for writingFor teachers, school leaders, and the government to make empirically substantiated responses to the continuation of this test, robust research concerning the relationship between the EGPS Test and children's writing outcomes must be undertaken. This will be of particular concern for school leaders as assessment becomes a driver of curricula; what is being tested is often a precursor to what will be taught. If the recommendations of the Education Select Committee are to be adhered to in future, then decisions regarding the continued use of EGPS testing may soon lie with school leaders. Therefore, robust and current empirical research was undertaken in order to guide decision making about the extent to which the EGPS Tests are of use in the evaluation of children's writing skills.

The framework of this research was grounded in psychological approaches which aimed to both understand writing proficiency and to identify certain required competencies which predict writing competency. This research explored the cognitive factors which are employed in the production of written text, and particularly the role of grammar, punctuation, and spelling in relation to children's written products.

1.4 Overview of the structure of research

This first chapter presents the rationale for research regarding the relationship between the EGPS Test and children's writing outcomes and provides an institutional context for this study.

The review of the literature was then carried out in five parts. Chapter two begins with the relevant literature on writing within primary schools. Then a short discussion on the writing process and writing products is offered. Finally, struggling writers and their differences from typical writers are explored.

Chapter three explores the cognitive models of writing development which, after a critique, offers two theorised distal factors of writing. This chapter ends with a discussion of the various dimensions of written products.

Chapter four offers an overview of writing assessment, exploring the various means of scoring writing. This chapter includes a discussion of genre and different means of collecting writing samples.

Chapter five then addresses one potential form of writing assessment: The EGPS Test. The historical context of the EGPS Test is presented, followed by an exploration of the test content. This chapter then proceeds to critique the EGPS Test and explains differing opinions on whether the EGPS Test can be used as a dependable proxy for writing.

Chapter six further unpacks issues of validity and reliability in regard to the EGPS Test. Validity, reliability, and associated terms are discussed in the context of the EGPS Test.

Chapter seven makes the aims of the study explicit, offering four main research questions. Overall research hypotheses are then presented.

Chapter eight discusses methods. This begins with the research strategy and design, followed by detailed exploration of the sample; choice of methods; analysis; and concludes with ethical considerations.

Chapter nine offers in-depth quantitative analysis of data gathered within this study. Data are presented in seven sections: section one offers descriptive statistics and distributions; sections two to seven address each of the research questions.

Chapter ten discusses the data presented and offers theoretical insights regarding the statistically significant relationships that have been found. Theoretical and practical insights, and implications for professional practice are shared before limitations of this study are discussed and suggestions for future research are offered.

Chapter eleven concludes the study by summarising the findings from this research project. Final conclusions are drawn before References and Appendices are offered.

CHAPTER 2: Writing - What is it and does it matter?

2.1 Introduction

This chapter explores writing in primary schools. The challenges faced around writing in English primary schools are discussed and issues around struggling writers are presented. This chapter then considers writing as a process and writing as a product.

2.2 Writing in Primary Schools

The capacity to produce writing is a transferable skill, which is integral to academic achievement, employment, and communication more widely (Dockrell, Connelly, Walter, & Critten, 2015). While the writing process is complex (Harmey, 2021), written products are the primary vehicle for demonstrating knowledge, both in education and everyday life (Graham & Harris, 2005). Writing serves as a means of personal meaning-making and as a construct to support learning (Bangert-Drowns, Hurley, & Wilkinson, 2004). In addition, writing can act as a consolidation process, both in relation to reading comprehension (Hebert, Simpson, & Graham, 2013) and in developing specific subject content knowledge (Bangert-Drowns et al., 2004). The rewards for learning to write remain rich: writing is a skill which later allows for engagement in society and contribution to the economy (Dunsmuir et al., 2015). While the implications for the EGPS test and its relationship regarding writing are limited to the English context, the relationship between grammar knowledge, to which the EGPS may potentially offer insight, and its relationship with writing may be universal to the writing process and the wider instructional context.

Prior to the withdrawal of the KS2 writing test and subsequent introduction of the EGPS Test, writing was historically the subject with the worst performance in comparison with reading, maths, and science at Key Stages 1 and 2 within primary schools in England (Department for Education, 2012). At the time of the introduction of the EGPS Test, there was reportedly little evidence to explain either why children perform worse in writing compared to other core subjects, or what predictors can offer a picture of writing achievement (Department for Education, 2012). However, since the introduction of the EGPS Test, attainment in writing, as measured by teacher assessment, has risen year on year (Department for Education, 2019a) with national average attainment in EGPS either exactly the same or one percentage point different from that of writing.

The historic gap in attainment is unsurprising as writing is a complex task, involving an array of determinants (Wijekumar et al., 2019) where writers must make appropriate choices of audience and purpose based on the task required (McCutchen, 2000) in order to achieve the desired outcomes for writing. The production of written text also requires the writer to represent ideas in symbolic form in order to transmit textual information (Olive, 2004). Writers are then faced with competing factors such as idea generation, handwriting, spelling, punctuation, sentence construction, textual organisation, and consideration of audience and purpose. This multifaceted cognitive task is underpinned by a number of linguistic processes, motor skills, and executive functions (Hayes, 2012; Kim & Schatschneider, 2017), each of which influence writing both directly or indirectly (MacArthur, Graham, & Fitzgerald, 2006).

Children's ability to write develops over time, drawing on the writer's cognitive resources, the context in which the writing is situated, and the respective demands of the written product, which vary by age and phase of writing (Dockrell & Connelly, 2021). Becoming a competent writer is a developmental process which requires personal regulation and strategic behaviour (Bereiter & Scardamalia, 2013; Zimmerman & Risemberg, 1997) and writing's ultimate goal is 'the act of communication in all its myriad forms and genres' (Dockrell & Arfè, 2020: 16). It is therefore unsurprising that the development of competency in writing is understood to take considerable time (Wijekumar et al., 2019).

There are reported global concerns in the Anglophone countries, as well as the EU, about children's ability to compose written text (Myhill & Jones, 2009; Persky, Daane, & Jin, 2003), and there is consensus within the community of educational researchers for a more developed body of research regarding writing. There remains a lack of research around the skills that underpin writing, namely grammar and punctuation, particularly in comparison to that of spelling or reading (Cameron, 2012; Stannard & Huxford, 2007; Wyse, Jones, Bradford, & Wolpert, 2013). The limited articulation of the role of grammar and punctuation in text generation processes within current research presents potentially significant implications as it may have resulted in the limitation of informed thinking regarding the text generation processes in children's writing. As will be demonstrated in this study, there is limited understanding of how writing convention skills (grammar and punctuation) can be deployed to improve children's written outcomes within the primary classroom.

2.3 Writing: Process or Product?

When exploring the development of writing, multiple approaches have been followed by writing researchers. Writing can be considered from two perspectives: process or product (Berninger, Fuller, & Whitaker, 1996). For practitioners and researchers to fully understand children's development in writing, it is important to be able to conceptualise the components and processes of writing at the varying stages of a writer's development. The writing process considers the interaction between the various cognitive functions and abilities which contribute to written composition (Becker, 2006). This approach explores the component skills which underpin the production of written text (Arfé, Dockrell, & De Bernardi, 2016; Berninger et al., 1992; Juel, 1988; Mackie, Dockrell, & Lindsay, 2013). The influence of component skills such as handwriting, spelling, vocabulary, and grammar have been studied to varying degrees to understand their influence on substantive quality, described as macro-structure, (Arfé et al., 2016; Berninger et al., 1992; Juel, 1988; Kim et al., 2011; Limpo & Alves, 2013; Mackie et al., 2013) as well as on word- and sentence-levels, described as micro-structure (Arfé et al., 2016; Mackie et al., 2013).

Alternatively, writing products relate to the final outcomes of written text. Researchers have explored the contents of children's writing products through text analysis and factor analytic methods (Arfé et al., 2016; Puranik, Lombardino, & Altmann, 2008; R. Wagner et al., 2011) in order to develop an understanding of the dimensions or factors that account for variance in children's writing (Puranik et al., 2008; R. Wagner et al., 2011). If writing is seen as a multi-dimensional phenomenon, then it should not be explained simply through substantive writing quality alone (Kim, Al Otaiba, Wanzek, & Gatlin, 2015). Exploration of writing products through micro-structure can involve examining content, composition length, or spelling. In understanding the relationship between the EGPS Test and writing, alternative textual dimensions such as accuracy, productivity, and complexity should be considered (Arfé et al., 2016).

In gaining a better understanding of the component skills in the writing process and their impact upon writing products, educators and researchers can begin to better support the development of typical writers. Furthermore, understanding of the challenges and bottlenecks in the writing process can begin to offer insight to early identification and intervention for struggling writers. For example, developments in more recent cognitive models of writing have demonstrated the constraints that weaker handwriting and spelling place upon writing quality.

By identifying handwriting as a bottleneck in the writing process, practitioners can provide support and intervention for that component skill. This, in turn, can release cognitive resources for other aspects of the writing process. However, one challenge that faces researchers and practitioners is how to translate theoretical models into practice. While researchers may identify the component skills of writing, how to develop and enhance those skills is not always clear for practitioners as the ways in which researchers operationalise component writing skills does not always translate easily into interventions or curricula.

2.4 Struggling writers (SW)

Many children have difficulties in engaging in the writing process (Graham & Harris, 2005), and similarly teachers have reported that they too find the instruction of writing a challenge, citing limited information and support in making instructional adaptations to meet student's needs, particularly around spelling instruction (Graham et al., 2008). When considering the component skills involved within the writing process, it is not surprising that it has been identified that many children struggle with writing (Dockrell, Connelly, & Arfè, 2019). While up to 22% of primary students in England find writing challenging (Department for Education, 2019b), struggling writers (SW) are a heterogenous group (Walter, Dockrell, & Connelly, 2021) whose difficulties may stem from a range of direct and indirect factors in writing (Dockrell et al., 2019).

The writing products of SWs are reportedly shorter, more error prone in spelling, punctuation, and grammar, and demonstrate poorer compositional quality than typical writers (TWs) (Dockrell et al., 2019; Dockrell et al., 2015; Sumner, Connelly, & Barnett, 2014). In line with wider understanding of the writing process, transcription skills (spelling and handwriting) have been found to be the greatest barrier for SWs (Arfé et al., 2016; Berninger, Nagy, & Beers, 2011; Juel, 1988; Kim et al., 2011). Owing to issues with transcription (Dockrell et al., 2019), and the resulting difficulties with fluency (Berninger, 1999; García, Crespo, & Bermúdez, 2017; Graham, Berninger, Abbott, Abbott, & Whitaker, 1997), SWs typically present difficulties at word-, sentence-, and text-levels (Graham & Harris, 2005).

There are also several indirect factors within the writing process which may impact upon the writing products of SWs. Reading ability may impact SWs' writing products as reading has been found to be closely associated with spelling abilities (Abbott & Berninger, 1993; Berninger, Winn, et al., 2008). Oral language also supports text generation within the

writing process (Babayiğit & Stainthorp, 2010; Kim, Al Otaiba, Folsom, Greulich, & Puranik, 2014a; Savage, Kozakewich, Genesee, Erdos, & Haigh, 2017) and subsequent difficulties with oral language are frequently transferred to difficulties in writing (Graham, Hebert, Fishman, Ray, & Rouse, 2020; Mackie & Dockrell, 2004; Scott & Windsor, 2000).

Yet identification of struggling writers requires an assessment which is both reliable and can accurately capture the various demands of writing. While SWs have typically been identified using standardised writing measures (e.g. Wechsler Objective Language Dimensions (WOLD) (Rust, 1996); Wechsler Individual Achievement Test - Second UK Edition (WIAT-II^{UK}) Written Expression Subtest (WES) (Wechsler, 2005), these specialist and individualised assessments can be costly and time consuming. Assessments which are quickly able to demonstrate the specificity and sensitivity to identify struggling writers are limited (Dockrell, Connelly, Walter, & Critten, 2017), with standardised assessments additionally often lacking the ability to sensitively measure changes over time (Dockrell et al., 2015). One means of sensitively identifying and assessing struggling writers is through curriculum-based measures (CBM; see Chapter 4). While CBM is a useful tool for exploring a range of writing factors (e.g. productivity, accuracy, complexity), which is quick to administer, the scoring is arguably overly-complex for many teachers.

Owing to the identified difficulties that SWs have in grammar, punctuation, and spelling, this study explores the predictive ability of the EGPS Test. If the EGPS Test offers sufficient specificity and reliability, it could serve as an affordable and efficient tool for the identification of struggling writers.

2.5 Overall Summary

Writing is a complex skill made up of an array of component factors. Writing can be explored through the skills that contribute to writing, or through the assessment of writing products. Writing can be challenging, and a number of children struggle with writing, yet very little is known about the contribution of writing convention skills (grammar and punctuation) to the writing process, beyond the widely established contribution of spelling.

CHAPTER 3: The Writing Process: Models of Writing Development

3.1 Introduction

Cognitive psychology explores the cognitive processes that children engage in when they write. Cognitive models of writing attempt to both explain the skills required of writers in producing written texts, and the wider factors which influence these skills. While it has been argued that cognitive models of writing are of limited pedagogic use (Bazerman, 2018), they can provide a framework to think about, challenge, discuss, and discard knowledge about writing.

This chapter first explores the established cognitive models of writing. After considering critiques of these established models, alternative models and frameworks are presented. Finally, a discussion around two further theorised distal factors of writing are offered: grammar and punctuation.

3.2 Cognitive Models of Writing

The cognitive processes of writing have been described as a problem-solving process (Dockrell & Arfè, 2020) and many attempts have been made to conceptualise this process. Models seek to capture the constituent skills which are required for producing writing products, as well as the indirect skills and wider environments involved in the production of text (Graham et al., 2018). However, the process of writing which leads to written outcomes is complicated and multifaceted (Dunsmuir et al., 2015).

Primarily, models of the writing process seek to offer explanations of higher-order thinking which are able to unpick the complexities of the human mind (Kellogg, 1996). Yet more importantly in the context of a professional doctorate, models of writing offer a theoretical basis to the practitioner-researcher to find interventions and to develop appropriate curricular to enable children to be successful in producing written products (Bazerman, 2018).

Despite their importance, an understanding of the cognitive processes contributing to children's written products is still considered to be in its infancy (Graham & Harris, 2005). While a number of theoretical models of the underpinning factors of writing have been the focus of research over the last 40 years (MacArthur et al., 2006), reliably capturing writing competence remains a challenge (Dockrell & Connelly, 2021). A brief overview of the established cognitive models of writing is now offered.

3.2.1 Hayes and Flowers

In their seminal model of adult writing, Hayes and Flowers (1981) proposed a conceptual model of writing whereby writing is dependent on the interplay between three main cognitive processes: planning, translating, and reviewing (see Figure 1). The initial cognitive process - planning - consists of three further sub-strategies: ideas, organising ideas, and goal setting. The generation of ideas involves retrieval of pertinent information from the long-term memory and applying it to the context of the task. The organisational element of planning involves the categorising and sequencing of the retrieved ideas. Alongside this interactive process of retrieval and organisation, goal setting is the process of ensuring that the organised ideas meet the goal specified for the text. The second cognitive process is translating, which is the action of turning ideas from a linear or hierarchic plan into the alphabetic code as words and sentences (Flower & Hayes, 1981). This requires a writer to put ideas into language which can then be translated into the written word as a cohesive and coherent text (Berninger & Swanson, 1994). The final cognitive process is reviewing. This requires the writer to evaluate writing to ensure that the text written matches the intended goal. This may involve revision of ideas or the correction of transcription errors. This is a complex and conscious process which requires writers to systematically evaluate their product to aid further translation or text revision (Hayes, 1996).

The Hayes-Flowers model (1981) remained the most prominent model of adult writing for a number of years. However, Hayes (2012) further developed the Hayes-Flowers model in response to the accumulation of research and critique regarding the original model, the greatest levelled against a lack of recognition against transcription processes. While there were several significant developments in the Hayes model, the most pertinent were (a) the addition of transcription in the writing process, (b) the inclusion and emphasis of working memory, and (c) the addition of motivation to the writing process. While these elements are now included in the updated model, the model still lacks specificity in how these processes develop in children's writing. For example, transcription is referred to as handwriting, with no mention of the role and development of spelling skills. Regardless of these omissions, the Hayes and Flowers (1981) and Hayes (2012) models were significant in their contribution to the literature around cognitive models of writing. However, the underdeveloped skills that underpin writing are now explored through other cognitive models.
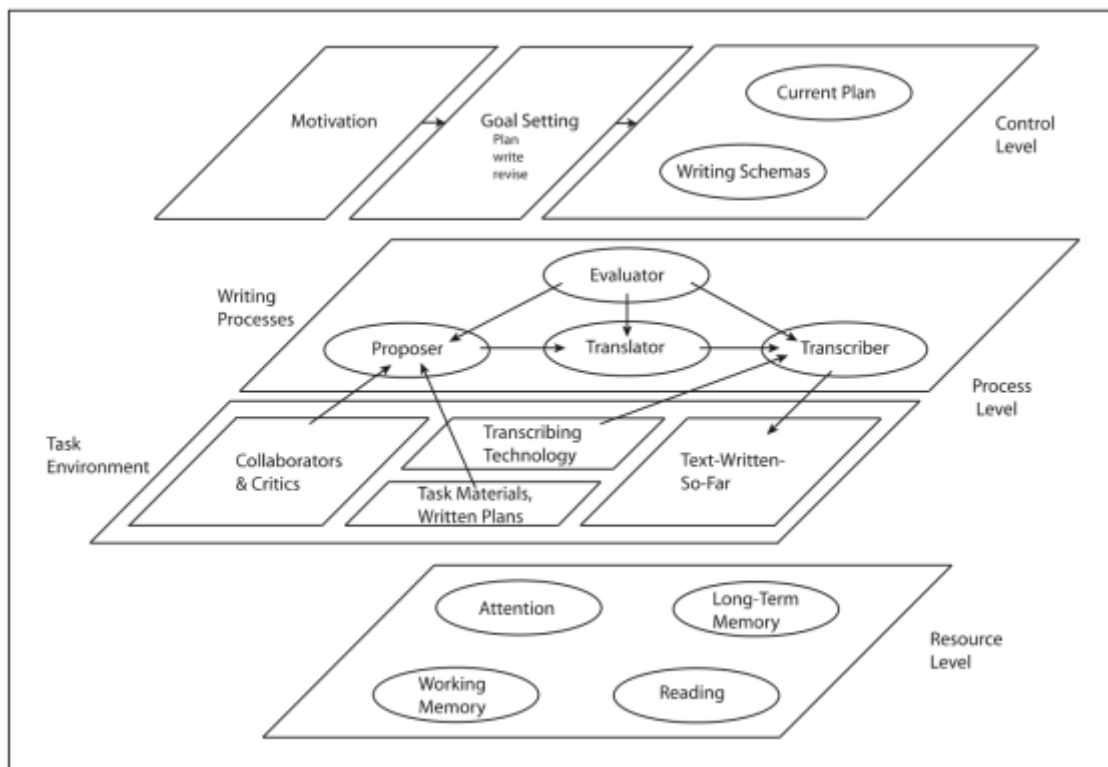
Figure 1 - Hayes (2012) update of the Hayes & Flower (1980) model of the cognitive writing processes.

3.2.2 Simple View of Writing

In their Simple View of Writing, Jue, Griffith, and Gough (1988) propose that the writing process is comprised of transcription processes (spelling and handwriting) and ideation processes (generation and organisation of ideas into sentences and texts). This framework stressed the importance of phoneme-grapheme correspondence, and of lexical knowledge on transcription processes; however, little detail was offered in relation to the process of ideation (Juel, Griffith, & Gough, 1986). Until this study, the understanding regarding children's cognitive processes in writing was assumed to be a simpler version of adult processes. As an exploratory study into the writing of children in elementary education, this study identified the importance of not simply seeing developing writing skills as a less complex version of adult writing (Berninger, 2019).

Berninger's revision of the simple view of writing (Berninger et al., 1996) extended the original model and proposed three major components, rather than two: transcription, text generation, and executive functions (see Figure 2). Transcription is comprised of the production of words and sentences through handwriting and spelling. Text generation is concerned with the achievement of writing goals at word-, sentence-, or text-levels. Finally,

13

executive functions refer to the processes of planning, monitoring, and revising. The interactions between transcription and executive functions, in the support of text generation, are supported by the working memory, synthesising contributions from both short-term and long-term memory. Berninger (2003) explains that children's early writing is reliant on transcriptions skills. As children gain increasing automaticity in transcription skills, executive functions become a greater factor in the writing process. It has subsequently been identified that inefficient transcription processes present as the most prevalent barrier to writing development in beginning writers (Abbott, Berninger, & Fayol, 2010; Berninger, 1999; Kim et al., 2011).

TEXT GENERATION

words, sentence, discourse

WORKING MEMORY

activates:
long-term memory (composing)
short-term memory (reviewing)

TRANSCRIPTION

handwriting,
keyboarding and
spelling

EXECUTIVE FUNCTIONS

conscious attention, planning
reviewing, revising,
strategies for self-regulation
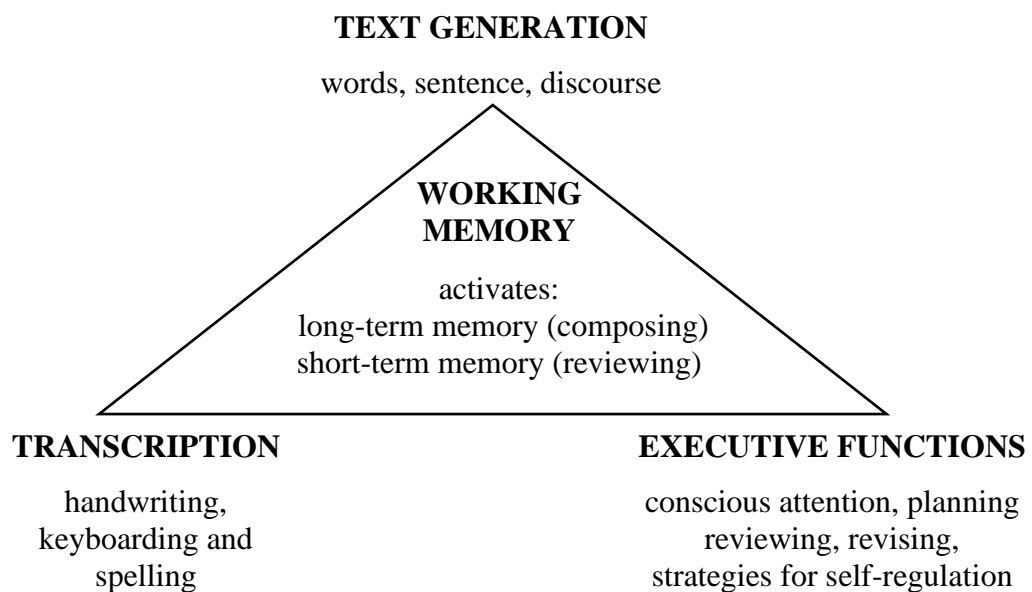
Figure 2 - The Simple View of Writing Model (Berninger & Amtmann, 2003)

With advances in brain imaging technology, Berninger and Winn's Not So Simple View of Writing identified a greater understanding of the role of the lesser developed areas of text generation and executive functions (see Figure 3). In addition, a greater understanding was offered in relation to the role and the component makeup of the working memory. The role of working memory was identified as to activate long-term memory within the processes of planning, composing, reviewing, and revising; similarly, working memory was also identified to activate short-term memory only during the processes of reviewing and revising. This greater understanding of the components of working memory contributed to the increased understanding of the wider role that working memory plays within the writing process. Finally,

the executive functions were expanded to include supervisory attention - the process of determining what is deemed worthy of attention or not.

TEXT GENERATION

WORKING MEMORY

Cognitive Flow

TRANSCRIPTION

handwriting,
keyboarding and
spelling

EXECUTIVE FUNCTIONS

supervisory attention,
goal setting, planning
reviewing, revising,
strategies for self-monitoring
and regulating
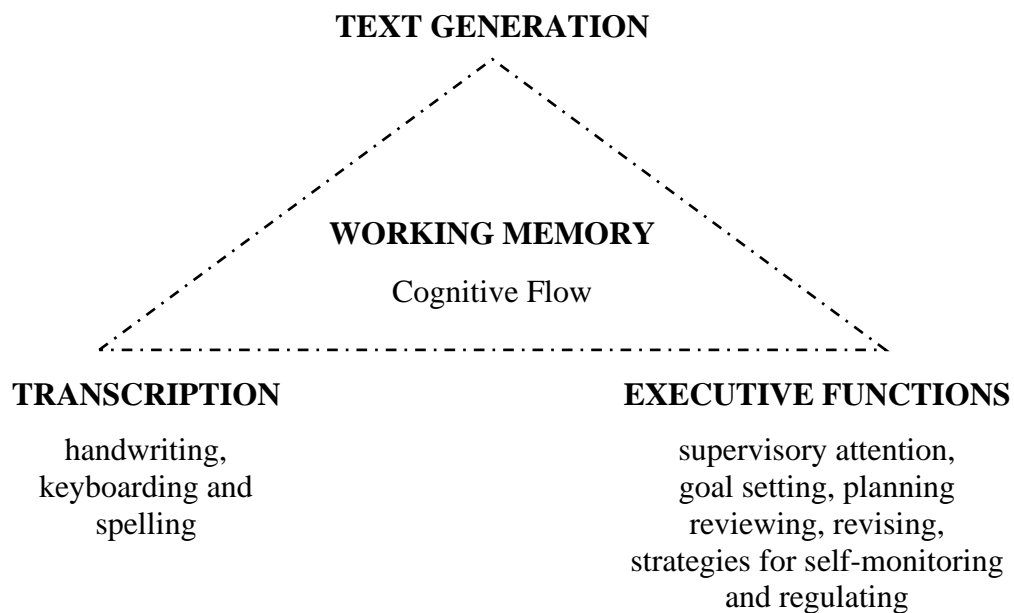
Figure 3 - Not-So-Simple View of Writing Model (Berninger & Winn, 2006)

3.2.3 Critiques of Established Cognitive models of writing

Models of writing within psychology seek to describe the psychological phenomenon that takes place within the writing process. In doing so, accounts of what writers need to know and address can be identified (Bazerman, 2018; Deane & Song, 2014; Graham, 2018).

Each of these cognitive models have identified the importance of idea generation, text generation, transcription, and a range of executive functions. Yet the cognitive models of writing explored above have been criticised more generally on several accounts. While these established cognitive models have been highly useful frameworks for understanding the writing process, critical elements remain underdeveloped. It is becoming increasingly understood that products and processes in children's writing within primary school are influenced directly or indirectly by a range of skills (MacArthur et al., 2006). A key critique of the established cognitive models of writing is that they do not fully explore the component skills which impact upon writing. This weakness in the models is particularly important in younger writers as children in primary schooling develop multiple levels of skills which potentially constrain writing development (Berninger, Mizokawa, & Bragg, 1991). Kim and Schatschneider (2017) argue that the not-so-simple view of writing specifically is

15

underspecified, citing a lack of specificity around the inter-relations between component skills. It is argued that the area of these models in greatest need of expansion is text generation. Text generation is a complex and dynamic process (Juel et al., 1986), but the cognitive models of writing explored above offer little explanation of the skills involved in text generation and the different levels of language required. In particular, the component skills of grammar and punctuation are not considered in any of these seminal cognitive models of writing.

3.2.4 Bereiter and Scardamalia

Bereiter and Scardamalia (1987) made a distinction between expert and novice writers through their knowledge telling and knowledge transforming models of writing which acknowledged different levels within the writing process. Beginning writers make use of a knowledge telling level strategy whereby text production is influenced by retrieval of content from the long-term memory; expert writers use a knowledge-transforming level strategy by elaborating upon representations of the rhetorical or communicative problems or goals to generate and evaluate writing content during the writing process.
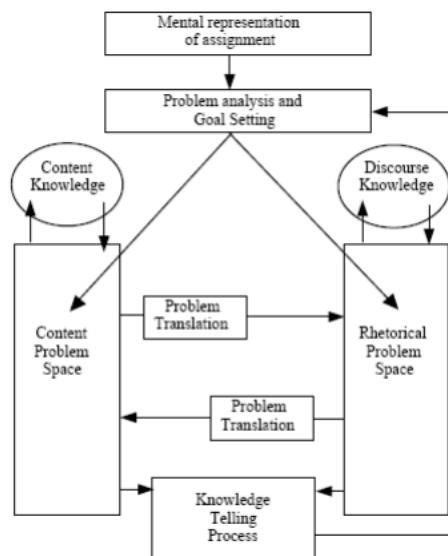
Figure 4 - Bereiter & Scardamalia's (1987) knowledge transforming model of writing

In their knowledge transforming model of writing (see Figure 4), two higher-level key processes are identified: the interaction between content knowledge and the content problem space and the interaction between discourse knowledge and the rhetorical problem space. Content knowledge refers to the knowledge of concepts and vocabulary required for the topic

or the goal of the writing product. The content problem space is where issues of belief and knowledge are worked out using the content knowledge available to the writer (Bereiter & Scardamalia, 1987). Alternatively, discourse knowledge refers to knowledge of discourse schema. In the rhetorical problem space, knowledge of specific literary forms or genre are worked out to specify the elements required for the text and how to use this knowledge to arrange the text. The successful repeated interaction of these problem spaces results in a transformation of knowledge. Goals are used to support this text generation process and to regulate the production of relevant information. However, the role of these higher-level skills can only be accessed when transcription reaches an adequate level of fluency (Berninger & Swanson, 1994). Whilst not strictly a cognitive model, this expansion in the understanding of text generation, and the importance of transcription skills is of note.

3.2.5 Levels of Language

The levels of language approach is an expansion of the cognitive models that considers writing quality at the sub-word, word, sentence and text/discourse levels (Benton & Kiewra, 1986; Whitaker, Berninger, Johnston, & Swanson, 1994). At sub-word and word levels, children begin to master alphabetic principles and phoneme-grapheme correspondences, which in turn are used to transcribe letters and words (Ehri, 1987). As words are formed, they begin to be separated with spaces and punctuation at the sentence level (Tolchinsky, 2006). This requires awareness of writing conventions. Writing conventions has been identified as an essential trait in writing (Culham, 2003; Diederich, 1974; Purves, 1988) which consists of accuracy of spelling, word separation, punctuation, usage and accuracy of grammar, capitalisation, and paragraphing (Culham, 2003; Salas & Caravolas, 2019). Alternatively, register is the combination of grammatical and lexical resources to meet expectations of a given context (Derewianka & Christie, 2008). At text level, knowledge of genre, register, and content becomes increasingly important as children write longer and more complex texts (McCutchen, 2006; Ritchey et al., 2016). The significance of this distinction allows for exploration of individual differences across each of the levels. For example, a child may be weak at sentence-level, but strong at the word-level. Understanding these individual differences enables researchers to understand bottlenecks in writing development (Hayes & Berninger, 2014). When exploring the process of translation within writing (transcription and text generation), the levels of language approach is particularly useful as it allows for the separation of textual processes from the information context of the texts being explored (Whitaker et al., 1994). It

17

is of additional use when considering the factors that cause intraindividual differences in writing (see Figure 5). For example, where a child may be strong at the word-level, but poor at the sentence level, explorations at different levels of language may help to explain factors that predict writing competence.
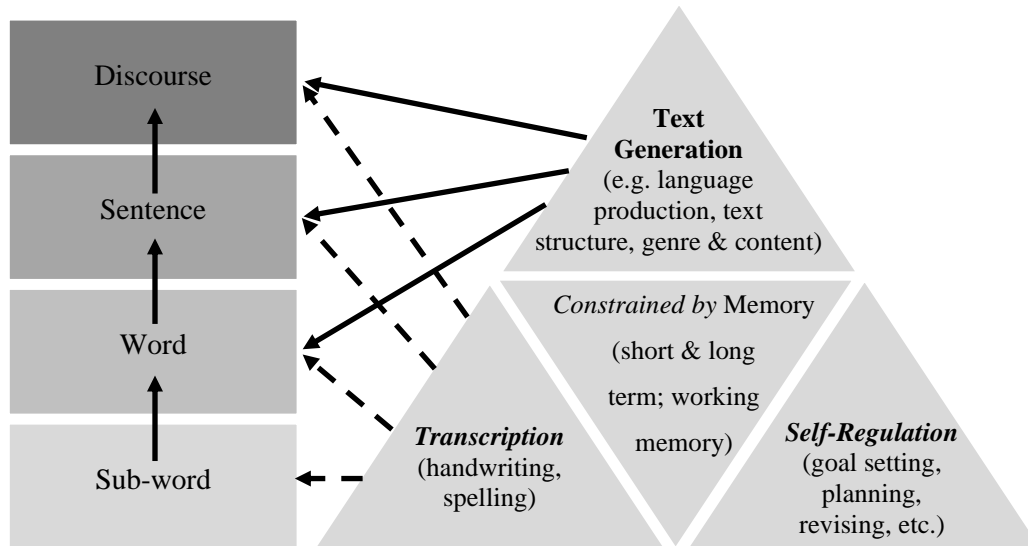
Figure 5 - Ritchey et al's (2016) adaptation of the Simple View of Writing with Levels of Language (Berninger & Amtmann, 2003)

However, while the levels of language approach is a helpful elaboration, Kim and Schatschneider (2017) argue that still little explanation is offered in regard to the component skills involved in text generation. Where the transcription aspect of the simple-view of writing has been specified in greater detail, citing morphological skills, orthographic knowledge, and phonological processing (Berninger, Abbott, Abbott, Graham, & Richards, 2002; Juel et al., 1986), specificity around the skills underpinning translation has been neglected (R. Alves, 2012). While word-level grammar in the form of morphology is considered in the levels of language approach, the role of grammar and punctuation remains underdeveloped and underspecified. With a lack of specificity, the relationships between the skills that underpin writing are ill-defined (Dockrell et al., 2019). The understanding around these relationships were further developed in Kim and Schatschneider's (2017) Direct and Indirect Model of Developmental Writing.

3.2.6 Direct and Indirect Effects Model of Developmental Writing (DIEW)

In their DIEW model of writing, Kim and Schatschneider (2017) posit that the previous cognitive models of writing mentioned above lack specificity regarding the specific components of the wider writing process. While Kim et al. acknowledge the development of understanding around transcription skills, they highlight the under-specification of the critical aspects of text generation and the skills that underpin its role in the writing process.

Drawing on theoretical models of discourse comprehension and production (Fletcher & Chrysler, 1990; Graesser, Singer, & Trabasso, 1994), DIEW draws upon levels of mental-representation in relation to text-generation. Component skills are situated within hierarchical structures (see Figure 6) whereby lower-order skills are required for higher-order skills, which in turn partially or completely mediate the effects of the lower-order skills upon writing. Within the model, executive functions support two key parts of the simple view of writing: transcription and text generation.

Within DIEW, Kim et al. operationalise text generation as discourse-level oral language skills in line with Juel et al. (1986). Text generation involves generating and organising ideas (Juel et al., 1986) which are required to be encoded and represented as oral language before writing is produced through transcription skills (Berninger, Abbott, et al., 2002; Kim et al., 2011). DIEW identifies discourse-level oral language as an upper-level skill, which is predicted by vocabulary and grammatical knowledge, as well as a range of higher-order cognitive skills (e.g. inference, perspective taking, monitoring).

The DIEW model expands on the simple view and the not-so-simple views of writing in several ways. Firstly, DIEW specifies the relationships between the skills that underpin writing, identifying the direct and indirect factors, such as vocabulary, grammatical knowledge, and inference making skills, that contribute to text generation. Secondly, DIEW expands the specified skills required for writing to include background knowledge, motivation and affect, as well as several higher-order cognitive and regulation skills.

Kim et al. hypothesised that as children develop direct skills in writing, such as transcription skills, then indirect skills, such as grammatical knowledge, become of greater importance in text production (Kim & Schatschneider, 2017). DIEW also identified that discourse-level oral language and transcription skills both had direct relations to writing (Kim

& Schatschneider, 2017). Alternatively, all other component skills contributed to writing indirectly through either transcription or discourse-oral language skills.
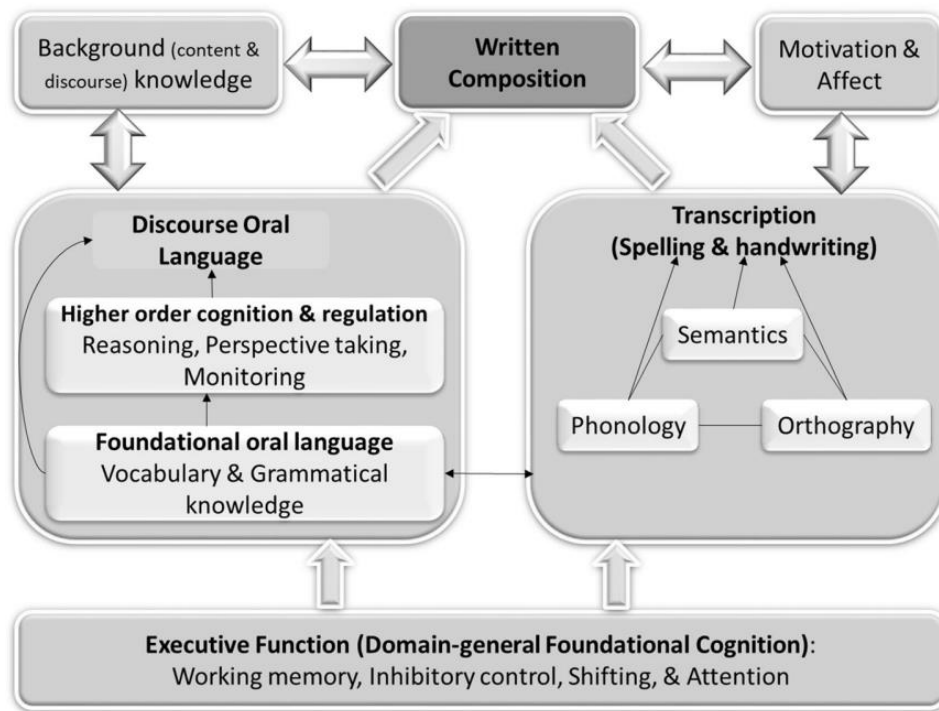
Figure 6 - Direct and Indirect Effects Model of Writing (Kim 2018)

This study offers great insight into the role of discourse-level oral language and expands upon previous cognitive models by exploring the component skills involved in translation processes. However, two critiques could be considered. Firstly, considering the levels of language approach described above, what remains unexamined is the role of the skills that underpin writing at the multiple levels of writing. If the writing process is a parallel process (Olive, 2014) working at word-, sentence-, and discourse-levels, then DIEW's operationalisation of oral language skills (grammar and vocabulary) at discourse level only is underspecified and does not consider the potential bottlenecks at each level with oral language. Research should take place to explore operationalisation of grammar and vocabulary at word-, sentence-, and text-levels. Secondly, while DIEW begins to explore some of the factors that underpin text generation through discourse oral language, there remain a number of direct factors (e.g. punctuation) and indirect factors (e.g. reading) that may impact upon writing which are still not considered within the model. Despite these critiques, the DIEW model of developmental writing remains a key theoretical influence within this thesis.

3.2.7 Proximal and Distal Factors in Writing

Drawing upon the literature regarding reading development (Zoccolotti, De Luca, Marinelli, & Spinelli, 2014), Dockrell (2019) draws parallels with the writing process, distinguishing the proximal (direct) and distal (indirect) factors that support writing development. In congruence with the DIEW model, factors that directly impact on writing, such as transcription skills, are conceptualised as proximal factors. Alternatively, skills that indirectly support the production of writing, such as oral language, reading or working memory, can be conceptualised as distal factors. The role of additional factors that may indirectly impact upon children's writing are of importance to the present study. Therefore, each of these factors are now explored in turn.

3.2.7.1 Proximal factors

Proximal factors are the transcription skills directly involved in producing written text, and both handwriting and spelling are the mechanisms required for translation of ideas into written products (Berninger et al., 1996). These transcriptional skills are regarded as important low-level skills in writing that require early mastery, as they can place significant constraints on writing development if automaticity is not achieved (Limpo & Alves, 2013; Limpo, Alves, & Connelly, 2017). Automaticity is regarded as the ability to recall information effortlessly to not burden the limited capacity of the working memory.

Proximal skills have been found to contribute to written expression in early writing (Connelly, Dockrell, Walter, & Critten, 2012; Puranik & Al Otaiba, 2012). Early writers invest a greater proportion of cognitive resources on transcription and, as such, do not attend to more global problems of text production. Older writers, who are less focused on transcription, become more attentive to meaning and global textual problems (Bereiter & Scardamalia, 2013). Inversely, where transcription skills are impeded in older writers, writing productivity is reduced (Olive, Alves, & Castro, 2009). This transcriptional impediment was also found in a study of oral storytelling in middle primary years where children's syntactic complexity was found to be greater in oral retellings than in written retellings (Fey, Catts, Proctor-Williams, Tomblin, & Zhang, 2004), confirming the constraints of transcription skills upon text generation. However, it is argued that the development of children's writing is not linear and involves simultaneous development of a variety of skills (Gillespie & Graham, 2014) and both of the proximal factors of writing tap into different processes (Abbott & Berninger, 1993). Yet mastery of proximal skills remains important as automaticity in these skills enables writers to

21

simultaneously activate the higher-level distal factors that underpin writing in the writing process (Berninger, Winn, MacArthur, Graham, & Fitzgerald, 2006; Kellogg, 1996; Olive & Kellogg, 2002). Each of the proximal skills are now explored in greater detail.

3.2.7.1.1 Handwriting

Handwriting is regarded as a complex skill that integrates orthographic and motor skills (Christensen, 2004), requiring both attention and memory as well as linguistic and motor skills (Bara & Gentaz, 2010; Graham & Weintraub, 1996). Handwriting can be assessed in relation to fluency and legibility (Graham, Berninger, Weintraub, & Schafer, 1998). Handwriting fluency can be operationalised through the number of letters or words written in a given time, considering both accuracy and speed. If children do not demonstrate automaticity in handwriting, the writing process will become more effortful, leaving fewer working memory resources to be devoted to text generation (Santangelo & Graham, 2016).

Handwriting fluency has been shown to predict compositional fluency (i.e., how much is written in a given time) and writing quality for younger primary school writers (Berninger, 1999; Berninger, Abbott, et al., 2006; Berninger, Vaughan, et al., 1997; Graham et al., 1997; Kim et al., 2011; R. Wagner et al., 2011). The time taken by children to transcribe words into text has been found to relate to the quality of written products in later primary years (Sumner, Connelly, & Barnett, 2013). Handwriting also impacts more than overall writing quality: letter writing automaticity has also shown significant relationships with both writing productivity (Graham et al., 1997; Kim et al., 2011; R. Wagner et al., 2011), and writing complexity (R. Wagner et al., 2011).

The contribution of handwriting fluency has been found to reduce between ages 6-7 to 8-9. In Kim and Park's (2019) recent longitudinal study of the DIEW model, transcription skills were found to present a constraining role in the writing process at 6-7 years of age, completely muting foundational oral language and higher order cognition skills. By ages 8-9, the constraints of handwriting skills were reduced, enabling direct and indirect contributions of working memory and higher-order skills which resulted in greater writing quality. While this study was carried out with South Korean children, it is argued within the study that the developmental nature of relations between transcription and higher-order skills would be language general, meaning that language relies on universal human information processing and therefore applicable to hypothetically applicable to English children.

22

What is not currently fully understood from the literature is the specific relationships between handwriting and other skills that may underpin text generation processes such as grammar and punctuation, particularly at the end of primary school. In a previous exploratory study into writing composition at 6-7 years of age, Kim et al. (2011) identified weak statistically significant relationships between letter writing ability and spelling, vocabulary, and grammar. However, in the later exploration of DIEW explored above (Kim & Park, 2019), handwriting and spelling completely mediated oral language (inclusive of grammar and vocabulary) at ages 6-7. By 8-9 years of age, the reduced constraints of handwriting allowed for greater contributions of working memory and higher-order cognitive skills. If this is the case, then it would be reasonable to expect that at least a comparable or lesser constraint from handwriting in Year 6 (ages 10-11) at the end of primary education. This would subsequently allow for greater potential direct or indirect contributions of grammar and punctuation to writing quality. If the findings of the study above are in-fact universal, then there remains a gap in the literature exploring the relationship between handwriting and the other component skills that underpin writing in Year 6.

It could be hypothesised that in the same way that automaticity in handwriting predicts compositional fluency and quality, the reduced burden on working memory would also allow for activation of the skills which underpin translation processes in writing such as grammar and punctuation. In parallel, automaticity in the application of grammar and punctuation knowledge may allow for a greater focus on compositional fluency and quality. However, this relationship is not yet understood and should be the focus of further study.

3.2.7.1.2 Spelling

Spelling is a transcription skill which encodes spoken language in to written forms (Perfetti, 1997), and is considered to be a significant component in the writing process (Crystal, 2012; Ehri, 1987). Spelling itself is a complex process which involves phonological and morphological understanding (Berninger, Abbott, Nagy, & Carlisle, 2010) in order to build up orthographic representations of words (Critten, Sheriston, & Mann, 2016). As a component of transcription skills, spelling facilitates efficient writing (Limpo et al., 2017) and is regarded as a prerequisite to expressing vocabulary in writing (Sumner et al., 2014: 293). It is argued that spelling 'bridges idea generation and text generation' (Abbott et al., 2010: 296).

Phonological awareness at 5 years of age has been found to predict writing competence at age 11 (Savage et al., 2017) and spelling ability at ages 5-6 has been found to be a predictor of both narrative and expository writing for children aged 8-9 (Kim, Al Otaiba, & Wanzek, 2015). Spelling has also been found to account for differences in text-level composition from ages 6 to age 13, owing to children with stronger spelling skills being able to translate and combine ideas in to written text. There is a widely established relationship between spelling ability and writing productivity (Graham et al., 1997; Kim et al., 2011), and there is also a limited amount of research identifying spelling as the only predictor of writing complexity (Kim, Al Otaiba, Folsom, Greulich, & Puranik, 2014b). Furthermore, there is a further body of research indicating that spelling is related to writing quality (Juel, 1988; Parker, Tindal, & Hasbrouck, 1991), with spelling at the word-level found to be a predictor of handwriting fluency and written sentence generation (Dockrell et al., 2019), although this study explores children with specific language impairment. Finally, the percentage of spelling errors at sentence and text levels has also been found to account for variance in writing quality (Berninger, Nielsen, Abbott, Wijsman, & Raskind, 2008; Graham et al., 1997; Mackie & Dockrell, 2004; Olinghouse, 2008).

The relative cognitive demands of spelling may limit children's productivity (Sumner et al., 2013) or vocabulary choices within writing products (Sumner et al., 2014). Furthermore, it is argued that students who struggle with spelling when composing writing products are more prone to forget their textual ideas and disregard rules around semantics and sentence construction (Limpo et al., 2017).

Writing is not a unitary construct, but it is clear from a strong body of evidence that spelling is a vital component in almost all the operationalisations available for the study of writing competency (quality, accuracy, productivity, and complexity). Therefore, when exploring the role of the EGPS Test, it will be important to establish the contributions of dissociable factors above and beyond spelling upon writing. In addition to handwriting, a wider range of distal factors should be considered.

3.2.7.2 Distal Factors

While research has identified the importance of the role of proximal factors in children's written production, proximal factors should not necessarily be seen with greater importance than distal factors (Harmey, 2021). As explored above, when constraints of the

transcription skills are reduced, it is hypothesised that distal skills play a larger role in production of written text, including oral language, reading, and working memory (Abbott & Berninger, 1993; S. Kent & Wanzek, 2016; Kim & Schatschneider, 2017). Each of these distal factors are now explored in turn.

3.2.7.2.1 Oral Language

In order to produce written texts, children are required to draw on multiple levels of language (Abbott et al., 2010). Children's ability to do so, particularly in relation to generation of oral and written sentences, has been found to underpin the production of written text (Arfé et al., 2016; Arfé & Pizzocaro, 2016) and oral sentence fluency influences text production over time (Savage et al., 2017). Two common operationalisations of oral language skills are measures of vocabulary and grammar.

Vocabulary is required to articulate ideas in writing accurately so that idea may be comprehended by the reader of the text. Vocabulary is often operationalised as lexical diversity (variance of word choice) and lexical density (proportion of content words). However, The established cognitive models fail to identify the role of vocabulary as a distinct component in the writing process, although DIEW (Kim & Schatschneider, 2017) does include this as a component of discourse oral language.

While grammar will be explored in greater detail in later chapters, a brief discussion is offered. From the little research that has taken place regarding the role of grammar in cognitive models of writing, there is some evidence of a relationship between oral grammar and children's writing (Arfé et al., 2016; Kim & Schatschneider, 2017; Olinghouse, 2008), although evidence has been conflicting in places (Kim, Al Otaiba, Sidler, & Gruelich, 2013; Puranik & Al Otaiba, 2012). Oral grammatical understanding has been found to act as a predictor of writing compositional quality (Olinghouse, 2008), with those students demonstrating a higher grammatical understanding achieving higher compositional scores in writing. In a study of first-grade learners (ages 6-7), Kim (2013) identified that children with highly developed grammatical knowledge, as operationalised through oral language completion tasks, produced writing of higher substantive quality than those with lower grammatical understanding. Furthermore, oral grammatical skills have also found to explain more variation in text quality than spelling in Italian children aged 7-8 (Arfé et al., 2016). Greater text quality may result from less constrains of transcription skills in middle and later primary years.

A number of studies have indicated that oral grammatical competence and vocabulary can explain a significant proportion of the variance of narrative quality in children's writing products (Kim et al., 2011; Kim, Al Otaiba, et al., 2013; Olinghouse, 2008). Suggestions have been made that efficient translation processes (lexical retrieval and syntactic generation) are based upon oral language processes (Bourdin & Fayol, 1994; Hayes, 2000) and automaticity in translation enables a greater proportion of cognitive resources to be made available for other aspects of the writing process (Bourke & Adams, 2003). However, it is argued that studies considering oral language fail to fully measure oral language skills such as receptive vocabulary, expressive vocabulary, and listening comprehension (Wise, Sevcik, Morris, Lovett, & Wolf, 2007).

Inversely, oral language has been found to constrain the writing process at word-, sentence- and text-levels (Berninger & Swanson, 1994; Berninger, Vaughan, et al., 1997). However, the limited body of knowledge around oral language and writing products (Apel & Apel, 2011; S. Kent & Wanzek, 2016; Olinghouse & Graham, 2009) support the notion that oral language skills are a distal factor in writing production that should be explored further when exploring the component skills that underpin writing.

While promising in support for the case of grammar, research in this area has focused purely on oral-language or receptive grammar tasks. It has been acknowledged that use of oral-language grammar completion tasks might not be best suited to capture the various dimensions of writing (Kim et al., 2014b). Such completion tasks rely on children's implicit knowledge for correctness against standard English. What is not determined by these tasks is how declarative knowledge of grammar and punctuation can inform or improve written products.

3.2.7.2.2 Reading

The relationship between reading and writing is well established (Berninger & Abbott, 2010; Berninger, Abbott, et al., 2002; Fitzgerald & Shanahan, 2000; Schoonen, 2019; Shanahan, MacArthur, Graham, & Fitzgerald, 2006) and reading has been found to be more uniquely related to writing, above and beyond the contribution of oral language skills (Berninger & Abbott, 2010; Kim, Al Otaiba, et al., 2013; Kim, Park, & Park, 2013). Reading skills are reportedly most closely associated with spelling skills, influencing writing directly and indirectly through spelling (Walter et al., 2021). Yet when considering reading's role in the writing process, many studies fail to account for the developmental nature of reading

development and the different ways that reading can be measured (Wise et al., 2007). In an attempt to conceptualise the reading process, the simple view of reading (Juel et al., 1986) offers two components to effective reading: word reading and comprehension.

Word reading, a component reading skill underpinned by phonemic awareness (Juel, 1986), has long been identified as associated with the component transcription skill , spelling (Abbott & Berninger, 1993; Ahmed, Wagner, & Lopez, 2014), whereby word recognition ability predicts spelling ability. Moderate associations have been identified between measures of word reading and writing (S. Kent & Wanzek, 2016; Kim, Park, et al., 2013) and word reading ability has been found to be a predictor of narrative quality in younger writers (Berninger, Abbott, et al., 2002; Olinghouse, 2008). While it has been identified that reading and writing both implement the use of similar knowledge (von Koss Torkildsen, Morken, Helland, & Helland, 2016) through a bidirectional relationship (Abbott et al., 2010), they remain as dissociable skills with differing developmental progressions (Berninger, Abbott, et al., 2002; von Koss Torkildsen et al., 2016). While addressing a different age-group to this study, word reading has also been identified as uniquely relating to writing quality for children aged 8-9 (Olinghouse, 2008). However, more widely evidenced relationships between reading comprehension and writing have been identified than for the relationship of word reading and writing.

The second component skill in the simple view of reading is reading comprehension. Reading comprehension is a process whereby meanings of words are integrated into textual structures through a single underlaying process, which is underpinned by both reading and listening comprehension (Juel, 1986). Reading comprehension has been found to have a direct impact on written compositional fluency and quality across different genres of writing (Berninger, Abbott, et al., 2002) and reading fluency has also been found to demonstrate a significant relationship with writing productivity (the amount of writing produced)(Kim et al., 2011; Williams & Larkin, 2013). One explanation offered is that reading fluency represents the automaticity of lexical access and this rapid interaction between orthographic and semantic information may impact upon translation processes in children's written productions (Williams & Larkin, 2013). Yet this same study also found no significant relationship between reading measures and quality or content of written products. When considering the relationship between reading and writing of weaker readers and writers, there is clearer agreement that

weaker readers have been shown to produce weaker content and organisation in their written products (Cragg & Nation, 2006).

While it is recognised that reading and writing use common and unique cognitive systems (Berninger, Cartwright, Yates, Swanson, & Abbott, 1994), it has been theorised that both reading and writing draw upon shared linguistic knowledge and orthographic knowledge (Berninger, Abbott, Thomson, & Raskind, 2001; Fitzgerald & Shanahan, 2000; Proctor et al., 2020). More recent work exploring cognitive models of both reading and writing has found that both processes also draw upon declarative and non-declarative linguistic knowledge (Schoonen, 2019). If this distal factor, reading comprehension, is underpinned by the ability to integrate meaning of words into sentence and text structures (Juel, 1986), this is likely to impact upon children's subsequent production of written sentences and text structures. It could be hypothesised that greater comprehension of written texts and the associated linguistic knowledge of grammar and punctuation may lead to greater accuracy and complexity in writing. As with oral language skills, the limited moderate associations between measures of reading and writing products support the notion that reading skills are a distal factor in writing production.

3.2.7.2.3 Working Memory

The status and importance of working memory is captured in the Simple View of Writing (Berninger, 1996) and the Not-so-simple View of Writing (Berninger, Abbott, et al., 2006). Where short-term memory facilitates the retention of incoming information involved in the processes of reviewing and revising (Berninger et al., 1991; Swanson & Berninger, 1996), long-term memory draws upon subject-specific content knowledge, discourse structures, and schemas of differing genres of writing (Berninger, 2000). Extraneous demands are placed on working memory during text production, owing to challenges presented by dual processing (McCutchen, 2000). In their DIEW model, Kim (2017) situates working memory as a lower-order executive function which can be mediated by transcription skills.

Working memory facilitates the temporary storage and processing of information retrieved from the long-term memory (Kellogg, 1996). Within this multi-store memory model there is only a limited capacity. The phonological loop and the visuospatial system are switched between by the central executive (Baddeley & Hitch, 2000). In the context of writing, the working memory store is required to demonstrate efficiency as it switches between holding

ideas generated from the long term memory whilst engaging in transcription skills (Kellogg, 2008)

An increasing body of research has demonstrated the relationship between verbal working memory and children's written narrative products (Berninger et al., 1992; Bourke & Adams, 2003), and a number of studies have demonstrated the significant relationship between working memory and narrative length (Berninger, 1999). In addition, the listening span measure of working memory has been found to be a significant predictor of text generation within later primary school years (Swanson & Berninger, 1996). Owing to its indirect role in the writing process, working memory can be considered as a distal factor in writing production.

3.2.7.3 Proximal and Distal Factors Critique

Within Dockrell et al.'s (2019) proximal and distal factors framework, the relationship between proximal and distal factors was only considered at the sentence level through an exploratory sentence generation task with struggling writers. Further research should take place with typical writers exploring proximal and distal factors whereby the specific components of text generation could be explored through either multiple levels of language (word, sentence, discourse) or through a multi-level composite assessment of writing.

While we are beginning to understand the range of proximal and distal factors of writing for younger writers, the framework remains underspecified and there is still some further research required to understand the potential wider range of distal factors of writing, including two domains of concern in this thesis: grammar and punctuation. Yet, in relation to grammar and punctuation and writing, few studies offer insight. The only study into these potential distal factors of writing in Year 6 available at the time of this research explains that punctuation measures accounted for 19% of variance in writing quality (Daffern, Mackenzie, & Hemmings, 2017). There was no significant relationship between grammar and writing beyond Year 4. Interestingly, in both Years 3 and 4, where grammar explained 19% and 28% of writing variance respectively, this relationship was only evident in females when controlling for gender. This study deployed the NAPLAN writing conventions test, a visually oriented, decontextualised test of spelling, grammar, and punctuation. Children are required to identify spelling errors in isolation and in short phrases, and to identify and label grammar and punctuation conventions from multiple choice questions. However, the measure of grammar and punctuation used in this study has been critiqued owing to the level of proofreading and

editing required (Willett & Gardiner, 2009). As a result, the assessment used in this study may tap editing and proofreading rather than grammar and punctuation knowledge. By contrast, in the writing convention test of concern to this study, EGPS Paper 1 places greater demands on various forms of 'constructed response' (STA, 2015:25) than the NAPLAN, requiring children to 'write a short answer of their own within a specified format'. This potentially requires a greater explicit knowledge of conventions around grammar and punctation that demonstrated from the multiple-choice response required within the NAPLAN tests. EGPS Paper 2 also requires children to produces spellings in a dictated sentence, rather than find errors in isolated words or phrases, which again may offer greater insight into the role of spelling knowledge than the NAPLAN Test. With all this in consideration, the role of grammar and punctuation as distal factors should be further explored through further research.

3.3 Theorised Distal Factors: Grammar & Punctuation

Both Kim and Schatschneider (2017) DIEW model of developmental writing and Dockrell's (2019) framework of proximal and distal factors of writing are key theoretical influences on this thesis. However, both fail to specify the role of both grammar and punctuation at different levels of language. Grammar and punctuation are key skills which underpin writing that serve as visible indicators of writing quality (Daffern et al., 2017) whereby texts with errors are perceived to be poorly constructed. Inversely, proficient writers demonstrate the ability to use these skills accurately when producing writing products (Fang & Wang, 2011). It has been noted that difficulties with these skills which underpin writing not only affect children's attitudes towards writing (Snowling, 2000), but also the overall quality of writing products (Berninger, Vaughan, et al., 2002). Daffern et al (2017) hypothesised that, if a child can demonstrate autonomy and agency with these given writing skills, they may in turn become more motivated and confident writers. Yet, as mentioned above, Daffern et al.'s (2017) operationalisation of grammar and punctuation has been questioned. Therefore, before we establish the means of assessing these writing conventions, it is beneficial to explore the hypothesised distal factors, grammar and punctuation, as well as the associated terms such as grammars, meta-language, and mechanics.

3.3.1 Grammar and Grammars

The term grammar, while in regular use in common parlance, can still be understood in various and differing ways (Hartwell, 1985). In most public and educational contexts, grammar is widely understood as an agreed way of speaking and writing; it is a rulebook for

communication. However, such simplicity in definition does not allow for the polysemic nature of the term. Within its own specialist dictionaries, grammar also defines itself in multiple ways: (a) a language system; (b) the rules about how words change their form and combine with other words to make sentences; (c) a book of grammatical rules; and (d) an individual's use of these rules (Chalker & Weiner, 1998: 176). Crystal and McLachlan (2004: 7) define grammar simply as 'the business of taking language to pieces to see how it works'. Alternatively, Lefstein's (2009: 379) definition encompasses the complexity of grammar and defines it as 'the study of language patterns and structure: not only morphology and syntax, but also elements of semantics and pragmatics'. This definition is congruent with the levels of language approach (Whitaker, Berninger, Johnston, & Swanson, 1994), whereby grammar is present and impacts upon writing quality at various levels of language. Morphology is an element of grammar which deals with modification of words, demonstrated in both inflectional and derivational morphemes. Within writing, morphemes impact upon writing at the word level through spelling (Bryant & Nunes, 2006). Both semantics (knowledge of meanings in sentences) and syntax (knowledge of sentence structure) are essential grammatical elements in sentence level writing (Abbott et al., 2010) which can impact upon the writers choice in ideation processes through sentence combination (R. Andrews, Torgerson, Beverton, Freeman, et al., 2004). Finally, pragmatics (knowledge of language in context) is knowledge of grammar at text level which involves meeting the intended goals of specific discourse schemas or genres; writing quality is affected by a writer's ability to meet these expected schemas (Cullen, 1999).

When looking at the term grammar as defined by the grammar dictionaries, there are competing and potentially contrasting definitions offered, with the two most notable being grammar as a rule book, or grammar as an observation of the individual's use of language. At this point, it is helpful to draw the distinction between grammar and grammars: the former being the study of language and the latter being language systems which can be explored and categorised through various assumptions around how language works. Pullum (2018: 12) surmises this distinction with grammar which refers 'to the structural properties and organisation that the sentences in a given language have' as opposed to grammars which are 'a description of those properties worked out by a linguist'.

3.3.2 Grammars as Language Systems

Prescriptive grammars are language systems concerned with form, and are predicated on the assumption that grammars should prescribe the required standards and benchmarks by

which deviations of language should be judged (Myhill & Watson, 2014). Prescriptive grammars are applied to spelling, syntax and register, and such grammars seek to formulate a single standard form of a language. Within prescriptivism, to deviate from this standard form is to use language incorrectly; in English primary schools, this prescriptivist form is called Standard English.

Conversely, descriptive grammars offer an analysis of how language users engage with language in different environments (L. Andrews, 2013), regardless of linguistic rules. Descriptivism argues that language is constantly changing, reflecting age, class, gender, and social status of those using language. As such, there is no argument of correctness within descriptivism as there is within prescriptivism.

More recent, and reportedly more coherently theorised (Myhill & Watson, 2014), rhetorical grammars are concerned with crafting and shaping language in order to achieve both rhetorical and stylistic effects (Paraskevas, 2006) in order to enable writers to understand and control their written products (Kolln, 2006). Rhetorical grammars seek to develop an understanding about the use of language, rather than to prescribe or describe as with other grammars offered above. This knowledge within rhetorical grammars can be theorised as meta-linguistic knowledge (Myhill & Watson, 2014).

For teachers, these opposing grammars offer pedagogical conflict, as many linguists and researchers are aligned with descriptivism in grammar (L. Andrews, 2013), while many teachers are concerned with making meaning and application of knowledge in practice through rhetorical grammars. Yet criteria of English Primary National Curricula (DfE, 2014: 5) asserts a prescriptivist pedagogical approach whereby pupils 'should be taught to control their speaking and writing consciously and to use Standard English'. However, this subscription of the National Curriculum to prescriptivism is not substantiated by empirical research. Rather, the alignment of the national curriculum over the last 30+ years has been described as rooted in folk theorisation of grammar (Myhill & Watson, 2014). Such views are entrenched in historical discourses which have held schools accountable for correcting the evil habits of speech contracted in the home and the street (Board of Education, 1921), whereby teachers are held responsible for compensating for the linguistic disadvantage of the home (Bernstein, 1971). One of the challenges for teachers is that to encourage a positive culture of the teaching of grammar which is not rooted in prescriptivist rules is frequently misinterpreted as a lowering

of standards (Pullum 2018). Therefore, considerable time in English primary schools is spent attending to prescriptive grammar and meta-linguistic knowledge.

3.3.3 Grammar Knowledge & Metalinguistics

Knowledge of grammar is essential for writing competency at word-, sentence-, and text-levels. Word-level grammar (operationalised as vocabulary in the EGPS Test) is seen within morphology through derivational suffixes (Berninger et al., 2011). Sentence-level grammar requires the manipulation of syntactic-level structures that produce clauses and makes use of syntactic understanding to combine multiple clausal structures into a single sentence (Berninger, Abbott, & Alsdorf, 1997). Finally, grammar influences the text-level through syntactic choice which impacts on the academic register of writing products (Derewianka & Christie, 2008) and increasing grammatical knowledge can lead to an improvement in writerly choice (Love & Sandiford, 2016). However, cognitive models of writing do not yet fully account for the role of grammar in the writing process (Arfé et al., 2016).

In establishing the contribution of *grammar knowledge* as a component writing skill, one of the greatest challenges is how researchers operationalise grammar for its analysis. What researchers consider as *grammar* is varied and so it could be argued that its measurement is inconsistent. For example, in the current body of research about writing, grammar has most typically been operationalised as either oral receptive grammar or visually oriented grammar (Daffern et al., 2017; Kim et al., 2011; Olinghouse, 2008). Oral grammar is measured either through grammatic completion tasks, whereby children must listen to a sentence with missing parts and are required to provide grammatically correct responses, or through grammatical understanding whereby children listen to a sentence with errors in syntax and determine if they are correct or incorrect. Alternatively, grammar has also been captured through visually oriented tasks which require either the identification of correct syntax, or production of syntactically correct sentences.

Within the measures described here, a distinction can be drawn between receptive grammar and productive grammar knowledge (Shintani, 2018). Receptive grammar knowledge refers to the understanding of oral or written language. Alternatively, productive grammar knowledge refers to the production of spoken or written code.

Alternatively, both receptive grammar and productive grammar can be implicit or declarative. Implicit knowledge is the automatised and unconscious understanding about

33

language which can distinguish correctness without awareness. Declarative knowledge is the conscious and deliberate sharing of knowledge about language which is produced with awareness of decision making. With these distinctions drawn when considering potential measures of grammar, it is understandable that a consistent operationalisation of grammar has not reached consensus.

Metalinguistics is a term which is used across the disciplines of psychology and linguistics, which can also lead to different definitions (Myhill, Jones, & Watson, 2013) and ambiguity in understanding (L. Andrews, 2013). Psychology frames metalinguistics as a thinking process, whereas linguistics distinguishes language as an artefact. In the pursuit of an understanding suited to writing within schools, Myhill (2011) defines metalinguistics as:

> '[The] explicit bringing into consciousness of an attention to language as an artefact, and the conscious monitoring and manipulation of language to create desired meanings grounded in socially-shared understandings'.
>
> (Myhill 2011, p250)

The difference between meta-linguistic knowledge and grammar knowledge is not mere semantics (Canniford, 2019) and, while the two terms are often used interchangeably, there are distinctions to be made. According to Myhill et al. (2013), meta-linguistic knowledge is knowledge about all elements of language. Alternatively, grammatical knowledge is the declarative knowledge of grammar, including morphology and syntax, which is articulated through the meta-language of grammatical terminology. It is within this framing that many teachers, children, and wider society, practically conceptualise the term *grammar* – as an element of metalinguistics that makes use of a declarative understanding of meta-language which can be used to explore, discuss, and understand language. This common language *about language* enables writers, learners and teachers to explore language using explicit terms which enable understanding (Waugh, Warner, & Waugh, 2019).

From what little research has been conducted exploring the various operationalisations of grammar, several reviews and meta-analyses have informed us of the effect of grammar *teaching* on students' learning about writing. Of such reviews, few have been favourable for the case of grammar, with most arguing that explicit teaching of grammar has little to no effect, if not a harmful effect on students' writing (Braddock et al., 1963; Hillocks Jr and Smith, 1991; Weaver, 1996; Wyse, 2001; Andrews et al., 2004; Andrews et al., 2006; Sheard et al., 2012; Myhill and Watson, 2014). While acknowledging the lack of recent research, the Evidence for

Policy and Practice Information and Coordinating Centre's review of the effect of grammar teaching on the accuracy and quality of written composition (Andrews et al., 2004) concluded that no evidence could be found which supported the notion that grammar *teaching* has any positive impact on the quality of children's writing. In a further and more recent meta-analysis, Graham, McKeown, Kiuhara, and Harris (2012) also confirmed that grammar instruction did not statistically influence writing quality. Yet all the studies previously mentioned focused on prescriptive grammars, rather than more recent theorisations of grammars. By contrast, most recent research by Myhill (2016) explains that research has shown 'a positive relationship between teaching grammar and improving writing', citing that where learning is not decontextualised, it can be beneficial. However, this study has been fiercely disputed, citing a lack of robust evidence regarding the benefits of the teaching of grammar (Wyse & Torgerson, 2017).

However, the present study is not concerned with the teaching of grammar. Rather it is focused on the relationship between grammar and punctuation and writing outcomes. More recent studies that seek to understand predictors of writing outcomes are developing a body of evidence demonstrating the impact of implicit and explicit grammar knowledge upon writing products. For example, two distinct aspects of grammar, grammatical knowledge and grammatical understanding, have been identified as predictors of writing quality (Daffern et al., 2017; Kim et al., 2014a; Kim, Al Otaiba, Wanzek, et al., 2015; Olinghouse, 2008), accuracy (Arfé et al., 2016; Fayol, Hupet, & Largy, 1999; Mackie et al., 2013; Negro, Chanquoy, Fayol, & Louis-Sidney, 2005), productivity (McMaster, Ritchey, & Lembke, 2011; Nelson & Van Meter, 2007; R. Wagner et al., 2011), and complexity (Mackie et al.; Van der Lely & Christian, 2000; R. Wagner et al., 2011). It is argued that, in embedding a conscious knowledge of differing linguistic structures, writers are more able to produce clear, well-structured unambiguous sentences (Derewianka & Christie, 2008), leading to greater writing quality at sentence- and text-levels. This study will explore the impact of one operationalisation of grammar knowledge upon writing: The EGPS Test.

3.3.4 Punctuation

Punctuation is best defined as a framework of non-alphabetical symbols which are situated within writing to group and partition words in order to aid comprehension through the demonstration of relationships in writing (Cullen, 1999). Punctuation can be seen as a means of establishing and distinguishing sense within writing. Curtis (2007: 3) argues that

punctuation's basic function is to organise words to enable the reader to understand 'precisely what the writer meant to say'.

However, Fayol (1997) posits that punctuation is integral to the process of transforming thought into a linear dimension. Punctuation aides coherence and presentation of complex ideas (Dockrell et al., 2015), yet use of punctuation can be both a constraining and enabling tool for writers (Pontecorvo, 1997) as effective use of punctuation requires a competent understanding of syntactic and semantic requirements of text organisation (Ferreiro & Pontecorvo, 1999). While acknowledging the importance of mastering punctuation in the journey towards being a competent writer, Waugh et al. (2019) argue that discrete punctuation exercises are unlikely to enable children to punctuate writing well. Instead, Waugh cites the need for an understanding within the context of the 'craft of writing'.

However, there is little empirical evidence examining knowledge of punctuation and its impact on writing products. While many aspects of punctuation are key in learning to write, they are not key in learning to speak (Waugh et al., 2019). Without explicit knowledge of punctuation, there would be no logical reason to transfer this skill from speech into writing products, as speech does not require punctuation marks. Dockrell (2015) considers punctuation to be a transcription skill as it involves the process of mark-making, and it is widely accepted within cognitive models that transcription skills precede writing quality (Berninger & Swanson, 1994; Hayes, 2012; Sumner et al., 2014). It is hypothesised in this study that if the ability to identify and label conventions of punctuation is a predictor of fluency in transcription, punctuation knowledge may in turn lead to greater writing quality.

3.3.5 Writing Conventions as Mechanics

Mechanics is the embodiment of grammar in practice: the process of translating grammatical understanding into written products. Where we could reasonably expect to see grammar and punctuation knowledge demonstrate transferability into writing is through writing mechanics. Mechanics can be distinguished from grammar in as much that grammar can be regarded as the conventions around verbal communication which create meaning (Anderson, 2005), whereas mechanics can be seen as the existence of features within written forms of communication which are not apparent in verbal communication, including spelling, capitalisation, punctuation, and organisational features of writing such as paragraphs. What is important from this distinction is that the effective application of mechanics requires the ability

to take grammatical understanding and to translate it into written text (Canniford, 2019). Children who score highly in measures of mechanics are likely to find their writing is perceived as eloquent, well informed and capable of high-order thought; conversely, children who score poorly in measures of mechanics are likely to be perceived as careless or incompetent in regards to the writing process (Appelman & Bolls, 2011; Jessmer & Anderson, 2001; Schindler & Bickart, 2012), even by their peers (Varnhagen, 2000). Yet it has been suggested that children who find difficulty in the application of mechanics may find the processes of getting ideas on to paper taxing on the working memory (Graham et al., 1997), which in turn interferes with higher-level writing skills.

3.3.6 Grammar and Punctuation Summary

Historic divisions in the importance of grammar within the primary curriculum are hinged upon the various conceptualisation of grammar described within this section. If a rhetorical understanding of language can be made explicit, knowledge of grammar and punctuation should enable children both control and choice over their language use, thus putting forward a case for both explicit teaching and explicit knowledge of these writing convention skills. While a wide body of research has considered the place of the *teaching* of grammar and punctuation, to the best of our knowledge, only one peer-reviewed study of a specific writing conventions test and its relationship with writing has been published (Daffern et al., 2017). The present study seeks to further that body of knowledge through the examination of the EGPS Test, which is further explored in Chapter 5.

Exploring the skills that underpin writing through proximal and distal factors is a helpful means of understanding the writing process. An alternative means of exploring writing is through writing products.

3.4 Dimensions of Written Products

The production of writing is a multidimensional phenomenon (Arfé et al., 2016; Kim, Al Otaiba, Wanzek, et al., 2015). It has been argued that writing development cannot be fully captured with a single dimension such as writing quality, as is often the case in cognitive models of writing (Kim, Al Otaiba, Wanzek, et al., 2015; Puranik et al., 2008). The writing products that children produce in the primary classroom will conform to the goal, norms, aims and values of the class, which are in turn dictated by school policy and national curricular (Bazerman, 2016; Wijekumar et al., 2019). While we have explored the cognitive and motor

processes that contribute to writing (Berninger et al., 1992; Juel, 1988; Olinghouse, 2008), we know less about the factors involved deemed important in *product* measures (von Koss Torkildsen et al. 2015).

Early studies regarding written composition identified quality and productivity as key dimensions for assessment (Berninger & Swanson, 1994; Graham et al., 1997; Olinghouse & Graham, 2009), whereby the quality of writing products was typically operationalised through the content and organisation of information (Graham et al., 1997; Wechsler, 2005) and productivity is typically operationalised as the number of words, sentences, or ideas produced within the writing products (Abbott & Berninger, 1993; Kim et al., 2011). However, it is argued that assessing quality or fluency alone, even within the levels-of-language approach, fails to capture the multidimensional phenomenon offered through written text production (Kim, Al Otaiba, & Wanzek, 2015).

A complementary framework for conceptualising writing as a multi-level process is through distinguishing the macro-structural and micro-structural dimensions of written products (Arfé et al., 2016). In exploring children's writing through macro and micro-structural factors, dimensions within writing can be considered as dissociable factors (Puranik et al., 2008; R. Wagner et al., 2011). Macro-structure of writing products is typically concerned with the quality of a text, whereby dimensions of writing are considered together as defined through holistic or analytic rating scales for specific genres or tasks. Alternatively, writing competencies can equally be understood though a range of micro-structural measures. The most commonly employed factors which are used by researchers at a micro-structural level are productivity, accuracy, and complexity (Puranik et al., 2008). These micro-structural dimensions are a complementary framing for extant frameworks (Arfé et al., 2016) as they can offer insight into writing at each of the word, sentence, and text levels.

3.4.1 Wagner's model of the Development of Written Language

In an attempt to comprehensively model the development of written language, R. Wagner et al. (2011) offer three a priori models of writing for children between ages 6 and 10: a general model; a hierarchical model; and a specific model. Wagner's general model explores a single factor of written production and its attributable factors whereby variance from the observed measures is explained by a single factor. Next, his hierarchical model explores two factors of written production: macro-level and micro-level, distinguishing the text-level factors

from the sentence- and word-levels. Finally, his specific model explores related latent variables: macro-organisation (text-level); complexity (sentence-level; syntactic density and mean-length of T-units); productivity (word-level; the number of words and number of new words); and accuracy (word- & sentence-level; spelling and punctuation). This specific model of writing development can capture multiple dimensions of writing through dissociable products with diverging predictors (Kim, Al Otaiba, Wanzek, et al., 2015; Sénéchal, Thomas, & Monker, 1995). According to Arfe (2016), each of these dimensions are able to explore the lexical, grammatical, and discourse features of children's writing. Each of these factors of the specific model of writing are explored in turn.

3.4.2 Macro-organisation (Text/Discourse-level)

Macro-organisation is concerned with the overall adherence to expected conventions of writing structures, which are often defined by genre, and the higher-level organisation of text. Assessment of macrostructure is operationalised through bespoke scales or holistic and analytical rubrics which can vary by genre (Koutsoftas & Gray, 2012; Nelson & Van Meter, 2007). While this dimension is by far the most prevalent in writing research (see Chapter 4 for a full discussion on holistic and analytic scoring) as it allows for quick judgement of longer pieces of writing, it alone may fail to capture the various levels of language (word and sentence) which are reported to affect translation skills (Whitaker et al., 1994). Macro-structural factors are also critiqued for their subjectivity, and this is where micro-structural factors can aid researchers in understanding children's writing.

3.4.3 Micro-structure (Word- & Sentence-level)

3.4.3.1 Productivity

The first micro-structural factor is productivity, which considers measures of length of work, often measuring words, clauses or sentences produced, as well as measures of lexical diversity. With these operationalisations, productivity is a more objective measure than macro-structure. Productivity has also been found to be a robust measure of children's writing development (McMaster & Espin, 2007; Nelson & Van Meter, 2007; R. Wagner et al., 2011) which is able to differentiate children's writing achievement for both struggling and typical writers (Scott & Windsor, 2000), and typically older writers write more. While length is not the ultimate goal of writing products (Kim et al., 2014b), studies exploring productivity through exploratory and confirmatory factor analyses have identified handwriting fluency as a significant predictor (R. Wagner et al., 2011). Spelling abilities have also been found to affect

writing productivity (Graham et al., 1997; Kim et al., 2011). Finally, productivity measures have been found to be related to quality of writing products in young writers (Berninger, Nielsen, et al., 2008; Connelly et al., 2012; Graham et al., 1997; R. Wagner et al., 2011).

3.4.3.2 Accuracy

Accuracy in writing can be operationalised in a number of ways. The use of correct word sequences (CWS) (Romig, Therrien, & Lloyd, 2017), defined as any two adjacent words deemed acceptable to the context of the sample engaged in writing (Dockrell et al., 2019), is an increasingly common method of exploring the micro-structural aspect of accuracy. CWS can be explored as a proportion score of the total word sequences, or as a product of correct word sequences minus incorrect word sequences. In a novel study of writing accuracy, Mackie et al (2013) found that receptive grammar predicted writing accuracy. Inversely, where children have been asked to divert their attention from writing whilst engaging in a secondary task (serial recall/produce sentences with particular linguistic features), an increase in grammatical errors was reported (Fayol et al., 1999; Negro et al., 2005), thus resulting in a reduction in accuracy of writing products (Mackie et al., 2013). While a large body of work exists exploring productivity measures, this is not apparent for the micro-structural measure of accuracy.

3.4.3.3 Complexity

Complexity of writing products can be measured through mean length of T-units, and through measures of clause density. Clauses per T-unit, which refers to a minimal terminable unit, can be used as a measure of syntactic complexity in order to establish maturity of writing (Wagner et al., 2011). Clause density (Hunt, 1965) refers to the ratio of clauses divided by the number of T-units in the text. A higher score of words per t-unit is characteristic of more complex syntactic structures which are able to express complicated relationships among ideas (Coirier, 1996). Sentence complexity has been related to syntactic skills (Van der Lely & Christian, 2000), whereby greater syntactic knowledge leads to increased construction of complex sentences.

In a study of children with specific language impairment (SLI) (Mackie et al., 2013), receptive grammar was able to explain the greatest proportion of variance in writing complexity. Furthermore, written construction for children with SLI was inhibited by the acquisition of grammatical knowledge.

3.4.4 Summary of Dimensions of Written Products

While less widely theorised than cognitive models of writing, macro-and micro-structural features of writing offer a complementary view of children's writing competencies that accounts for the multi-dimensional nature of writing. However, of the few studies do exist around the micro-structural factors of writing, some limitations of sampling are apparent. For example, both Van der Lely & Christian's (2000) and Mackie's (2013) studies of micro-structural factors both involved small or very small samples of children identified with specific language impairment, which cannot be generalised to typical writers. Whereas Kim's (2011) study explored writers in kindergarten and so cannot be generalised to all writers in the primary years. Furthermore, Wagner et al's (2011) study of microstructural factors only deployed one writing sample, not accounting for different genres. The authors also identify that the study failed to consider the role of oral language and reading within micro-structure.

Of significance to this study of writing conventions is the role that receptive grammar has in predicting both writing accuracy and complexity, and the potential inhibiting factor that poor grammatical knowledge can have on writing. Therefore, it would be interesting to understand if declarative visually oriented grammar, as operationalised through the EGPS Test, is also able to predict writing accuracy and complexity. However, the impact of the component skills upon writing must be measured appropriately.

3.5 Overall Summary

Children need to become fluent in lower-level writing processes such as transcription and text generation to enable working memory to focus on higher-level processes (Sumner, 2013). Multiple models of the writing process for typically developing children have been presented. However, while recent models of the writing process have begun to explore the specific direct and indirect factors that impact on writing competence, current models do not consider declarative writing conventions skills such as those tapped by the EGPS Test. Greater understanding of the role of some of the writing convention skills (grammar and punctuation) may be useful in better understanding the processes involved in text generation. However, writing is not a unidimensional phenomenon and measures of quality can be subjective. Alternative measures of macro- and micro-structural features of writing support extant frameworks but enable researchers to capture the multiple dimensions of writing. The various considerations of assessing children's writing products are now explored.

CHAPTER 4: The Writing Product: Assessment of Writing

4.1 Introduction

Writing involves a range of skills and processes (Harmey, 2021), which can be assessed through various operationalisations. The assessment of written products is vital in the effective teaching of writing (Jones, 2002). When operationalised in an appropriate way, writing assessment can offer practical and conceptual support for teachers (White, 1994), helping to identify, and to provide targeted support for learners (Dockrell & Connelly, 2021). However, what was made clear through the Bew Review (2011) and its underpinning research is the potential subjectivity of the assessment of writing, particularly in the case of National Curriculum assessment tasks.

Within the primary curriculum, two methodologies for assessment are commonplace: summative and formative. Summative assessment refers to a summary of a learner's attainment through tests, examinations and end of year teacher-judgements (Looney, 2011). Conversely, formative assessment gathers information through assessment to identify needs and to adjust instruction accordingly (Mansell & James, 2009). Historically, writing assessment for research and statutory assessments has been conducted through summative assessment of a single text (Dockrell et al., 2015).

However, assessing writing is regarded as problematic as writing assessment has been described as a 'the game of reducing a complex profile of achievement to a simple number' (Richmond, Burn, Dougill, Raleigh, & Traves, 2017). Claims have been made that the assessment of writing forms the single greatest barrier to writing instruction and research (Cole, Haley, & Muenz, 1997), and Dockrell (2021) cites this difficulty as a lack of a common definition of what constitutes writing proficiency at different developmental stages. Alternatively, B. Miller and McCardle (2011) posit that research into the assessment of written products, in general, has been neglected. Yet research regarding assessment of writing is shrouded in controversy and uncertainty (Daffern et al., 2017; McMaster & Espin, 2007) and is comparatively under-researched when compared with reading. This has been attributed to the complexity of writing assessment and a lack of amenability to the scientific study of the acquisition of writing skills (De Lemos, 2002).

Questions have also been raised regarding the dimensionality of writing and how these factors should be assessed (Kim et al., 2014b). If researchers are to be able to accurately capture

the components of written products as described above, then they may better understand the cognitive factors of the writing process (Dockrell & Connelly, 2021). Furthermore, understanding of such factors could guide writing instruction and pedagogy (Mackenzie, Scull, & Bowles, 2015). Yet a writing assessment's ability to determine proficiency depends on what and how the assessment is operationalised (Schoonen, 2012). Operationalisation will subsequently vary depending on researchers' assumptions regarding the dimensionality of written composition. Currently in English primary schools, Teacher Assessment Frameworks (STA, 2016) set the dimensions that are required to meet a national expected standard in writing, although the empirical basis of such expectations has not been verified or validated through UK norm-referenced testing. Therefore, there is merit in understanding the empirical research behind the assessment of writing products.

This chapter explores how writing can be measured effectively. It proceeds to define holistic and analytic scoring, before discussing Curriculum Based Measures (CBM). Then the importance of genre is discussed and methods of collecting writing samples are offered.

4.2 How can writing be measured objectively?

Effective assessment is rooted in the identification of core dimensions through developmentally sensitive means (Dockrell et al., 2015). Therefore, teachers and researchers exploring writing must make conscious decisions about the dimensions of writing which are to be evaluated (Sadler, 2013). Understanding what and how to assess within writing is fundamental in the understanding of children's writing proficiency, and as there is no consensus on what constitutes writing proficiency at different points in writing development, decisions around appropriate assessment of writing at the end of Key Stage 2 are not asserted by an agreed definition of what competent writing looks like for an 11-year-old.

As writing is not unidimensional, but rather a number of dissociable dimensions, writing assessments are reliant upon sensitive measures of writing competencies which are able to capture characteristics of proficient written products (Dockrell et al., 2019). Assessing a single element of writing products has also been found to fail to capture the realistic demands of differing writing tasks (Scott & Windsor, 2000) and so differing methodologies seek to operationalise writing through different component skills.

Researchers and teachers have employed a number of approaches in an attempt to capture writing competency. Each of these approaches vary in their methods of exploring

assumptions around the various dimensions of writing, yet all explore competency at either word-, sentence-, and text-levels, or a composite of all three, through collections of children's writing. Once collected, children's writing is open to evaluation in a multitude of ways, typically falling into one of two categories: holistic or analytic, each of which are explored in turn.

4.2.1 Holistic Scoring

Holistic scoring is a measure of a rater's overall judgement of a written product (Huot, 1990) and is based on the assumption that writing is unidimensional (Kim et al., 2014b). This global quality scoring of texts is a widely used measure of writing for teachers, researchers and psychometric assessments (Dockrell et al., 2015), and it is this form of assessment that is typically used for macro-structural assessment or measures of global writing quality. Holistic measures seek to measure multiple aspects of writing, often employing a rubric, but then to assign a single score that captures global performance (Frey, 2018). A significant benefit of holistic measures is that they enable relatively quick scoring of texts (Dockrell & Connelly, 2021), and holistic measures can be helpful at quickly identifying children at upper and lower ends of attainment on the writing spectrum. However, little is learnt from a single score about the majority of students who perform variably across the numerous criterion of generalised writing criteria (Frey, 2018), with limitations in differentiation between changes over time and performance in the various performance domains in writing. It has also been argued that it is unclear what holistic rubrics actually assess (Olinghouse, Santangelo, & Wilson, 2012) and, depending on genre, holistic quality is a poor predictor of paragraph, sentence, or vocabulary abilities (Olinghouse et al., 2012). Holistic measures are reported to be susceptible to unintentional overweighting of particular genre-specific features of writing, rather than general components such as sentence or paragraph construction (Banerjee, Yan, Chapman, & Elliott, 2015; Schoonen, 2005), which is unsurprising in an assessment measure which assumes that the whole writing product is more than the sum of its parts (Myers, 1980).

Even after support from training and benchmarking resources, there remain concerns regarding the reliability of holistic scoring. Meta-analysis of holistic scoring has shown that over half of the studies of holistic writing assessments examined demonstrated less than an acceptable .80 inter-rater consistency score (Graham, Hebert, & Harris, 2011). Holistic assessments limited ability in differentiating between performance on particular components of writing (Espin et al., 2000) is of particular concern when seeking to understand the writing

process, and this approach has been suggested to limit both theory development and the identification of skills that writers need to improve to achieve competency in writing (Dockrell & Connelly, 2021).

In English primary schools, previous statutory assessment frameworks where teachers have marked student's written products have been reported to be prone to inaccurate or incorrect scoring of up to 30% (Doyle & Godfrey, 2005). There are questions to be raised regarding the validity of such assessments, as validity is dependent on teacher judgements being reliable. Currently, in English primary schools, teachers are required to carry out a holistic assessment of writing in both Key Stages 1 and 2, using the Teacher Assessment Frameworks. Teachers assign a qualitative banding based on a scoring grid (see Appendix 5) to a portfolio of written products as a form of statutory summative assessment. While there is merit in the use of multiple writing products across a variety of genres (Bouwer, Béguin, Sanders, & Van den Bergh, 2015; Scott & Windsor, 2000), the statutory national framework makes it clear that 'The frameworks are not a formative assessment tool: they are not intended to guide individual programmes of study, classroom practise or methodology'(STA, 2016:2). Herein lies possibly the greatest limiting factor for holistic scoring. The production of a single score or statement within holistic scoring results in an inability to distinguish the various components required of a competent writer within their writing products, thus limiting practitioners in providing suitable intervention or researchers in generating suitable theory. An alternative to holistic scoring, which addresses a number of the critiques around holistic scoring, is analytic scoring.

4.2.2 Analytic scoring

Analytic scoring is an alternative means of evaluating written products through the assignment of separate scores for predetermined dimensions of a written task which are deemed to be features of quality writing (Espin, Weissenburger, & Benson, 2004). Analytic scoring has been found to provide validity in judging the complex competencies of writing products (Frey, 2018) and analytical assessment of writing has been recognised as more reliable than its holistic counterpart due to its focus on specific writing abilities (Hayes et al., 2000). By identifying comparative strengths and weaknesses in children's writing, analytic measures provide useful diagnostics, enabling differentiated instruction based on the relative weaknesses of each learner (Dunsmuir et al., 2015). Unlike holistic scoring, analytic offers high levels of inter-rater agreement when using analytic scoring rubrics (Espin et al., 2004; Frey, 2018) and analytic scales have been found as easier method to train raters when compared with holistic measures,

owing to the specific features to be identified in writing products (A. Cohen, 1994). In turn, the specific nature of analytic scoring provides a useful means of providing formative feedback in order to improve future writing performance (Fathman & Whalley, 1990). In addition, the rating of individual competencies in writing through analytic scoring increases the generalisability of assessment tasks (Frey, 2018).

Early work on analytic scoring from Diederich (1974) offers six analytic dimensions of quality writing: ideas, organization, spelling and conventions, wording, style, and handwriting. With these as a basis, numerous alternative analytic measures have been useful in better understanding both the writing process and writing products. Yet while the benefits of analytical scoring are evident, few tests of this nature are based on norms from the UK population (Dunsmuir et al., 2015). It is also worth noting that often the different dimensions by which writing is assessed on scoring rubrics are often highly correlated, thus not necessarily offering independent data about each different dimension (McMillan, 2013). Additionally, it is reported that construct validity of the varying dimensions used in analytical measures is often weaker than that of many holistic measures (Gansle, VanDerHeyden, Noell, Resetar, & Williams, 2006; Weissenburger & Espin, 2005). Analytic scoring can also be susceptible to the halo effect, whereby ratings of one domain influence scores of another (Lai, Wolfe, & Vickers, 2015), thus reducing the independence of each of the domains.

4.2.3 Curriculum Based Measures of Writing

Curriculum based measures of writing (CBM-W) is a means of measuring a writer's competency through direct assessment of particular writing skills. This sensitive measure of children's written text production (Espin et al., 2000) is rooted in established assessment practices in reading and mathematics (Stecker, Fuchs, & Fuchs, 2005). CBM writing assessments were designed to offer global indicators of children's writing performance (Deno, Marston, & Mirkin, 1982) and have been shown to be reliable and valid (Dockrell et al., 2015; Gansle et al., 2006; Kim, Gatlin, Al Otaiba, & Wanzek, 2018; McMaster, Du, et al., 2011; McMaster & Espin, 2007). In these tasks, children are typically required to write for brief periods (3-7 minutes) in response to a probe or prompt (Coker Jr & Ritchey, 2010; McMaster, Du, et al., 2011; Weissenburger & Espin, 2005). Written products are then evaluated on measures of productivity, accuracy, and complexity. Productivity measures can consist of total number of words written; correct word sequences; and total number of punctuation marks. Accuracy measures include incorrect word sequences or derived scores such as correct word

sequences divided by total number of words written (%CWS), or correct word sequences minus incorrect word sequences (CIWS). Complexity measures typically explore mean length of t-units (K. Wagner et al., 2019). While appropriateness of these measures varies with age, the use of the variety of measures within CBM-W offer face validity (see Chapter 6) for teachers and researchers (Dockrell et al., 2015).

Yet CBM-W has been criticised as a measure, due to difficulties in scoring as well as its inability to measure macro-structural aspects of writing. Furthermore, limited research has examined CBM-W within English primary schools. However, this measure and approach to assessing writing is of particular use to this study as it enables the examination of writing products at word and sentence levels, and CBM-Ws ability to examine 'connected-text' (Kim, Al Otaiba, Wanzek, et al., 2015: 3) enables grammatical accuracy to be captured within indicators of correct word sequences. In particular, the %CWS accuracy measure offers strong validity evidence (Jewell & Malecki, 2005; McMaster & Espin, 2007; Weissenburger & Espin, 2005), especially for children in upper primary (Lembke, Deno, & Hall, 2003). Finally, studies using CBM-W have been able to reliably differentiate between narrative and expository texts (Apel & Apel, 2011; Dockrell et al., 2015; Koutsoftas & Gray, 2012; Scott & Windsor, 2000), and pupils have been found to produce less text and to be less accurate in the expository genre compared with narrative writing (Dockrell et al., 2015). Therefore, a fuller explanation of the role of genre in writing assessment is now offered.

4.3 Genre and Assessment

A further criticism regarding the cognitive models of writing is that writing draws upon and requires individuals to produce an indeterminate number of text types and genres (Bazerman, 2018). What is currently regarded as competency in children's writing can only be generalised to the context of the genres taught in school-writing, as 'processes are … tied to the target product, and how that product will be used' (Bazerman, 2018:303). Within educational institutions such as primary schools, there is significant variance in expectation, procedure, and genre of writing (Carroll, 2002; MacDonald, 1994). Even at class level, it can be argued that individuals will approach writing with variance in the strategies that they have learnt (McCarthy, 1987), regardless of the often well-defined criteria offered through national curricular (Herrington & Curtis, 2000). As this study does not seek to dispute Bazerman's assertions, it must assert that any relationships found between grammar and writing are situated within the writing products of typical primary school-writing.

Genre can be defined as a set of social practices that enable us to achieve goals (Derewianka, 2003; Martin, 1985; Painter, 2001). Derewianka (2003: 135) explains that the varied and numerous situations that we engage in within daily life 'involve predictable and recurring patterns of language use'. Within our own social contexts, we begin to identify the generic expectations of each situation and we respond in the appropriate genre as we seek to shape discourse through varied communicative goals and functions (Grimshaw, 2003; Paltridge, 1997). In the context of writing, writers use knowledge of text and schema, alongside patterned language, to meet the needs of a writing goal within specific genres.

Common genres within the primary school context are narrative and expository writing (Dockrell et al., 2015). Narrative texts are agent-orientated and explore events through a temporal framework (Berman & Katzenberger, 2004; Berman & Nir-Sagiv, 2007). Narratives tell a story and seek to form discourse which conveys a series of events. These events may or may not be fictional and fit within a spacio-temporal context (Peterson & McCabe, 1994; Polanyi, 1982). There is an increasing body of literature which demonstrates support for the link between written narrative ability and academic performance (Boudreau, 2008). Narratives based on personal experience are regularly produced through informal dialogue and use of everyday language structures (Ravid & Katzenberger, 1999) which may make them easier for children to produce in comparison to other genres. By contrast, expository texts are topic-oriented and explore concepts and issues through the development of arguments and the sharing of ideas (Britton, 1994; Katzenberger, 2005). Through expository texts, children must be able to write in an accurate and clear manner in order to inform and persuade others (Graham, Harris, & Hebert, 2011), and as such, mastery of expository writing takes longer (Berman & Nir-Sagiv, 2007). Berman (2007) argues that the requirements of complex vocabulary make expository writing more challenging for younger writers, and such challenges are evident in student's texts (Beers & Nagy, 2011; Koutsoftas & Gray, 2012). In expository writing, children produce shorter, more error-prone texts as they attempt to use increasingly complex text structures (Scott & Windsor, 2000). Expository writing is rarely dialogic and is rarely produced outside of the academic, educational, or professional spheres where products are expected to adhere to the 'literate lexicon' and written-like language (Blank, 2001; Ravid & Katzenberger, 1999; Scott, 2004).

Olinghouse, Graham, and Gillespie (2015) have also identified that *knowledge of genre* makes a contribution to the prediction of writing outcomes, and research has suggested that

children are required to reach a certain level of competence within each genre before using productivity as a predictor for writing outcomes (Bereiter & Scardamalia, 2013).

While both holistic and analytic assessments can provide information about some genre-specific elements of writing products (Olinghouse et al., 2012), this requires the use of genre-specific criterion, as what is a strength in one genre can be considered a weakness in others (Brien, 2012). Therefore, this should be addressed either through genre-related assessment criterion, or by only exploring factors which not influenced by genre (e.g. micro-structural factors).

More specifically to the narrative genre, researchers have been challenged when seeking to create ways of assessing narrative writing which demonstrates reliability and validity (Rezaei & Lovorn, 2010). This has presented difficulties in relation to holistic based scoring of writing as used by teachers for statutory assessments in England (Espin et al., 2004; Hayes, 2000; D. Miller & Crocker, 1990). It is argued that the most appropriate means of assessing writing proficiency should not simply be through one, but multiple texts across genres. Bouwer et al. (2015) offers that only 10% of variance in children's writing products is related to individual skill, with genre and task accounting for a greater proportion of the variance in writing quality. Furthermore, judgements of writing vary not only by individuals writing skills, but also due to differences between tasks and raters (Huot, 1990).

Regardless of holistic or analytical methodologies, the assessment of a single writing product can lead to failure in capturing the unique demands of the various genres of writing (Bouwer et al., 2015; Scott & Windsor, 2000). Writing performance has been found to differ substantially between the various writing tasks carried out in schools (Reed, Burton, & Kelly, 1985; Schoonen, 2005; Van den Bergh, De Maeyer, Van Weijen, & Tillema, 2012). When considering writing tasks within genre, Scardamalia and Bereiter (1987) have identified that younger writers find transformation of ideas and knowledge for specific audiences challenging. As surmised by Bouwer et al. (2015: 95) 'writing about oneself is easier than writing for someone else'. There is not agreement regarding the root of this challenge, although Bouwer et al. do suggest that this may result from topic knowledge or task familiarity. Therefore, choices around collection of writing samples are a vital consideration in the exploration of children's writing and this is now explored more fully.

4.4 Collecting Samples of Writing

In addition to the different measures which can be used to understand children's writing competencies, the most effective methods of collecting writing samples for the various forms of analysis has failed to reach consensus (Puranik et al., 2008). Without careful consideration, prompts which are not facilitative have the potential to restrict compositional fluency, thus limiting quality of children's writing products (R. F. Hudson, Lane, & Mercer, 2005). Prompts can vary in format, differing at word-, sentence-, and text-levels. Moreover, the options available to researchers for text-level prompts are numerous, with researchers drawing on tasks such as spontaneous story generation (Nelson & Van Meter, 2007) or story generation based on a stimulus (Pearce, 2003).

The nature of the task selected for writing assessment can impact the additional constraints which children must manage when producing writing products. For example, where children are required to produce texts spontaneously or generate ideas in an unknown context, idea generation can become an inhibiting factor, particularly for those with language impairments (Puranik et al., 2008). Alternatively, when texts involve personal retelling, additional burdens are placed on working memory. As with any means of assessment, trade-offs must be made by researchers to tap the dimensions of interest within research.

When exploring the role of writing conventions among the array of additional contributing factors in text production, it is important to reduce potential confounds to the text generation process. Therefore, in the collection of samples of writing within this study, personal retellings within both narrative and expository genres will reduce limitations within idea generation, enabling greater exploration of the impact of explicit writing convention knowledge on children's writing products. Even though it is acknowledged that personal-retellings can constrain memory load, the literature suggests that this constraint is less significant than those created by transcription skills (Graham et al., 1997).

4.5 Overall Summary

It has been demonstrated that writing can be assessed in several ways. What is clear from the literature is that if a researcher wants to capture the multifaceted nature of writing at a single point in time, then a variety of measures involving differing genres and tasks that capture the multiple dimensions of writing (productivity, accuracy, complexity) is most appropriate. Furthermore, if any given study is concerned with exploring the cognitive

processes involved in translation within writing, personal retelling products are least likely to place additional burdens upon the cognitive processes within primary aged children. However, there is little evidence that the EGPS Test is a reliable measure of writing, yet it is often used as a proxy for writing. Of particular interest to this thesis is the impact of grammar and punctuation skills as tapped by the EGPS Test upon the various measures and dimensions of writing mentioned above. Therefore, the EGPS Test is now explored in greater detail.

CHAPTER 5: The English Grammar Punctuation and Spelling (EGPS) Test

5.1 Introduction

The English Grammar, Punctuation and Spelling Test (EGPS) is a statutory assessment of writing conventions, carried out in the final year of English Primary education. According to the Standards and Testing Agency (STA), the purpose of the EGPS Test is to 'ascertain what pupils have achieved in relation to the areas of the national curriculum (2014) describing grammar, punctuation and spelling' (STA, 2015: 4). However, the test has been criticised as an inappropriate proxy for writing assessment (Myhill, 2016), yet only one limited empirical evaluation of the relationship between the EGPS Test and children's writing is currently available (Canniford, 2019).

Within this chapter, the historical context of grammar in the English national curriculum offers an initial rationale for the introduction of the EGPS Test. A full exploration of the EGPS Test is then offered, exploring its purpose, domains, and response formats. Critiques of the EGPS Test are presented and discussions around using the EGPS Test as a proxy for writing are considered.

5.2 Pre-EGPS Test: Historical Context

When considering the use of writing conventions within the context of national policy, grammar, punctuation, and spelling have long been influenced by strong societal pressures for correctness, which is dominated by history and convention (Wyse et al. 2013). Since the early 19th century, politically charged discourse around the role of writing conventions in school curricular has loomed over the teaching of each of the domains of grammar, punctuation and spelling (Carter, 1990; Education, 1921; R. Hudson & Walmsley, 2005; Watson, 2015), where explicit grammar knowledge was believed to be too difficult for children to learn (Macauley, 1947). Writing conventions were considered both irrelevant and boring within the progressivist education movements of the 1960s and 1970s (Lefstein, 2009), and even when acquired, implicit acquisition of grammar awareness was believed to render explicit knowledge useless (Elbow, 2002). Furthermore, the national council of English Teachers declared that there is limited correlation between grammar teaching and good writing (Wyse, 2006), and so the anti-grammar zeitgeist of the 1970s led to the demise of both teaching and research of grammar and

language, which was reported as 'non-existent' within the anglophone universities that were training primary school teachers (R. Hudson & Walmsley, 2005).

A renaissance proceeded during the 1980s and 1990s, during which an interest in writing conventions developed. This overturned the pedagogical trends of the previous decades, and this period has been termed as the 'rebirth of grammar teaching' (Hudson & Walmsley, 2005:594). However, this rebirth was not greeted openly across the sphere of education, and the reintroduction of writing conventions was viewed as controversial (Lefstein, 2009). With the commissioning of the Kingman Report (1988), the need for an emphasis on 'Knowledge about Language' (KAL) was highlighted. KAL can be defined as a pedagogical approach to language teaching rooted in meaning and context, which is aligned with descriptive grammar (Myhill and Watson, 2014). Yet, a tension existed between the findings of this report and the government of the time which reported a desire to re-introduce increasingly formal teaching of grammar (Abrams, 1991). Blenkin et al (1992: 8) ascribed this conflict between researchers and government to fear of change, offering: 'for some, change – and especially social change – has been viewed as a process of deterioration from a kind of golden age of perfection, change is the process by which things get worse rather than better'. However, through the 1988 Education Reform Act, the controversial imposition of prescriptive grammars as an element of the English curriculum brought grammar back into primary schools across England.

In 2011, The Bew Curriculum Review (Bew, 2011) recommended the removal of the Writing Standard Assessment Tasks (SATs) at the end of Key Stage 2 (KS2; upper primary). With the identification of misclassification rates within the results of English levels reported in Year 6 Writing SATs as high as 32% (Wiliam, 2000), the Bew Review described significant concerns with the writing tests of the time, citing the inconsistency and subjectivity of the assessments. Hereafter, the introduction of a draft National Curriculum brought about an end to the writing SATs in 2012. Instead, a test of English grammar, punctuation and spelling (EGPS) for children in KS2 was introduced. This was in line with the recommendations of the Bew Review, citing the successful and reliable use of spelling and grammar tests internationally. The EGPS Test became a statutory assessment for English primary schools in 2013, in conjunction with a teacher-assessed holistic writing assessment which would be moderated by local education authorities. Subsequently in 2014, further development in policy and the introduction of National Curriculum 2014 considerably raised the demands upon

teachers' and pupils' grammar knowledge. For the first time in the history of the English primary curriculum, a glossary of grammar terms was 'enshrined in statute' (Bell, 2016: 149), placing strains on teacher's subject knowledge and confidence. By 2016, an even more challenging EGPS Test was introduced in-line with the newly increased demands of National Curriculum 2014.

5.3 What is the EGPS Test?

The English Grammar, Punctuation and Spelling test (EGPS) is a KS2 SAT of writing conventions. The test is administered in Year 6, the final year of primary education in England when children are typically 10-11 years old. The test is produced by the Standards and Testing Authority, an executive branch of the Department for Education, and a new test and mark scheme is produced each year (STA, 2017c). The test population for all EGPS Tests includes all pupils registered in maintained schools, special schools and academies in England, and exemptions or special arrangements are detailed fully for test administrators.

The EGPS assessment of the EGPS domains as presented in the national curriculum consists of two components, which are presented to participants in two test papers of visually presented decontextualised tasks. Paper one covers seven grammar and punctuation content domains (STA, 2015) as can be seen in Table 1.

G1	Grammatical terms/word classes
G2	Functions of sentences
G3	Combining words, phrases and clauses
G4	Verb forms, tenses and consistency*
G5	Punctuation
G6	Vocabulary[#]
G7	Standard English and formality

* G4 Verb forms, tenses and consistency includes elements of word-level grammar as seen within inflectional morphemes, which can change what a word does in terms of grammar but does not result in a new word

[#] *G6 Vocabulary* refers to word-level grammar as seen within morphology, synonyms and antonyms, and word families. This includes derivational morphemes, whose use can create new words/vocabulary.

Table 1- Paper 1 EGPS Content Domains – STA, 2015

The official testing guidance for the EGPS Tests, provided by the STA (2015), explains that the content domains ensure testing coverage of the relevant elements of the National Curriculum for English in Key Stage 2, drawing on both the programme of study for English, as well as the statutory appendices, which dictate the compulsory elements of the National Curriculum.

Paper one offers up to 50 marks (the number of questions varies year on year) and consists of four response formats, which range from a closed response to an extended response, as can be seen in Table 2. Paper 2 presents participants with 20 spellings, which are administered aurally within 20 contextualised sentences. The spelling test covers the spelling content domains from the National Curriculum Spelling Appendix (see Appendix 6) and aims to test the spelling patterns taught throughout the National Curriculum 2014 included in the Spelling Appendix.

Table 2 – Paper 1 EGPS Response Strategies – STA, 2015

Response format	Selected response	Constructed response: data transformation	Constructed response: prompted	Constructed response: independent
Explanation of format	Selecting the correct response or identifying a feature from a given field of data	Transforming a given word, phrase or sentence	Inserting a word or phrase within a given target sentence, following a specific prompt	Open response, without a prompt or frame within which to write

Table 3 highlights the required distribution of marks for each content domain in the EGPS Test of any year. The data in brackets shows the specific number of marks allocated for each content domain in the 2018 test used in this research. Tests are designed based on number of marks, rather than questions. The number of questions varies from year to year as some questions may award multiple marks. Up to 50% of the combined EGPS score results from the grammar domains (with vocabulary consisting of word-level grammar), with around 20% for punctuation and 30% for spelling.

Table 3 - Profile of marks by content area – STA 2015; 2018

Paper	Content domain reference	Number of marks	Percentage of total marks
Paper 1: questions	grammar	25-30 (30)	36-50% (43%)
	vocabulary[#]	3-7 (5)	4-10% (7%)
	punctuation	10-20 (15)	14-29% (21%)
Paper 2: spelling	spelling	20	29%
	Total	70	

[#] Vocabulary refers to word-level grammar as seen within morphology, synonyms and antonyms, and word families.

To maintain consistency for year-on-year comparisons of performance, the EGPS Test translates raw scores into a scaled score. At the extremes of the scaled score distribution, scores are truncated to minimise potential confounds where the test is unable to measure optimally. Test scores are finally allocated one of the following performance descriptors which are reported to local education authorities to assess attainment and inter-key stage progress: working towards the expected standard; working at the expected standard; working at greater depth within the expected standard.

In English primary schools, children are explicitly taught writing conventions. The testing of the EGPS domains through the EGPS Test requires children to demonstrate declarative knowledge of writing conventions to be successful within closed and extended visually oriented responses. The EGPS Test is a potentially suitable tool for children to demonstrate declarative knowledge of the writing convention skills of grammar, punctuation, and spelling.

5.4 Critiques of the EGPS Test

Professionals from within primary education (Safford, 2016), as well as critics external to primary classrooms within the media, have since contested the role of grammar within the curriculum and have been vociferous regarding the EGPS Test (Mansell, 2017). Proponents such as Crystal (2013:1), who explicitly states that the EGPS Test 'turns the clocks back half a century', clarifies that a detrimental emphasis is now placed on metalinguistics, rather than using grammar as a starting point for effective writing. Furthermore, ex-Children's Laureate

Rosen (2015) argues that the EGPS Test is shrouded in 'terminology-itis', an ill comprising of three symptoms: lack of universal agreement on content; lack of evidence as to whether coherence and effectiveness in children's writing comes from metalinguistic knowledge; and time spent on teaching metalinguistics, rather than teaching children to write coherently and interestingly. But in addition to political discourse, theoretical debates over the inclusion of the teaching of grammar in the classroom have endured long-standing discord (Andrews et al., 2006; Braddock et al., 1963; Andrews et al., 2004; Hillocks Jr and Smith, 1991; Myhill and Watson, 2014; Sheard et al., 2012; Wyse, 2001; Weaver, 1996), with the role of grammar in research being dominated by 'a desire to prove that it "works" or not, rather than by a desire to understand the nature of children's learning and the multiple complexities of any individual classroom context' (Myhill, 2021: 265).

Myhill and Watson (2014) argue that this very focus on the exclusion or inclusion of grammar within the curriculum has resulted in a lack of developed articulation or research regarding the actual contribution and impact that writing convention knowledge and understanding might have on children's writing. At the time of writing this study, the only available published peer-reviewed research regarding the EGPS Test has reported that the test has led to an increase in decontextualised teaching of grammar (Safford, 2016). However, no studies have actually addressed the *impact* of increasing knowledge of grammar, punctuation, and spelling upon children's writing processes or products. It is clear that learning about grammar remains under-theorised (Wyse, 2001). While this is true of wider understanding about the teaching of grammar, it is also true of the EGPS Test and the role of writing convention skills, especially grammar and punctuation. Yet to date, the EGPS Test is still used commonly as a proxy for writing attainment in statutory reporting within primary schools.

5.5 EGPS as a proxy for writing assessment

In 2017, one of the original four authors of the EGPS curriculum called for the EGPS Test to be discontinued, citing a lack of evidence regarding the validity of the content domains to improve the use of written language (Mansell, 2017). In earlier evidence submitted to the Education Select Committee (2016), Myhill stated that:

> 'as a proxy for assessing children's writing, it is wholly inappropriate: children can identify errors in isolated tests but far fewer can transfer that learning to their writing, so it does not provide any information about accuracy in writing'.

While Myhill's (ibid.) argument regarding children's ability in isolated tests and the limited transferability of learning into writing is understandable, her assertation that such assessments do not provide any information about children's accuracy in writing is empirically unfounded. In fact, writing convention assessments, whereby respondents are required to identify a correct response or an error in a passage, have been found to be successful in enabling judgements of accuracy in writing (Largy, Dédéyan, & Hupet, 2004) and in predicting writing competence (Daffern et al., 2017).

It could be argued that the EGPS Test is a measure of how well a child is able to apply their knowledge of writing conventions. Yet, what is of concern in relation to this research, and to our understanding of writing and grammar, is what effect does knowledge of writing conventions have on children's written products and what constraints can writing conventions place within the writing process. To our knowledge, only a single study has explored the transferability of knowledge of writing conventions into writing quality (Daffern et al., 2017). In their novel study, it is identified that there is a relationship between children's ability to identify or label disassociated writing conventions and an individual's capacity to coordinate the vary demands of the writing process required to produce well-written text. In addition, recent research into the convergent validity between EGPS scores and children's scores on a standardised writing assessment (Canniford, 2019) has shown strong convergent validity between the EGPS Test 2017 and writing quality. However, using such tests as a proxy for writing remains to be heavily criticised (Dockrell & Connelly, 2021) and is largely unsupported as writing is a complex process which is difficult to reduce to a single score.

One area of study that is also currently not understood is the validity (convergent or content) of the spelling paper of the EGPS Test. As far as we are aware, no studies have taken place to explore the relationship between *EGPS Paper 2: Spelling* and norm-referenced tests of spelling. As mentioned in Chapter 2, spelling is typically measured through a standardised assessment of single words in dictated sentences. This is the same method for EGPS Paper 2 – Spelling. If the EGPS spelling were to demonstrate convergent validity with existing standardised measures of spelling, then it would be reasonable, based on the extant literature mentioned within this chapter, to expect the EGPS to predict writing productivity.

5.6 Overall Summary

The introduction of the EGPS Test is rooted in a long history of fiercely held views between researchers and politicians. Yet opinions have often been based on 19th century attitudes rather than empirical research (Carter, 1990). Until further exploration takes place into the relationship between the EGPS Test and writing outcomes, judgements regarding the validity and accuracy of the EGPS Test as a proxy for writing remains to be folk theorisation. Where a relationship between writing conventions and writing outcomes has been identified through writing convention tests, further exploration should take place to explore the dependability of the EGPS Tests as a measure of writing and writing conventions. If the EGPS can measure writing conventions in English, and writing conventions are found to play a significant role in the writing process, then this will have potentially significant implications across anglophone countries

CHAPTER 6: EGPS Validity and Reliability

6.1 Introduction

This thesis seeks to understand the relationship between the EGPS Test and children's writing outcomes. Reliability and validity are an essential factor in the exploration of this relationship. This chapter begins by exploring terms around dependability. It proceeds to offer a discussion around concepts of reliability and validity. Finally, issues of validity and reliability regarding the EGPS Test are explored.

6.2 Dependability

When considering the EGPS Test, it is helpful to consider how reliable and valid it is as a measure. While dependability of a measure offers no actual statistical value, it is a helpful term in describing the purposefulness of a measure, and it is through the intersection of reliability and validity that we define the unifying term: dependability (Wiliam, 1993). In seeking out the optimum dependability of a measure, equilibrium must be reached between increasing reliability and decreasing validity.

When describing the stages required in investigating the dependability of an assessment, Fuchs (2004) offers three stages required in establishing an evidence base. Stage 1 is to establish the static technical features of the assessment – reliability and validity. Stage 2 seeks to establish the sloping technical features of the assessment. Stage 3 seeks to establish the usefulness of the assessment in the classroom context. Owing to the lack of research into the EGPS Test, this thesis seeks to explore the EGPS Test through stage 1 of this theoretical model, exploring reliability and validity from a single data point.

6.3 Reliability

Within social sciences, reliability refers to the consistency of a measure of a given concept (Bryman, 2016; L. Cohen, Manion, & Morrison, 2013). Reliability ensures that research is acceptable within the research community (Biggam, 2015). In the quantitative research paradigm, reliability is of upmost importance for the researcher as it demonstrates whether research and its findings are repeatable (Bryman, 2016). In ensuring that a measure offers both precision and accuracy, key factors in the discussion of reliability include stability, internal consistency and inter-rater reliability.

Stability refers to the consistency of a measure, either over time or over similar sample groups. Any administration of a given measure should produce comparable data from similar participants if undertaken within an appropriate timeframe (L. Cohen et al., 2013).

Internal consistency refers to the coherence of a measure which consists of multiple items. Where replication of a measure is not possible (test-retest reliability), split-half methods can be utilised to determine reliability as it allows for a multi-item task to be administered in full in a single administration (Frey, 2018). Cronbach's alpha and Spearman-Brown split-half coefficients can test the internal consistency of the measure. The general acceptability level of internal reliability is above 0.8 (L. Cohen et al., 2013).

Equivalence is comprised of two elements: equivalent forms and inter-rater reliability. The ability of similar instruments to yield similar results is defined as equivalent form. The measurement of equivalent forms can be explored through difference of means in the form of paired samples t-tests (Field, 2009). Alternatively, inter-rater reliability explores the consistency and dependability of raters when making subjective judgements within a measure (Bryman, 2016). Where low levels of inter-rater agreement are apparent in research, this may be indicative of weakness in testing procedures or coding methods. In the context of writing, it is widely accepted that assessing writing and making secure judgement of writing products can be highly subjective where inadequate guidance or scoring procedures are provided (Bew, 2011; Graham, Bollinger, et al., 2012). Therefore, within writing, achieving high inter-rater agreement can be challenging. A number of statistical methods can be used to establish inter-rater reliability. While the most common method is percentage agreement (Robson & McCartan, 2016), this method is critiqued as it does not account for any agreement that has arisen by chance (Lavrakas, 2008). Considering this together with the known subjectivity of writing assessment, achieving absolute agreement when assessing writing is problematic. Intraclass correlation coefficient is a statistical method that can be used to establish reliability which accounts for variability between individual raters' scores for each participant in relation to total sample variability (Laerd, 2018).

6.4 Validity

Validity explores to what degree an instrument can measure a given concept (Duff, Mengoni, Bailey, & Snowling, 2015; Wiliam, 2001) and to what extent theory and empirical evidence

support interpretations of that measure (AERA, 1999). The specifics of validity that are applicable to this study are now further explored.

Face validity considers if an instrument truly reflects the intended content that it seeks to measure. This form of validity can be confirmed by experts within the field that the instrument seeks to measure (Gray, 2014). Alternatively, content validity confirms the ability of an instrument to tap the domains it purports to measure (Carmines & Zeller, 1979).

Defined as the 'unifying concept of validity' (Messick, 1980: 1015), *construct validity* captures the extent to which an instrument can measure given theoretical constructs (Brown, 1996; Cronbach & Meehl, 1955). Messick explains that construct validity is 'an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores' (ibid). In the examination of construct validity, proposed interpretations of given measures are elaborated upon by defining their scope through the delineation of the particular aspects that the measure seeks to represent (AERA, 1999). Construct validity can be further categorised as *convergent* or *discriminant validity*. Convergent validity confirms whether or not measures of the same theoretical construct correlate highly (Colman, 2015). Alternatively, discriminant validity confirms that measures of differing theoretical constructs do not correlate highly.

The measure of *criterion validity* explores a given measure's ability to predict real-life performance by exploring strength of association with an accepted measure of a concept (Gray, 2009). In the case of this research, criterion validity would explore the associations between the EGPS Test and a number of criterion measures (e.g., WIAT-II; CBM-W), quantitatively measuring the relationship, which provides a correlation coefficient (r value).

Validation of the EGPS Test can be viewed as developing a scientifically sound validity argument to support the intended interpretation of EGPS Test scores and their relevance to the proposed use (AERA, 1999). In the case of the EGPS Test, face validity would reasonably require the test to measure attainment in grammar, punctuation, and spelling. Content validity of the EGPS Test is reported through the domain information (STA, 2015). Criterion validity would explore the relationship between the EGPS Test and other established measures which demonstrate content validity with writing.

However, whether the EGPS Test offers construct validity is a more complex question depending on its intent. If the purpose of the EGPS Test is to purely offer an attainment score

of grammar, punctuation, and spelling for the domains prescribed in the test manual, then it is likely to offer convergent validity. However, if its purpose is to act as a proxy for writing, then the validity of the measure is questionable. In the context of exploring the relationship between the EGPS Test and children's writing outcomes, our concern is whether the measure of the content domains included in the EGPS Test can reasonably capture the multiple constructs of writing. The EGPS Test could be deemed as offering construct underrepresentation (AERA, 1999), as the test construction offers only closed answers, and it is currently unclear whether EGPS knowledge is transferable in writing.

This challenge can be operationalised using several already accepted conceptual measures of the theoretical construct of writing, in the various forms described above (WIAT-II^{UK} Written Expression Subtest; CBM*).* In a prior study of EGPS and writing by Canniford (2019), a strong correlation was established between EGPS scores and measures of sentence combination. Moderate correlations were established between paragraph production and measures of vocabulary, thus demonstrating convergent validity between the EGPS Test and children's writing. However, as we have explored above, writing is multidimensional and so a more comprehensive study should look at writing as a product at word-, sentence- and text-level to establish the validity of the EGPS Test.

6.5 Overall Summary

The first step in investigating the dependability of a measure is the exploration of reliability and validity. Construct and criterion validity are two measures of validity that are of particular importance to this exploration of the relationship between the EGPS Test and writing. Where moderate correlations have previously been established between EGPS and writing, this should be further explored using a variety of constructs of writing to address the multi-dimensionality of writing.

CHAPTER 7 : Aims of the Study

7.1 Introduction

The review of the literature has provided an overview of the theoretical underpinnings of the writing process and has established the importance of distinguishing between writing process and product. What is clear is that there is a lack of empirical research regarding the place of two of the writing conventions captured by the EGPS within the writing process: grammar and punctuation. At the time of publication, only one study (Canniford, 2019) was available regarding the relationship between the EGPS Tests and its link to writing outcomes.

7.2 Aims

The principle aim of the present study was to examine the potential use of the EGPS Test of writing conventions as a means of evaluating writing quality for children in Year 6 in English primary schools. From the review of the literature, it has been reasoned that writing performance can be operationalised and assessed at word-, sentence- and text-levels, through quality, productivity, complexity, and accuracy. This acknowledges the multi-dimensionality of writing and enables us to establish in which ways the EGPS Test demonstrates a relationship with writing outcomes.

This study extends the findings from my IFS research, offering a more comprehensive investigation of the relationship between performance on the EGPS Test and writing competence. Owing to the established contribution of spelling to the writing process (as captured by Paper 2: Spelling) the first step was to explore how many dimensions underpin performance on Paper 1: Grammar and Punctuation. The EGPS Test is thought to assess seven grammar and punctuation domains. Factor reduction was carried out through principal component analysis to explore which factors these domains load upon. This reduction of the data was used to inform subsequent analysis, informing exploration of how individual domains or combined factor scores contributed to patterns of association with overall text quality as operationalised through the Wechsler Individual Achievement Test - Second UK Edition (WIAT-II^{UK}) Written Expression Subtest (WES) (Wechsler, 2005).

The next question investigated to what extent the EGPS Test correlates with writing skills measured by an objective norm-referenced test. For text quality, analytical measures from the WES was used and a standardised score was computed. As the EGPS Test is reportedly used as a proxy for writing competency, this question investigated convergent validity of the

EGPS Test and writing quality at differing levels, as well as with a composite analytical measure.

A developing body of research has identified the importance of direct and indirect factors in writing (Dockrell et al., 2019; Kim, Al Otaiba, & Wanzek, 2015; Kim & Schatschneider, 2017). At the time of writing, no studies were available that investigate the specific contributions of grammar and punctuation as assessed at word- and sentence-levels through visually presented decontextualised tasks. Drawing on the existing literature around proximal and distal factors (Dockrell et al., 2019), performance on the EGPS Test was considered alongside proximal and distal factors in relation to predicting writing outcomes. The proximal factors that were explored include spelling and handwriting, while the distal factor was reading comprehension. Although it was planned that measures of oral language and working memory would also be collected, restrictions around COVID-19 meant that the assessment battery needed to be reduced. Due to the proximity (i.e., individual administration) that would be required with participants for these measures, the decision was taken to exclude the oral language and working memory assessments.

The importance of micro-structural factors of writing (productivity, accuracy, complexity) has also been explored in the literature. Addressing limitations with my IFS regarding the subjectivity of analytic writing assessments, CBM-W provided a score of written productivity (total words written, correct word sequences, number of punctuation marks, and number of T-units) and accuracy (proportion of correct word sequences and proportion of correct spellings) (Romig et al., 2017). T-unit analysis and the measurement of syntactic complexity in linguistics was used as an index for the complexity of writing (Hunt, 1964). Clauses per T-Unit explored the syntactic complexity of children's writing through analysis of the CBM-W tasks.

The literature review has also highlighted the importance of variance in genre in the assessment of writing (Berman & Nir-Sagiv, 2007; Koutsoftas & Gray, 2012; Scott & Windsor, 2000). For this reason, an examination of both expository and narrative writing, as assessed by the CBM-W (McMaster & Espin, 2007), was conducted. This enabled comparisons of the relationships between EGPS and writing outcomes across genre.

Finally, the study examined the sensitivity and specificity of the EGPS Test in identifying struggling writers. Very few children in England are exempt from taking statutory

assessments such as the EGPS Test or writing Teacher Assessment. In order to establish the contribution of writing conventions skills as assessed by the EGPS Test across the wider population, struggling writers and children and children with Education Healthcare Plans (EHCP) for Special Educational Needs (SEN) were included in this study. Currently, the EGPS Test is reportedly used as a proxy for writing assessment. However, as seen in the review of the literature, writing is multifaceted. Therefore, it should be established if the EGPS demonstrates appropriate diagnostic accuracy in identifying struggling writers. This area of study may prove more challenging to operationalise since there is no 'gold standard' for the identification of writing difficulty, and any cut-off between 'typical writers' and 'struggling writers' is arbitrary. To explore the place of the EGPS Test as an identifier of 'at risk' writers, a cut-off of <85 (>1 SD) on the WIAT-II^{UK} Written expression subtest was used to identify struggling writers. Standard deviations are typically considered the cut-off point for the identification of underperformance (Hanley, 1997; Snowling, 2000). Once struggling writers were identified, ROC analysis was then be carried out to explore the sensitivity and specificity of the EGPS Test in identifying struggling writers.

Therefore, this thesis explored the validity of the EGPS Test to better understand its relationship with children's writing outcomes through the following four research questions:

- RQ1. How many dimensions underpin performance on the EGPS?
- RQ2a. Does the EGPS Test correlate with writing quality?
  - RQ2b. To what extent does the EGPS Test predict writing quality once proximal and distal factors have been accounted for?
- RQ3a. To what extent does the EGPS Test account for performance in micro-structural factors of writing products?
  - RQ3b. To what extent can the EGPS Test predict micro-structural performance in expository and narrative genres in writing once proximal and distal factors have been accounted for?
- RQ4. Can the EGPS Test identify struggling writers?

7.2 Overall Research Hypotheses

The review of the literature identified several cognitive models and frameworks of writing which could reasonably offer further insight into the relationship between children's

writing and the potential contribution of grammar and punctuation as captured by the EGPS Test. This thesis has not set out to test these established models and frameworks; these models provide an empirical framework to compare the influence of predictors of writing competencies. Rather, it is expected that the impact of writing conventions captured by EGPS upon children's writing would support these models of writing by offering further specificity around the proximal and distal factors involved in text generation processes. Hypotheses related to each of the four research questions are mapped out below.

7.2.1 The EGPS Test Paper 1 will result in two domains: Grammar and Punctuation

Existing research identifies writing conventions as consisting of grammar and spelling (Abbott et al., 2010; Arfé et al., 2016; Berninger, Vaughan, et al., 2002; Daffern et al., 2017; Kim, Al Otaiba, et al., 2013). Additionally, punctuation and vocabulary have also been included within measures of either writing conventions (Berninger et al., 1994; Hayes, 2012; Kim & Schatschneider, 2017; Mackenzie & Hemmings, 2014; Sumner et al., 2014) or of oral grammar (Kim, Al Otaiba, et al., 2013; Olinghouse, 2008).

EGPS Paper 1 reportedly covers 7 distinct domains. Upon exploration of these domains (see Chapter 3.3) most of these domains would typically be regarded as an aspect of grammar in existing literature. Only one domain is regarded as distinct in existing literature - domain G5: Punctuation. As explained in Chapter 3, domain G6: Vocabulary explains knowledge of word-level grammar. Therefore, it was hypothesised that the EGPS Test Paper 1 will be explained by two factors within a principal component analysis: grammar and punctuation.

7.2.2 The EGPS Test Paper 2 will correlate highly with measures of spelling

There is currently no data exploring the convergent validity of the EGPS Test Paper 2 and a standardised measure of spelling. Given the nature of Paper 2, which is administered in a dictation task with guidance sentences placing the word in context, and the 24 spelling content domains which share similarities with standardised measures, it is hypothesised the EGPS Test Paper 2 will offer good convergent validity with a standardised spelling measure,

7.2.3 The EGPS Test Paper 1 will be strongly correlated with writing quality: predicting writing quality at sentence and text levels.

Overall, it was predicted that a higher total score on the combined EGPS Test will predict greater writing quality. At this combined level, the test theoretically taps spelling through Paper 2, which is an established predictor of writing quality from the literature (Abbott

& Berninger, 1993; Daffern et al., 2017). It was then predicted that children who demonstrate greater grammar and punctuation knowledge, as operationalised by the EGPS Test Paper 1, will produce writing of higher quality, given the findings of extant literature which identifies the role of oral grammar (Abbott & Berninger, 1993; Arfé et al., 2016; Berninger & Whitaker, 1993; Kim & Schatschneider, 2017; Olinghouse, 2008), grammatical understanding (Arfé et al., 2016; Mackie et al., 2013; Van der Lely & Christian, 2000), and language conventions (Daffern et al., 2017) in childrens writing quality as measures by sentence- and text-level tasks, as well as composite measures of writing that comprise of performance across three tasks.

It was then hypothesised that the EGPS Test would be able to predict writing to various degrees depending on the level of language. At word-level, it was hypothesised that the proximal transcription skills of spelling and handwriting would be the unique predictors. At sentence- and text-levels, and on the composite writing score, it was predicted that the proximal skill of spelling would be the most significant predictor, followed by the theorised distal factors of grammar and punctuation.

7.2.4 The EGPS Test Paper 1 will predict accuracy and complexity of writing.

Limited studies have established the influence of receptive grammar upon writing (R. Andrews, Torgerson, Beverton, Freeman, et al., 2004; Mackie & Dockrell, 2004). Where children have greater knowledge of *verb forms and tenses* and greater abilities with *combing words, phrases and clauses* (domains 3 and 4 of the EGPS Test Paper 1), it was likely that writing products would demonstrate higher degrees of accuracy through inflection and word choice, as well as potential for increasingly complex sentence combinations through varied syntactic structures. Therefore, it was reasoned that the EGPS Test Paper 1 would demonstrate a strong association with writing accuracy in both narrative and expository writing. It was also reasoned that greater grammar and punctuation knowledge might result in wider syntactic choices, leading to more complex writing products in narrative and expository writing. Regarding writing productivity, the established role of handwriting was expected to be a significant predictor. However, it was reasoned that by age 11, grammar and punctuation skills may become a greater constraint on translation processes than transcription. Children with greater grammar and punctuation skills will be better equipped to produce longer writing products with less constraints as they have a greater repertoire of language structures to draw upon. Advanced writing is associated with the expansion of words and phrases with adjectives

and adverbs (Beard, 2000); therefore, strength in grammatical knowledge and grammatical vocabulary may enhance fluent retrieval of appropriate content, potentially aiding productivity.

Owing to the increasing difficulties of expository writing explored in the literature, no a-priori prediction was offered regarding the relationship between grammar and punctuation skills and genre. It could be reasoned that as with productivity, less constraints from the proximal factors could lead to greater contribution of these theorised distal factors tapped by the EGPS Test to text generation processes; alternatively, it could be reasoned that greater knowledge of writing conventions may negate some of the additional challenges of expository writing. However, the likely impact across genre was unclear.

7.2.5 The EGPS Test will offer high discriminant validity in identifying struggling writers

Struggling writers typically find difficulty at word-, sentence- and text-levels (Graham & Harris, 2005), and transcription skills, which are inclusive of spelling, have been found to limit writing fluency (Berninger, 1999; García et al., 2017; Graham et al., 1997). Therefore, it is reasonable to assume that the combined EGPS Test, which includes word- and sentence-level tasks involving grammar, punctuation, and spelling would offer high levels of sensitivity and specificity in identifying struggling writers. Additionally, a strong association between the EGPS Test and sentence combining was identified by Canniford (2019), mirroring the findings of Dockrell (2019) which identified the diagnostic accuracy of sentence generation in identifying struggling writers. As the EGPS Test is inclusive of multiple skills that underpin writing, namely grammar and spelling, it was likely to offer greater diagnostic accuracy than spelling alone. However, based on the literature review, the role of spelling, handwriting, and reading was also explored in identifying struggling writers.

Chapter 8 : Methodology and Methods

8.1 Introduction

This chapter begins with a brief discussion on research paradigm, strategy, and design. A description of the sample is then offered. The measures used in the study for data collection are explored and the administration and scoring procedures are explained before analytical choices are discussed and ethical considerations are presented.

8.2 Research Paradigm

Research paradigm refers to a given set of common beliefs and agreements which can be shared between scientists to enable understanding of how problems should be understood and addressed (Kuhn, 2012). Paradigm is an intellectual framework which enables researchers to identify the specific worldviews that they hold when engaging in research, which can characterised by ontology, epistemology and methodology (Guba, 1990).

Positivism states that there is a single reality which can be measured through scientific quantitative methods (Gray, 2014). Positivism requires objectivity in research, separating facts and values in order to pursue value-free inquiry (Snape & Spencer, 2003). This thesis is situated within a positivist paradigm, which shapes the ontological, epistemological, and methodological rationale of this research.

Ontology is the study of the nature of existence and reality (Crotty, 1998; L. Hudson & Ozanne, 1988) which explores the possibilities of what can be known about the world (Snape & Spencer, 2003). A variety of ontological positions exist to explain 'whether or not there is a social reality that exists independently from human conceptions and interpretations' (Ormston, Spencer, Barnard, & Snape, 2014: 4). The realist ontological position of this research asserts that the social world exists externally to social actors (Bryman, 2016).

Epistemology explores assumptions that the researcher makes about the nature of knowledge in order to make sense of the world (Crotty, 1998; Richards, 2003). Within social scientific domains, epistemology defines what should regarded as acceptable knowledge within certain disciplines (Bryman, 2016; Gray, 2014). Epistemological assumptions held by researchers affect the methods that they will use in order to uncover knowledge within the social sciences. This thesis is situated within a empiricist epistemological framework in which

it can be presumed that social phenomenon are measurable through the application of methods of the natural sciences (Gray, 2014).

Deductive theory enables a researcher to draw upon theory in order to create hypotheses which are subjected to empirical scrutiny (Bryman, 2016). The deductive nature of this research project, which is rooted in the collection of numerical data, is aligned with the quantitative research strategy.

Finally, this study employed a cross-sectional research design. Cross-sectional design comprises of the collection of data from samples at a single point in time for the purpose of gathering quantifiable data in order to elicit patterns of association for a given sample (Bryman, 2016). One of the challenges of cross-sectional design is potential for lack of internal validity, as cross-sectional data cannot always offer direction of causal influence (Gray, 2014). In the case of this thesis, the design will not always enable us to establish whether explicit grammar knowledge increases writing ability, or vice versa. However, exploration of patterns of association will still offer fresh insights into this much under-researched area of study. Furthermore, exploratory regression analyses will help to explore causal relationships between EGPS and writing.

8.3 Sample

A sample of English state-maintained primary schools and academies were recruited from across England. To ensure a representative sample, inclusion criterion of attainment within 10% of the national average for EGPS and writing was applied using published Ofsted statutory assessment data (Ofsted, 2020). After ethical approval was granted by the University Research Ethics Committee at University College London Institute of Education (Z6364106/2020/02/14), state primary schools and academies across England within teaching school and hub professional development networks who met the inclusion criterion were invited to take part in the study. Information regarding the aims and requirements of the study were sent to school headteachers (See Appendix 1.2.3), resulting in seven state-maintained primaries and academies being recruited, with a total of eight classes offering participation. The mean class size of the sample was 27.25 pupils (*SD = 4.86, range 21-33).* All participating schools in the study were identified as achieving within 10% of the national average for both EGPS 2019 (national = 78%; sample mean = 79.25%) and teacher assessment of writing

(national = 78%; sample mean = 79.88), with the range of prior attainment for the EGPS Test being 70-86 and of writing being 76-85, thus satisfying inclusion criterion.

The seven primary schools were situated in Inner (n = 3) and Outer London (n = 3), and shire counties in the north and the southwest of England (n = 2) with a total sample pool of N = 239 children. Participation in the study was supported by the teachers and headteachers but remained voluntary. Parental consent was sought for all children (see Appendix 1.2.2) to take part in the study through signed permission on a leaflet which children then also posted into a 'consent box' (Canniford, 2019) in order to show their own active informed consent (Robson & McCartan, 2016). Consent was provided by n = 224. Any children that did not complete all the assessment tasks were excluded from the final sample (n = 14), as this would not allow for complete within-samples comparison. An attrition rate of 6.25% resulted in a final sample size of n = 210.

To ensure a representative sample, children with an Education Healthcare Plan (EHCP) for Special Educational Needs (SEN) around language or cognition difficulties were included in the study, although this factor was considered in the analysis of the data.

Power analyses were carried out to indicate the minimum number of participants required. Owing to the multiple methods required to answer the range of research questions, several analyses were run for the requirements of each measurement using a statistical modelling software package (G-Power 3.1). To detect medium sized effects ($\rho H1 = .3$) in bi-variate correlations using the parameters of an alpha-level of $\alpha = .05$ and a power size of $1-\beta = .95$, a priori power analysis indicated that a sample of N = 138 would be adequate (lower critical r = -.17; upper critical r = .17; actual power r = .95). For the detection of medium effects (d = 0.3) within independent samples t-testing (L. Cohen et al., 2013), using the parameters of an alpha-level of $\alpha = .05$ and a power size of $1-\beta = 0.95$, a priori power analysis indicated that a sample of n = 176 is required (non-centrality parameter $\delta = 3.32$; critical t = 1.65; Df = 53; actual power r = .95). For the detection of medium effects ($f^2 = .15$) within linear multiple regression for R^2 increase, using the parameters of an alpha-level of $\alpha = .05$ and a power size of $1-\beta = 0.95$, with a total number of 10 predictors, a priori power analysis indicated that a sample of n = 172 is required (non-centrality parameter $\lambda = 24$; critical F = 2.0; numerator df = 10; denominator df = 161; actual power r = .95). Field (2009) explains that when sampling for regression, a sample should have 15 cases of data for every predictor variable. With a maximum of 10

predictor variables, this would require a sample of N= 150. Therefore, the required minimum sample size of N = 172 to ensure detection of medium effects across the range of measures was met.

8.3.1 Sample Descriptives

Background information regarding age, gender, socioeconomic status as measured by Free School Meals (FSM), English as an additional language (EAL), and special educational needs (SEN) was available for each participant. Within the sample, the mean age in months was 127.48 (SD = 4.94). There was an equal split of gender (male n = 105; female n = 105). Furthermore, 34.76% of the sample were eligible for Free School Meals (FSM), 31.9% were identified as EAL, and 22.4% were identified as on the SEN register. Of note, only one child in the SEN group had an Educational Health Care Plan (EHCP).

Averages for state-funded schools in England (including academies) for 2020 were 14.1% for FSM, 21.2% for EAL, and 13.5% for SEN. This means that the number of disadvantaged children in the sample is higher than the national average. Given the reported correlations between EAL, SEN and FSM, Table 4 presents patterns of co-occurrence of the demographic factors. Of note is the higher proportion of boys in both FSM and EAL, and the higher proportion of girls in SEN.

Table 4 - Sample Demographic Factor Co-occurrence

	N = 210		Male		Female	
	n	%	n	%	n	%
FSM	73	34.8	37	17.61	36	17.14
EAL	67	31.9	27	12.86	40	19.05
SEN	45	21.4	18	8.57	27	12.86

Note. n relates to total participants in each demographic group; n (%) relates to the % of the total sample.

8.4 Measures

As this study sought to explore the relationship between the EGPS Test and writing measures of productivity, accuracy, quality, fluency, and complexity, a number of measures were utilised. In addition, measures of proximal and distal factors of writing were operationalised through a selection of measures explored below.

The principal researcher carried out the scoring for the writing measures of this study and class teachers scored the EGPS Tests. While it is acknowledged that raters should ideally be blind to the aims and purpose of the study, evidence from R. Kent, O'leary, Diament, and Dietz (1974) suggests that knowledge of the hypothesis by raters has little impact on scoring.

8.4.1 English Grammar, Punctuation, and Spelling Test (EGPS) 2018

The EGPS 2018 Test consists of two papers. Paper One comprises of seven grammar and punctuation content domains (STA, 2015) presented through a variety of question types. Children are given 45 minutes for this paper, answering through selected response and constructed response. If children require help reading or understanding a question in the test, administrators can read the question to the child or rephrase a question, provided no subject-specific information is given away. Where a question requires the child to fill the blank from multiple possible answers, the answers can be read to the child, but this must not involve filling the gap with any potential answer. Paper Two is a test of spelling that encompasses 24 spelling content domains and seeks to report the spelling patterns included in the Spelling Appendix within the National Curriculum 2014 (see Appendix 6). Paper 2 is administered by the test administrator with scripted guidance. Twenty contextualised spellings are administered aurally within dictated sentences.

The prescriptive mark schemes provided raters with a selection of acceptable answers for each of the questions, with specific answers not to accept. Scores are totalled for Paper One and Paper Two, and a composite total raw score is calculated. EGPS Tests were marked by class teachers and a moderation of 10% of the EGPS 2019 tests was carried out (n= 30) by the principal researcher.

Combined scores from Papers 1 and 2 are then given a scaled score for statutory reporting and year-on-year comparisons. Scaled scores are allocated a qualitative performance descriptor: working towards the expected standard; working at the expected standard; working at greater depth in the expected standard. However, the scaled score and the performance descriptors are of little practical use in this study. On date of publication, no data on reliability were available for the EGPS 2018 Test. At the time of this study, no reliability data were available for this measure.

8.4.2 WIAT-II ^{UK} Written Expression Subtest (WES)

The WIAT-II UK Written Expression Subtest (WES) (Wechsler, 2005) is an assessment of children's writing proficiency. Designed for children aged 4-17, this assessment of writing competency is administered to the participants on a one-to-one basis following a standardised scripted procedure. The participants within this thesis fit within the age 8-11 category, which dictates the tasks that will be administered. The age-appropriate tasks required to give a standardised score are the word fluency, sentence combination and paragraph tasks (Wechsler, 2005) and the tasks take approximately 30 minutes to administer in total. Test-retest stability coefficient for the WIAT-II UK written expression subtest has been reported as .81, indicating stability over time. An average inter-rater reliability coefficient for the written expression subtest is reported as .85.

The first task within the WES is word fluency. Children are presented with a box where they are asked to write as many things they can think of that are 'round' for 30 seconds. The sentence combination task provides children with a succession of sentences that are required to be combined in various ways as directed by the administrator using a standardised script and picture prompts. The final paragraph task is a timed assessment presented with one of two prompts. Children are given up to 10 minutes to respond on lined paper, with a short planning time before writing. The prompt writing prompt used for the paragraph task was 'On a rainy day, I like to…', in line with the WES administration manual (Wechsler, 2005).

Canniford (2019) explored the reliability of the WES when administered in large groups, instead of on a one-to-one basis. The study found that there was no statistically significant difference between individual and group administration on the WES: $t(96) = -.431$, $p = .67$, $d = 0.08$. In line with Canniford, this study administered the WES to whole classes in order to reduce administration time, enabling the possibility of a larger sample.

Scripts were scored applying the scoring guidance from the WIAT-II^{UK} (Wechsler, 2005). The WES word fluency score is calculated from the sum of single-syllable and multi-syllable responses given. A single point is awarded for a mono-syllabic response and two points are given for poly-syllabic responses. Raw scores of the word fluency task are then required to be scaled using a four-point scale. The scaled score then subsequently contributes to the WES composite score. The WES sentence task uses an array of descriptive statements which allow for the numerous possible combinations of sentences. Written responses are awarded up to two

75

points for the highest level of competence and zero points for the lowest level of competence. The total score of this subtest also contributes directly to the composite writing score. For assessment at the paragraph-level, the paragraph task analytic scoring rubric consists of three domains: mechanics, organisation and vocabulary. A rubric of different descriptive statements of competence are given within each domain. For example, for sentence structure within the domain of organisation, the lowest level of competence is defined as 'the majority of sentences are run-on sentences' and is awarded zero points. Alternatively, the highest level of competence is characterized as 'every sentence is a complete sentence' and is awarded two points.

In all measures, students' names were anonymised, and their scores for each measure were recorded on a scoring grid. Once scores were allocated, no re-adjustments were made to originally assigned scores. Scores were entered into SPSS statistical package for analysis.

While the WES is able to address both analytical and holistic measures of a writing product which account for text quality, it does not offer any information on children's writing productivity or accuracy. Therefore, a variety of additional measures have been selected to fully explore writing products.

8.4.3 Curriculum Based Measures – Writing (CBM-W)

Curriculum Based Writing Measures (CBM-W) offers a standardised procedure for assessing writing proficiency for children aged 7 -12 (Weissenburger & Espin, 2005). Through the administration of a writing probe, children are given 3-7 minutes to respond, typically offering a time-efficient and reliable assessment of children's writing products. CBM-W offers a variety of measures with which writing can be evaluated: total number of words; total number of punctuation marks; total number of T-units; mean words per t-unit; correct word sequences (CWS); incorrect word sequences (IWS); and words spelt correctly.

For the purpose of this research, the administration and analysis of CBM-W followed that of Dockrell et al. (2015), which examined the use of CBM-W in English primary schools.. Using both narrative and expository probes, the writing prompts were presented in written form at the top of a writing response page. Additional sheets of paper were available where required. Owing to known difficulties for younger children in transforming ideas to a specific audience (Scardamalia & Bereiter, 1987), writing tasks for both genres based on person experience were selected to reduce task effects (Bouwer et al., 2015). For the narrative task, the prompt was "One day, I had the best day ever at school"; for the expository, the prompt was "Describe your

favourite thing to do and tell why you like it". For each task, children were invited to think about their answer for 30 seconds and were then given with 5 minutes to write their response. Children were told not to re-write the prompt.

For both genres, measures of productivity, accuracy, and complexity were sought. Productivity was operationalised as total number of words produced, total number of punctuation marks, and total number of T-units. Total number of words (TNW) was determined by a sum of total words produced within the 5-minute administration, exclusive of indicated errors by the child but inclusive of words spelt incorrectly (Dockrell & Connelly, 2015). The prompt, numerals and crossed out words are excluded from the measure. Total number of punctuation marks (TNP) offered a sum of punctuation, irrespective of correct placement. Repetition of punctuation marks at the same point in the text is scored as one punctuation mark (for example, !!!). Correct Word Sequences (CWS) were identified as any combination of two words in sequence deemed acceptable by a native speaker (Weissenburger & Espin, 2005). Within the context of CWS, end punctuation and initial capitalisation for new sentences were also deemed a correct sequence in line with Dockrell et al. (2015). Terminable units (T-units)(Hunt, 1965) refer to syntactic units of meaning which include main clauses and any bound dependent clauses (Arfé et al., 2016). The start of a new t-unit can be identified as following after closing punctuation, or as following a coordinating conjunction (e.g. and/but/or/yet/so). A single point is awarded for each main clause, inclusive of each subordinate clause, and non-clausal structure that is included within each t-unit (Hunt, 1970).

Accuracy was measured through proportion scores. For proportion of correct word sequences (%CWS) and proportion of correct spelling (%CS), scores were calculated in relation to the total score of each textual feature (Dockrell et al., 2015; Kim et al., 2018). For example, percentage of correct word sequences (%CWS) was calculated by dividing CWS by TNW.

Complexity was operationalised through mean length of T-units (MLT-Units). MLT-units were identified by dividing the total number of words by the total number of T-units.

8.4.4 Helen Arkell Spelling Test 2 (HAST-2)

The Helen Arkell Spelling Test Version 2 (HAST-2) is a UK standardised measure of spelling. This array of tests is designed to assess the attainment of spelling for children aged 5 through to adult using group or individual administration. Each test comprises of 30 to 50

words and is administered orally, presenting each word individually, followed by a sentence placing the word in context. Test A was used following group administration procedures which requires children to answer items 11-60, resulting in children writing 50 spellings.

Correct spellings are awarded 1 mark, allowing for letter reversal. When a child misspells 10 non-consecutive words, the scoring is discontinued. Handwriting or capitalisation is not penalised. Raw scores were converted to standard scores (M=100, SD=15). The inter-rater reliability is reported as r = .99 and test-retest reliability coefficient for this test is .91.

8.4.5 The Detailed Assessment of Speed of Handwriting (DASH)

The Detailed Assessment of Speed of Handwriting (DASH) (Barnett, Henderson, Scheib, & Schulz, 2007) is a UK standardised measure of handwriting speed. The Alphabet Writing and Copy Fast tests assessments were administered in this study.

Alphabet writing was administered through a script whereby children are instructed to write the letters of the alphabet (lower-case) in the correct sequence for a period of 60 seconds. This was then scored by allocating one point for each correctly sequenced letter. Letters out of the correct sequence, capital letters, and letter reversal were not allocated a point. Copy Fast was administered through a script whereby children were instructed to write a given sentence (printed at the top of the page) as quickly as possible, whilst ensuring that writing is legible over a period of two minutes. This was then scored by awarding one point for each correctly written word. Repeated, unfinished, misspelt and illegible words were not awarded a point. Raw scores were converted to standard scores (M = 10, SD = 3). The DASH technical manual reports α = .83-.89, demonstrating strong internal consistency.

8.4.6 The New Group Reading Test (NGRT)

The New Group Reading Test (NGRT) is a measure of reading comprehension (GL Assessment, 2010). Participants are required to decode, engage in both sentence completion (20 items) and passage comprehension (32 items) in a booklet. All responses are multiple choice. The prescriptive mark schemes provided raters with the acceptable answers for each of the questions. Children are awarded one point for each correct answer, with a total possible score of 52. Standard age scored were then calculated (M = 100, SD = 15). The technical manual reports test-retest stability coefficient for the NGRT as r = 0.83, which indicates stability across time, and validity as r = 0.81.

8.5 Procedure

All tasks were administered by class teachers as a whole group. For each school in the study, a script was provided for the teachers to administer the tasks based on the given order of their tasks. Due to the range of measures used, a Latin square design within a factorial design structure was used to reduce potential impacts of order effects (Breakwell, Hammond, Fife-Schaw, & Smith, 2006). Owing to COVID-19 restrictions, children were tested in their own classrooms with their own class teachers under test conditions. Sessions were spread over a number of days, not exceeding one school week. All schools completed the tasks within a testing window of three weeks during the Autumn term in 2020. Session lengths varied depending on tasks, however, no session lasted longer than 45 minutes. It was explained to all children that they were able to stop at any point in the tasks if they no longer wished to participate, although children who were hesitant were encouraged on their performance. No children in the study refused to complete any tasks.

The EGPS 2018 Test was administered in two parts, Paper 1: Questions and Paper 2: Spelling. Teachers administered the test in accordance with the Standards & Testing Agency administration guides (STA, 2017a, 2017b). Paper 1 is a 45-minute test and Paper 2 lasts approximately 15 minutes. The remainder of the measures were administered according to the procedures in the testing manuals through a provided script. For group administration of the WES, each item was in accordance with the standardised scripted procedure; however, a digitalised version of the stimulus book was used on the classroom interactive whiteboard as a prompt for elements of the sentence task during group administration.

8.6 Standard Scores, Standard deviations, and Scaled Scores.

Many of the measures used in this study are standardised tests. Where these tests have been used, strict adherence to the prescribed instructions has been maintained as best as possible. Training for schools and a script for administration was provided to ensure adherence to guidance for all measures. Where possible, raw scores were converted to standardised scores by the researcher. In doing so, standardised scores offered comparative information on levels of children's performance in comparison with same-aged peers. Where scores are norm-referenced, this enables comparison against nationally representative samples. These scores can also offer insight where groups deviate from means when reporting any findings and serve as confirmation that typically developing groups are in fact performing at the expected level.

The standard deviation represents the breadth of variability in scores. Most standard scores in this thesis have a mean of 100 and a standard deviation of 15. Therefore, any participants who score between 85 and 115 would fall within the standard deviation of 15 points from the mean. This standard deviation enables a judgment of whether the participant falls within the average range. When considering underperformance in a given area, scores below 1 standard deviation serve as a useful cut-off point (Snowling, 2000). When considering struggling writers or, more broadly, children with learning difficulties, standard deviations are typically regarded as a cut-off point for identification of underperformance (Hanley, 1997; Snowling, 2000). Equally, standardised scores enable comparison of the performance of individuals or groups against an expected level. In the case of the DASH measures, raw scores were converted to a mean of 10 with a standard deviation of 3.

Alternatively, the EGPS Test uses a scaled score, which serve a means of comparison between different versions of the same test. In the case of the EGPS Test, a scaled score of 100 gains a qualitative judgement of 'expected standard'. Use of the scaled score enables year-on-year comparison of test scores, despite variations in content. Scaled scores are criterion referenced and, as such, are not based on a quota of performance. Therefore, it is typical that the mean within a scaled score distribution will not be 100. While the scaled score of the EGPS Test is reported in the descriptive statistics, this is of little practical use and as such raw scores were used in analysis.

8.7 Analysis

8.7.1 Inter-Rater reliability

A moderation of 10% of the EGPS 2018 tests was carried out (n= 30) by the principal researcher. A random sample of 10% of all other assessments (n = 30) were assessed by an additional rater with Qualified Teacher Status to ensure inter-rater stability. All writing scripts were scored by the principal researcher and inter-rater judgements were carried out by one of two experienced English subject leaders. Prior to scoring, training was provided where example scripts were used to reach a minimum agreement of 80%, before scoring of the full random sample was carried out. A two-way random-effect model based on absolute agreement assessed the inter-rater reliability. A high degree of reliability was observed for all measures, with intra-class correlation estimates demonstrating at least strong (> .71) or very strong (> .91) inter-rater reliability (LeBreton & Senter, 2008).

8.7.2 Statistical approach

8.7.2.1 Overall Performance

Data collected within this study was analysed using the SPSS statistical package. Initial exploration of task performance sought to understand general patterns of distribution within the sample. Normal distribution was explored through a Shapiro-Wilk test of normality (Shapiro & Wilk, 1965), whereby a significance level greater than .05 indicates that the sample is not different from normal distribution. This was confirmed by explorations of the sample distributions which included visual inspections of histograms, normal Q-Q plots and box plots to understand the skewness and kurtosis of the data (Cramer, 1998; Doane & Seward, 2011). Levene's test for equality of variance were run to establish homogeneity of variance within the sample. A significance level greater than .05 indicates that the variances are approximately equal and thus the data meets the assumptions for parametric tests. Where this value was found to be significant ($< .05$), Welch t-testing was used as this method does not assume equal population variances. Bonferroni corrections were applied where multiple correlations were carried out to counteract inflated Type I errors.

Descriptive statistics, including means, standard deviations, and minimum and maximum scores are provided in the results tables. Group differences were then explored through inferential statistics. Assumptions regarding outliers and distribution for groups were explored through boxplots and further Shapiro-Wilks testing. Where outliers were present, t-testing was carried out with and without outliers. Where evaluation determined no appreciable effect (Laerd, 2018), outliers were retained in the analysis unless otherwise indicated in reporting of the data. Where further assumptions of parametric t-testing were not met, the non-parametric Mann-Whitney U test was used. After adjusting for multiple tests with Bonferroni's adjustment, t-values which reached significance were then tested for effect size using Cohen's D for parametric tests or Cohens r for non-parametric tests.

8.7.2.2 Research Question 1

Initial exploration of the EGPS Paper 1 and its constituent test domains made use of exploratory factor analysis with principal component analysis (PCA) factor extraction method to reduce data into a smaller number of meaningful factors. Factor analysis is a tool which can be employed within statistical methods to enable overlapping variables to be reduced into lesser set of factors based around specific characteristics of a given measure (Green & Salkind, 2012).

81

Convergent validity of the EGPS Paper 2 and the HAST-2 was assessed through bi-variate correlation.

8.7.2.3 Research Question 2

Inferential statistics were reported using tests of bi-variate correlation. Where the data demonstrated a linear relationship; no significant outliers; and bivariate normality (Laerd, 2015b), then zero-order Pearson's correlation were presented. The significance level of testing was adjusted for multiple correlations using Bonferroni's adjustment in order to counteract inflated Type I errors.

Where significant relationships were found between two or more variables of importance to the research questions, stepwise multiple linear regression analysis were then used to develop a model for predicting writing performance based on performance scores on the EGPS, as well as established proximal and distal factors. Distinct regression analyses were examined for different levels of writing to establish distinctions between different operationalisations of writing quality.

Finally, path analysis explored the direct and indirect contributions of the various measures, including EGPS, to overall writing quality. AMOS statistical package was used to explore path models. Measures of good fit were assessed by χ^2 ($p > .05$), goodness of fit index (GFI) ($>.9$), comparative fit index (CFI) ($>.9$) and root mean square residual (close to zero as possible).

8.7.2.4 Research Question 3

Initial exploration of the micro-structural measures of writing made use of exploratory factor analysis with principal component analysis factor extraction method to reduce data into a smaller number of meaningful factors.

Inferential statistics were reported using zero-order Pearson's correlation using Bonferroni's adjustment in order to counteract inflated Type I errors. Where significant relationships were found between two or more variables of importance to the research questions, stepwise multiple linear regression analysis was then used to develop a model for predicting writing performance based on performance scores on the EGPS, as well as established proximal and distal factors. Separate regression analyses were computed for the factors identified in principal component analysis to make further distinctions with regards to different microstructural factors. This process was repeated to further explore the individual factors that contribute to each narrative and expository genre.

7.7.2.5 Research Question 4

For research question 4, z-scores were reported to explore differences in EGPS and the range of writing measures. ROC analyses were then presented to explore the specificity and sensitivity of the EGPS Test in relation to struggling writers. Good sensitivity and specificity indicate a reliable discriminator of typical and struggling writers.

8.8 Ethical Considerations

This project has adhered to the British Psychological Society (2009) Code of Ethics and Conduct, and Code of Human Research Ethics (2014). As this study has taken place in schools, evidence of Disclosure and Barring Service Enhanced Certificates for all involved in the research project were confirmed. This research project was also achieved in line with institutional policies regarding safeguarding, child-protection, health and safety, and COVID-19 regulations.

Prior to undertaking the research, ethical approval was sought from the Institute of Education, University College London. This study was granted ethical approval by the primary and secondary supervisors and registered with UCL with the following data protection number: Z6364106/2020/02/14. Ethical considerations are vital due to the potential for harm, stress and anxiety that is possible in any research project involving participants (Robson & McCartan, 2016) particularly children. Final approval was granted in May 2018, prior to data collection. Re-submission of approval was granted in October 2020 due to updates in-line with COVID-19. For further information regarding ethical considerations, see Appendix 1.

8.9 Summary

This chapter has presented the research paradigm of the study, clarifying the key characteristics of the sample. Children were drawn from around England from schools that performed close to national average in Key Stage 2 statutory assessments. The measures presented enabled empirical exploration of the EGPS Test and its relationship with writing, and scoring procedures were set out. Finally, ethical considerations were presented. This thesis now presents the results in Chapter 9.

CHAPTER 9: Results

This chapter begins by exploring the overall performance of the sample. Descriptive statistics for the whole sample are offered, before an exploration of group differences is considered. This is subsequently followed by analysis of the data which responds to each of the research questions.

9.1 Overall Task Performance

The standard scores of the composite WIAT-II[UK] WES, which is norm-referenced to a UK population, were approximately normally distributed, with a skewness of -.86 (SE = .17) and kurtosis of .52 (SE = .33). Each of the component raw scores of the WIAT-II[UK] Written Expression Sub-test also demonstrated approximately normal distribution, excluding the word fluency raw score which was mildly leptokurtic (1.03, SE =.33), meaning that the sample was more concentrated around the mean than the normal distribution. As this is not the case for the scaled composite score, this may reflect the word fluency measure rather than the sample.

The EGPS 2018 Raw scores were approximately normally distributed. The measure demonstrated a skewness of -.16 (SE = .17) and kurtosis of .87 (SE = .33).

The CBM measures of total words produced and number of t-units for both narrative and expository genres were approximately normally distributed, with a skewness and kurtosis in the normal range. However, number of punctuation marks and mean words per T-unit were both positively skewed and leptokurtic, meaning that the mode and median were below the mean as demonstrated by a longer tail on the right side.

When exploring Proximal and Distal measures, DASH Copy Fast task scores were leptokurtic (1.51, SE = .33) and further inspection of papers indicate possible administration errors. However, DASH Alphabet were approximately normally distributed. HAST-2 standard scores were also approximately normally distributed. Finally, the standard scores for the NFER measure of reading comprehension were also approximately normally distributed. These preliminary analyses support the use of each of the selected measures, excluding DASH Copy Fast, for subsequent analyses.

Table 5 – Descriptive Statistics for EGPS, Writing, Proximal and Distal Measures (N=210)

	Mean	(SD)	Min	Max	Skew (SE)	Kurtosis (SE)
EGPS 2018 Test						
EGPS Scaled Score	100.98	(7.72)	84.00	120.00	.10 (.17)	-.46 (.33)
Paper 1: Grammar and Punctuation Raw Score	27.67	(10.23)	7.00	50.00	-.06 (.17)	-.98 (.33)
Paper 2: Spelling Raw Score	11.96	(4.92)	.00	20.00	-.60 (.17)	-.29 (.33)
EGPS Combined Raw Total Score	39.63	(13.82)	9.00	69.00	-.16 (.17)	-.87 (.33)
Writing Measures						
WIAT-II[UK] Composite Standard Score	103.63	(13.07)	67.00	135.00	-.15 (.17)	-.17 (.33)
WIAT-II[UK] Composite Raw Score	22.44	(5.312)	8.00	35.00	-.10 (.17)	-.19 (.33)
WIAT-II[UK] Word Fluency Raw Score	8.29	(3.57)	1.00	22.00	.79 (.17)	1.03 (.33)
WIAT-II[UK] Word Fluency Scaled	1.94	(1.09)	1.00	4.00	.75 (.17)	.82 (.33)
WIAT-II[UK] Sentence Combination Raw Score	6.10	(2.65)	.00	12.00	-.10 (.17)	-.45 (.33)
WIAT-II[UK] Paragraph Task	14.40	(3.55)	5.00	22.00	-.19 (.17)	-.47 (.33)
CBM-N — Total words produced	60.05	(25.41)	12.00	139.00	.67 (.17)	.29 (.33)
No. punctuation marks	4.77	(3.79)	.00	25.00	1.64 (.17)	4.18 (.33)
No. T-units	6.75	(3.34)	1.00	21.00	.99 (.17)	1.48 (.33)
Mean words per T-unit	9.58	(3.03)	4.10	25.00	1.58 (.17)	4.65 (.33)
Proportion CWS[#]	.89	(.11)	.42	1.00	-1.54 (.17)	2.50 (.33)
Proportion Correct Spelling[#]	.96	(.05)	.50	1.00	-3.84 (.17)	24.85 (.33)
CBM-E — Total words produced	54.49	(22.09)	9.00	128.00	.54 (.17)	.43 (.33)
No. punctuation marks	3.95	(3.03)	.00	19.00	1.86 (.17)	4.85 (.33)
No. T-units	5.32	(2.57)	1.00	13.00	.79 (.17)	.28 (.33)
Mean words per T-unit	11.07	(3.57)	5.00	27.50	1.38 (.17)	2.92 (.33)
Proportion CWS[#]	.90	(.12)	.26	1.00	-2.09 (.17)	13.64 (.33)
Proportion Correct Spelling[#]	.96	(.05)	.61	1.00	1.31 (.17)	2.92 (.33)
Proximal Factors						
DASH Handwriting – Copy Task Standard Score	9.19	(3.14)	3	17	.45 (.17)	-.07 (.33)
DASH Handwriting – Copy Task Raw Score	14.04	(4.736)	4	33	.82 (.17)	1.51 (.33)
DASH Handwriting – Alphabet Task Standard Score	9.13	(2.73)	3	17	.41 (.17)	.02 (.33)
DASH Handwriting – Alphabet Task Raw Score	43.65	(17.196)	12	104	.59 (.17)	.15 (.33)
HAST-II Spelling Standard Score	82.99	(11.91)	59	116	-.44 (.17)	-.13 (.33)
HAST-II Spelling Raw Score	34.42	(6.15)	10.00	49.00	-1.23 (.17)	.17 (.33)
Distal Factors						
NFER Reading Standard Score	98.66	(12.53)	60	141	.20 (.17)	.39 (.33)
NFER Reading Raw Score	27.90	(9.62)	6.00	51.00	.04 (.17)	-.75 (.33)

Note:
Standard Scores M = 100, SD = 15.
DASH Standard Score M = 10, SD = 3.
[#] Accuracy measures – proportion scores reported

9.1.1 EGPS National Comparison

As can be seen in Table 5, the mean scaled score on the EGPS Test 2018 was 101. This is lower than the average scaled score for all schools in England in 2018, which was 106 (DfE, 2019: 5). This difference in attainment may indicate that the sample in this study are working below the national average attainment in EGPS. However, this data must also be considered in light of the timing of the data collection. Data collection for this study took place in the Autumn term. This was eight months before children in the national data comparison would have sat this statutory assessment which takes place in the Summer term. Therefore, this could reasonably account for a reduction in attainment compared with national averages. Furthermore, the sample mean on standard score of the WIAT-II^{UK} indicates that the sample are broadly academically average, with a mean standard score of 104. However, this difference in EGPS attainment should be considered in the analysis of the data.

9.1.2 Group Differences Analytic Strategy

For each demographic group, independent t-tests were carried out. Where measures violated assumptions for parametric testing, each measure was individually assessed with the inclusion and exclusion of outliers to see if this resulted in any applicable effect (Laerd, 2015a). Where differences were not evident, outliers were retained in the analysis and t-tests were carried out. The exception to this was the CBM proportion scores, where non-parametric Mann-Whitney U-test scores are reported due to violations of assumptions of normal distributions. Bonferroni's correction was applied (p = .002) to all measures. To reduce the data for analysis, standard scores have been omitted. Raw scores are used for the testing of difference within subsequent analysis.

9.1.3 Gender Differences

Table 6 splits the sample into two groups: boys and girls. As can be seen in Table 6, there was a statistically significant difference between WIAT-II^{UK} Sentence Combination scores between males and females, although this failed to reach significance with Bonferroni's corrections. There were no other significant gender differences in the remaining EGPS, WIAT-II^{UK}, proximal or distal measures.

Table 6 - Gender differences in EGPS, Writing, Proximal and Distal Measures (N=210)

	Male (n=105) Mean	(SD)	Female (n=105) Mean	(SD)	Test Stat. T / U	Effect Size Cohen's D / *r*
EGPS 2018 Test						
Paper 1: Grammar and Punctuation Raw Score	26.86	(10.20)	28.49	(10.24)	-1.15	
Paper 2: Spelling Raw Score	11.91	(5.23)	12.00	(4.62)	-.13	
EGPS Combined Raw Total Score	38.77	(13.91)	40.49	(13.73)	-.90	
Writing Measures						
WIAT-II^{UK} Composite Raw Score	21.76	(4.85)	23.12	(5.68)	-1.87	
WIAT-II^{UK} Word Fluency Raw Score	8.57	(3.82)	8.01	(3.29)	1.14	
WIAT-II^{UK} Sentence Combination Raw Score	5.54	(2.46)	6.65	(2.73)	-3.08*	.43
WIAT-II^{UK} Paragraph Task	14.16	(3.40)	14.65	(3.70)	-.99	
CBM-N Total words produced	56.62	(24.99)	63.49	(25.48)	1.97	
No. punctuation marks	4.45	(3.44)	5.09	(4.10)	1.22	
No. T-units	6.40	(3.12)	7.10	(3.52)	1.51	
Mean words per T-unit	9.39	(2.58)	9.77	(3.42)	.90	
Proportion CWS[#]	.89	(.12)	.90	(.11)	5230	
Proportion Correct Spelling[#]	.96	(.05)	.97	(.06)	5120	
CBM-E Total words produced	53.35	(23.33)	55.62	(20.83)	.74	
No. punctuation marks	3.67	(2.86)	4.24	(3.18)	1.37	
No. T-units	5.25	(2.74)	5.39	(2.41)	.40	
Mean words per T-unit	10.95	(3.24)	11.20	(3.88)	.49	
Proportion CWS[#]	.89	(.12)	.91	(.11)	4934	
Proportion Correct Spelling[#]	.96	(.06)	.96	(.05)	5156	
Proximal Factors						
DASH Handwriting – Copy Task Raw Score	14.08	(5.15)	14.01	(4.30)	.10	
DASH Handwriting – Alphabet Task Raw Score	42.16	(17.10)	45.14	(17.24)	-1.26	
HAST-II Spelling Raw Score	34.34	(6.32)	34.50	(6.01)	-.19	
Distal Factors						
NFER Reading Raw Score	27.65	(8.86)	28.14	(10.35)	-.37	

Note:
* *p* < 0.01 ** Bonferroni's Adjustment *p* < .002
Standard Scores M = 100, SD = 15.
DASH Standard Score M = 10, SD = 3.
[#] Accuracy measures – proportion scores reported
Test Statistics: T = T-testing; U = U-score
Effect Sizes: Cohen's D for T-testing; *r* for U-Score

87

9.1.4 Differences between pupils with recorded SEN and their Peers

Table 7 splits the sample into two groups: those with SEN and those without SEN. There was a statistically significant difference between all EGPS and WIAT-II^{UK}, measures (see Table 7) whereby children not identified as having an SEN consistently performed better than those identified with SEN, except on the measure of WIAT-II^{UK} Word fluency, where group performance did not reach statistical significance. Statistically significant differences were also found with measures of number of punctuation marks, proportion of correct word sequences and proportion of correct spellings, as well as each of the proximal and distal measures, whereby children with SEN performed more poorly than those without SEN status. Therefore, SEN status was considered as a factor in future regression models.

Table 7- Performance of pupils with recorded SEN and their peers in EGPS, Writing, Proximal and Distal Tasks (N=210)

		SEN Status				Statistics	
		SEN (n=45)		Non-SEN (n=165)		Test Stat.	Effect Size
		Mean	(SD)	Mean	(SD)	T / U	Cohen's D / r
EGPS 2018 Test							
Paper 1: Grammar and Punctuation Raw Score		20.60	(8.45)	29.60	(9.84)	-5.60**	.98
Paper 2: Spelling Raw Score		8.80	(5.73)	12.82	(4.31)	-5.14**	.79
EGPS Combined Raw Total Score		29.40	(12.16)	42.42	(12.93)	-4.38**	1.04
Writing Measures							
WIAT-II[UK] Composite Raw Score		19.00	(5.11)	23.38	(4.98)	-5.20**	1.15
WIAT-II[UK] Word Fluency Raw Score		7.91	(3.60)	8.39	(3.56)	-.80	
WIAT-II[UK] Sentence Combination Raw Score		4.78	(2.43)	6.45	(2.60)	-3.89**	.66
WIAT-II[UK] Paragraph Task		12.38	(3.30)	14.96	(3.42)	-4.52**	.77
CBM-N	Total words produced	53.18	(27.49)	61.93	(24.57)	-2.06	
	No. punctuation marks	3.44	(2.90)	5.13	(3.93)	-3.17*	.48
	No. T-units	6.07	(3.13)	6.93	(3.38)	-1.55	
	Mean words per T-unit	9.45	(3.82)	9.62	(2.79)	-.28	
	Proportion CWS[#]	.82	(.13)	.92	(.10)	1965.5**	.33
	Proportion Correct Spelling[#]	.93	(.06)	.97	(.05)	2230**	.29
CBM-E	Total words produced	47.11	(20.75	56.50	(22.08)	-2.56	
	No. punctuation marks	2.78	(2.00)	4.27	(3.19)	-3.86**	.56
	No. T-units	5.07	(2.62)	5.39	(2.56)	-.74	
	Mean words per T-unit	10.11	(3.07)	11.34	(3.66)	-2.06	
	Proportion CWS[#]	.83	(.15)	.92	(.10)	2317**	.29
	Proportion Correct Spelling[#]	.93	(.07)	.97	(.04)	2359**	.27
Proximal Factors							
DASH Handwriting – Copy Task Raw Score		11.96	(3.87)	14.64	(4.77)	-3.47**	.62
DASH Handwriting – Alphabet Task Raw Score		36.76	(17.71)	45.53	(16.62)	-3.10**	.51
HAST-II Spelling Raw Score		29.80	(7.77)	35.68	(4.96)	-4.82**	.90
Distal Factors							
NFER Reading Raw Score		21.58	(8.35)	29.62	(9.23)	-5.28**	.91

Note:
* $p < 0.01$ ** Bonferroni's Adjustment $p < .002$
Standard Scores M = 100, SD = 15.
DASH Standard Score M = 10, SD = 3.
[#] Accuracy measures – proportion scores reported.
Test Statistics: T = T-testing; U = U-score.
Effect Sizes: Cohen's D for T-testing; r for U-Score.

9.1.5 Differences between children who had English as an Additional Language and Monolingual Peers

Table 8 splits the sample into two groups: those with EAL and those without EAL. There was a statistically significant difference between the performance of children with EAL and monolingual English-speaking children's performance on the EGPS Raw Combined Total and EGPS Paper 1, where medium effect sizes were observed (L. Cohen & Holliday, 1996), as shown in Table 8. For both measures, the mean score for children with EAL exceeded that of native English-speaking children. A statistically significant difference was found for narrative CBM measures of number of punctuation marks, proportion of correct word sequences, and proportion of correct spelling whereby children with EAL outperformed their native English peers. However, these differences were not apparent in the expository CBM measures where the two groups were comparable. A statistically significant difference was found for measures of reading, where again children with EAL out preformed their monolingual peers, although the effect size was small. There were no statistically significant differences between the two groups for any of the remaining measures.

Table 8 - Differences between children who had English as an Additional Language and Monolingual Peers in EGPS, Writing, Proximal and Distal Tasks (N=210)

| | EAL Status | | | | Statistics | |
| | EAL (n=67) | | Non-EAL (n=143) | | Test Stat. | Effect Size |
	Mean	(SD)	Mean	(SD)	T / U	Cohen's D / r
EGPS 2018 Test						
Paper 1: Grammar and Punctuation Raw Score	31.61	(11.05)	25.83	(9.30)	3.95**	.57
Paper 2: Spelling Raw Score	12.88	(4.99)	11.52	(4.85)	1.87	
EGPS Combined Raw Total Score	44.49	(14.80)	37.35	(12.76)	3.59**	.52
Writing Measures						
WIAT-II^{UK} Composite Raw Score	22.61	(5.95)	22.36	(5.00)	.32	
WIAT-II^{UK} Word Fluency Raw Score	8.54	(3.98)	8.17	(3.36)	.69	
WIAT-II^{UK} Sentence Combination Raw Score	6.09	(2.96)	6.10	(2.50)	-.02	
WIAT-II^{UK} Paragraph Task	14.46	(3.80)	14.38	(3.44)	.16	
CBM-N Total words produced	63.03	(26.99)	58.66	(24.61)	1.16	
No. punctuation marks	5.82	(4.37)	4.27	(3.39)	2.56*	.40
No. T-units	7.25	(3.67)	6.51	(3.16)	1.51	
Mean words per T-unit	9.33	(2.59)	9.70	(3.22)	-.81	
Proportion CWS$^{#}$	.92	(.10)	.88	(.12)	6211.5**	.24
Proportion Correct Spelling$^{#}$	.98	(.04)	.96	(.06)	6119**	.23
CBM-E Total words produced	57.91	(23.43)	52.88	(21.33)	1.54	
No. punctuation marks	4.64	(3.86)	3.63	(2.51)	1.96	
No. T-units	5.69	(2.76)	5.15	(2.47)	1.42	
Mean words per T-unit	11.10	(4.02)	11.06	(3.35)	.08	
Proportion CWS$^{#}$	.91	(.09)	.89	(.13)	5221.5	
Proportion Correct Spelling$^{#}$	.97	(.03)	.96	(.06)	5651.5	
Proximal Factors						
DASH Handwriting – Copy Task Raw Score	13.81	(4.53)	14.18	(4.81)	-.54	
DASH Handwriting – Alphabet Task Raw Score	43.76	(18.21)	43.60	(16.76)	.06	
HAST-II Spelling Raw Score	35.40	(6.13)	33.97	(6.13)	1.58	
Distal Factors						
NFER Reading Raw Score	29.90	(10.23)	26.96	(9.20)	2.08*	.30

Note:
* $p < 0.01$ ** Bonferroni's Adjustment $p < .002$
Standard Scores M = 100, SD = 15.
DASH Standard Score M = 10, SD = 3.
$^{#}$ Accuracy measures – proportion scores reported.
Test Statistics: T = T-testing; U = U-score.
Effect Sizes: Cohen's D for T-testing; r for U-Score.

9.1.6 Differences between children with Pupil Premium status and their peers

As can be seen in Table 9, there was a statistically significant difference between children eligible for Pupil Premium (PP) and children not eligible for PP on the EGPS Paper 1, whereby the mean score for children not eligible for pupil-premium was statistically significantly higher, although the effect size was small (L. Cohen & Holliday, 1996). A statistically significant difference between both narrative CBM accuracy measures demonstrated that the performance of children eligible for PP was lower than their peers. All other remaining measures demonstrated comparable scores between children eligible for PP and children not eligible for PP.

Table 9 - Differences between children with Pupil Premium Status and their Peers in EGPS, Writing, Proximal and Distal Tasks (N=210)

	PP Status				Statistics	
	PP (n=73)		Non-PP (n=137)		Test Stat.	Effect Size
	Mean	(SD)	Mean	(SD)	T / U	Cohen's D / r
EGPS 2018 Test						
Paper 1: Grammar and Punctuation Raw Score	24.90	(10.31)	29.15	(9.91)	-2.91*	.42
Paper 2: Spelling Raw Score	11.52	(4.73)	12.19	(5.03)	-.94	
EGPS Combined Raw Total	36.42	(13.57)	41.34	(13.69)	-2.48	
Writing Measures						
WIAT-II^{UK} Composite Raw Score	21.84	(5.78)	22.77	(5.04)	-1.21	
WIAT-II^{UK} Word Fluency Raw Score	8.64	(4.03)	8.10	(3.30)	1.05	
WIAT-II^{UK} Sentence Combination Raw Score	5.92	(2.61)	6.19	(2.67)	-.71	
WIAT-II^{UK} Paragraph Task	13.92	(3.93)	14.66	(3.32)	-1.38	
CBM-N — Total words produced	60.73	24.57	59.69	25.93	.28	
No. punctuation marks	4.62	3.98	4.85	3.70	-.42	
No. T-units	7.03	3.61	6.60	3.19	.89	
Mean words per T-unit	9.31	2.51	9.73	3.27	.96	
Proportion CWS#	.87	.12	.91	.11	3901**	.18
Proportion Correct Spelling#	.96	.05	.97	.06	3833**	.20
CBM-E — Total words produced	56.48	22.92	53.42	21.64	.95	
No. punctuation marks	3.73	3.11	4.07	2.99	-.79	
No. T-units	5.58	2.27	5.18	2.49	1.05	
Mean words per T-unit	11.15	3.93	11.04	3.37	.21	
Proportion CWS#	.88	.12	.91	.11	4296.5	
Proportion Correct Spelling#	.96	.05	.96	.06	4445.5	
Proximal Factors						
DASH Handwriting – Copy Task Raw Score	13.93	(5.25)	14.13	(4.42)	-.29	
DASH Handwriting – Alphabet Task Raw Score	43.77	(19.26)	43.59	(16.06)	.07	
HAST-II Spelling Raw Score	33.47	(6.30)	34.93	(6.04)	-1.65	
Distal Factors						
NFER Reading Raw Score	26.25	(9.88)	28.77	(9.39)	-1.82	

Note:
* $p < 0.01$ ** Bonferroni's Adjustment $p < .002$
Standard Scores M = 100, SD = 15.
DASH Standard Score M = 10, SD = 3.
Accuracy measures – proportion scores reported.
Test Statistics: T = T-testing; U = U-score.
Effect Sizes: Cohen's D for T-testing; r for U-Score.

9.1.7 Summary of overall task performance

In summary, this exploration of the data has identified some differences which were accounted for within subsequent analyses. Limited differences were identified when exploring gender. There was a clear difference between children identified as SEN status and their peers

across the range of measures, and this has been controlled for in all analyses accordingly. The effect of EAL status has identified differences with moderate effect sizes within measures of EGPS and some difference with small effect sizes in micro-structural writing and reading. For Pupil Premium status, statistically significant differences were identified for EGPS and micro-structural measures of writing, albeit with small effect sizes. As this study seeks to understand the contribution of the EGPS Test to writing outcomes, it is important to understand any factors that may influence EGPS scores; therefore, EAL and PP status were retained in the analysis.

9.2 RQ1a – How many dimensions underpin performance on the EGPS?

To examine the extent to which the EGPS Paper 1 reflects the content domains of grammar and punctuation, an exploratory factor analysis (EFA) with a principal component analysis (PCA) factor extraction method was run on the 7 domains (6 grammar and 1 punctuation,) examined in Paper 1.

Table 10 - EGPS Domains Factor Analysis: One Factor Solution

| | Extraction Sums of Squared Loadings | | | |
Component	Total	% of Variance	% Cumulative	Total
1	4.73	67.62	67.62	4.73
2	.55	7.83	75.45	
3	.47	6.66	82.10	
4	.41	5.87	87.97	
5	.35	5.06	93.03	
6	.27	3.83	96.86	
7	.22	3.14	100.00	

Extraction Method: Principal Component Analysis.

The EFA identified one factor with eigenvalues >1, which accounted for 67% of the variance in the EGPS Test Paper 1 (see Table 10). Factor loadings are reported in Table 11. Visual inspection of the scree plot (see Figure 7) confirmed that only one factor should be retained. All grammar and punctuation domains loaded on one factor, which was interpreted as *EGPS Writing Convention Skills (EGPS-WC).*

Table 11 - EGPS Grammar Domain Factor Loadings

	EGPS Writing Convention Skills
G1 Grammatical terms / word classes	.87
G2 Functions of sentences	.74
G3 Combining words, phrases and clauses	.80
G4 Verb forms, tenses and consistency	.85
G5 Punctuation	.89
G6 Vocabulary	.81
G7 Standard English and formality	.78

Extraction Method: Principal Component Analysis. 1 component extracted.
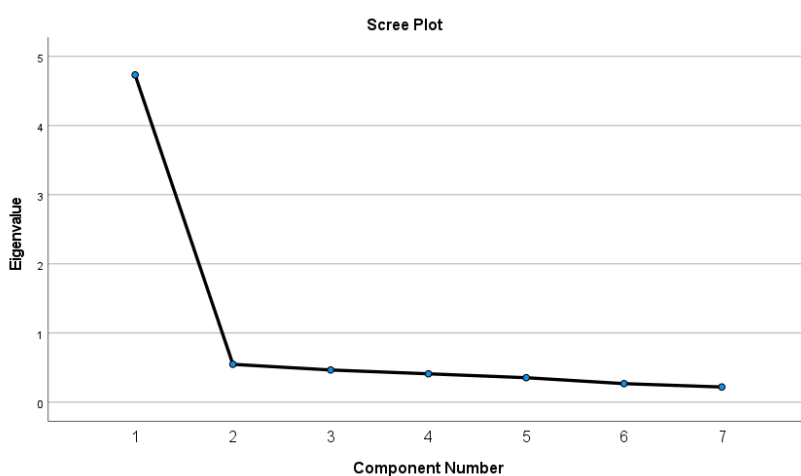
Figure 7 - Scree Plot for EGPS Grammar Domains

Considering that each of the constituent domains in Paper 1 load on Writing Convention Skills, no further exploration of each of the individual domains from Paper 1 was warranted. Factor scores were computed to represent EGPS Writing Convention Skills (EGPS-WC), which are used in place of EGPS Paper 1 scores in the subsequent analyses, unless otherwise indicated.

Paper 2 consists of a single total score of children's spelling attainment which offers a measure of the dimension of spelling. Convergent validity of the dimension of spelling, as assessed by the EPGS test, and the standardised HAST-2 spelling assessment was established through Pearson's correlation. EGPS Paper 2 correlates strongly with the HAST-2 measure of spelling: $r(210) = .84, p < .001$.

9.3 RQ2 - Does the EGPS Test Correlate with Writing Quality?

The next research question sought to explore the relationship between the EGPS Test and writing at multiple levels. Firstly, convergent validity between the EGPS Test (combined raw score) and writing quality as operationalised by the composite WIAT-II^{UK} was examined. Next, owing to the known contribution of spelling to children's writing quality established in the literature, the EGPS Writing Convention Factor score (EGPS-WC) is used to better understand the unique contribution of writing convention skills to writing quality, independent of spelling. Table 12 presents correlations between the EGPS combined raw, EGPS-WC, EGPS Paper 2 scores, and the WIAT-II^{UK} measures of writing quality.

Table 12 - Bivariate Correlations between EGPS and WIAT-II^{UK}

	1	2	3	4	5	6	7
1. EGPS Test Combined Raw Total	-	.90**	.81**	.27**	.43**	.65**	.70**
2. EGPS Writing Convention Skills		-	.58**	.21*	.37**	.58**	.62**
3. EGPS Paper 2 – Spelling Raw Score			-	.19*	.35**	.59**	.60**
4. WIAT-II^{UK} Word Fluency Raw Score				-	.05	.10	.28**
5. WIAT-II^{UK} Sentence Raw Score					-	.36**	.74**
6. WIAT-II^{UK} Paragraph Raw Score						-	.86**
7. WIAT-II^{UK} WES Composite Raw score							-
N = 210, Bonferroni's adjustment: *p = .006, **p < .001							

The EGPS Test 2018 combined raw score showed strong correlations with the WIAT-II^{UK} Written Expression Subtest (WES) composite measure of writing quality and paragraph scores. More moderate correlations were found with sentence writing and word fluency.

Exploring now the independent contributions of grammar and punctuation, the EGPS-WC score showed strong correlations with the WIAT-II^{UK} WES composite score and the WIAT-II^{UK} Paragraph sub-test. A more moderate correlation was found between the EGPS-WC score and the WIAT-II^{UK} Sentence sub-test. A small statistically significant correlation was found between the EGPS-WC and WIAT-II^{UK} Word fluency sub-test.

9.4 RQ2a – To what extent does the EGPS Test predict writing quality once proximal and distal factors have been accounted for?

The next research question explored how much variance in performance on the EGPS Test Paper 1 (EGPS-WC) accounts for children's writing quality at word-, sentence- and text-levels, followed by composite quality as operationalised by the WES. Zero order correlations between the EGPS-WC score, writing measures at word-, sentence-, paragraph-levels, and

composite task, and the proximal and distal measures were first explored. Multiple regressions were used to examine the independent contribution of EGPS-WC. Finally, path analysis was used to capture relationships between proximal and distal contributions to writing quality.

9.4.1 Associations between EGPS, Writing Quality Measures, and Proximal and Distal Factors

Table 13 shows the zero order correlations between the measures of EGPS-WC, writing measures, and the proximal and distal measures captured. Bonferroni's adjustment ($p \leq .006$) was used to control for multiple correlations and raw scores were used for each test measure. For proximal factors in writing, measures of handwriting and spelling were explored and for distal factors, reading comprehension was explored.

Table 13 - Zero order correlations between EGPS-WC, writing measures, proximal and distal factors

	1	2	3	4	5	6	7	8
1. EGPS Writing Convention Skills	-	.21**	.37**	.58**	.62**	.15	.55**	.73**
2. WIAT-II^{UK} Word Fluency Raw score		-	.05	.10	.28**	.27**	.23**	.11
3. WIAT-II^{UK} Sentence Raw Score			-	.36**	.74**	.22*	.35**	.28**
4. WIAT-II^{UK} Paragraph Raw Score				-	.86**	.19**	.58**	.51**
5. WIAT-II^{UK} Composite Raw score					-	.28**	.60**	.50**
6. DASH - Alphabet Raw Score						-	.26**	.08
7. HAST-2 Spelling Raw score							-	.57**
8. Reading Raw Score								-
N = 210, Bonferroni's adjustment: *p = .006, **p < .001								

The EGPS-WC score was found to correlate strongly with the distal measure of reading, and the proximal measure of spelling. While there was an association between the EGPS-WC score and handwriting speed, this failed to meet significance when Bonferroni corrections are applied. The EGPS-WC score is significantly correlated with all measures of writing quality, with the strongest association between the EGPS-WC and the WIAT-II^{UK} composite measure.

Spelling was significantly correlated with each writing measure at word-, sentence- and paragraph-levels. The measure of handwriting was significantly correlated with each level of writing and demonstrates a significant relationship with spelling. Reading was also significantly correlated with performance on the sentence- and paragraph-level tasks, but not at word-level. Spelling also demonstrated a strong relationship with reading.

9.4.2 Multiple Regressions

Multiple linear regression analysis was used to develop a model for predicting writing quality from performance on the EGPS-WC and the proximal and distal factors. Table 13

shows that each predictor variable demonstrated a significant zero-order correlation. Separate multiple regressions were conducted to examine the independent contributions of each of these measures on writing at the word, sentence, and paragraph level (see Table 14).

A stepwise approach was used for each regression analysis. Age, SEN, EAL and PP were entered at the first step (gender was also included at this step but for the sentence level regression only), followed by the proximal measures in the second step. For proximal measures of writing, HAST-II spelling was entered as the measure of spelling and the alphabet task was entered as a measure of handwriting fluency. Next the distal factor of reading was added as an additional step into the model. Finally, based on the findings of the earlier factor analysis, the EGPS-WC score was used as a distal predictor to establish the additional unique contribution of EGPS-WC. A summary of the stepwise regressions is reported in Table 14.

9.4.2.1 Predicting Word-Level Writing

The overall model for predicting word-level writing fluency was statistically significant, $F (8, 201) = 4.42$, $p < .001$; $R^2 = .13$, $R^2_{Adjusted} = .10$. Handwriting speed was the unique predictor of word fluency (at Step 4) with a beta weight of (.21).

9.4.2.2 Predicting Sentence-level Writing

The overall model for predicting sentence-level writing quality was significant, $F(9, 200) = 7.28$, $p < .001$; $R^2 = .23$, $R^2_{Adjusted} = .19$. EGPS-WC was found to be the unique predictor of sentence-level writing in the final model with a beta weight of (.28).

9.4.2.3 Predicting Text-level Writing

The overall model for predicting text-level writing quality was significant, $F(8, 201) = 23.92$, $p < .001$; $R^2 = .45$, $R^2_{Adjusted} = .43$. EGPS-WC and Spelling were found to predict text-level quality. The EGPS-WC beta weight (.38) was higher than the spelling beta weight (.33). The addition of EGPS-WC to the prediction of text-level quality (Step 4) led to a statistically significant increase in R^2 of .06, $F(1, 202) = 22.25$, $p < .001$.

9.4.2.4 Predicting Composite Writing Quality

The overall model for predicting composite writing quality was significant, $F (8, 201) = 30.81$, $p < .001$; $R^2 = .52$, $R^2_{Adjusted} = .50$. EGPS-WC and Spelling were found to predict composite writing quality. The EGPS beta weight (.46) was higher than the spelling beta weight (.32). The addition of EGPS-WC to the prediction of composite writing quality (Step 4) led to a statistically significant increase in R^2 of .09, $F(1, 202) = 36.03$, $p < .001$.

Table 14 - Regressions Examining Predictors of writing quality at word, sentence, and paragraph levels, and composite writing quality. (note: presented over two pages)

		Step 1				Step 2				Step 3				Step 4			
		B	Std error	Beta	t	B	Std error	Beta	t	B	Std error	Beta	t	B	Std error	Beta	t
Word Level	SEN	-.53	.61	-.06	-.87	.56	.63	.07	.90	.57	.64	.07	.89	**.62**	**.63**	**.07**	**.98**
	EAL	.31	.53	.04	.59	.22	.51	.03	.43	.21	.51	.03	.42	**-.02**	**.52**	**.00**	**-.04**
	PP	.58	.52	.08	1.12	.65	.50	.09	1.31	.65	.50	.09	1.31	**.80**	**.50**	**.11**	**1.60**
	Spelling					.12	.04	.20	2.77*	.12	.05	.20	2.37	**.09**	**.05**	**.16**	**1.89**
	Handwriting					.05	.01	.23	3.27***	.05	.01	.23	3.25***	**.04**	**.01**	**.21**	**3.09*****
	Reading									.00	.03	.00	.05	**-.05**	**.04**	**-.13**	**-1.31**
	EGPS-WC													**.82**	**.37**	**.23**	**2.25**
	R^2	.01				.11				.11				**.13**			
	R^2 Change					11.47***				.003				**.02**			
	F	.76				5.09***				4.22***				**4.42*****			
Sentence Level	SEN	-1.53	.43	-.24	-3.54***	-.72	.45	-.11	-1.60	-.61	.46	-.10	-1.34	**-.59**	**.45**	**-.09**	**-1.31**
	EAL	.02	.38	.00	.05	-.09	.37	-.02	-.26	-.15	.37	-.03	-.40	**-.37**	**.37**	**-.07**	**-1.02**
	PP	-.14	.37	-.03	-.38	-.04	.36	-.01	-.13	-.01	.36	.00	-.03	**.12**	**.35**	**.02**	**.35**
	Gender	-.98	.36	-.18	-2.75	-.96	.34	-.18	-2.82**	-.95	.34	-.18	-2.79**	**-.83**	**.34**	**-.16**	**-2.46**
	Spelling					.12	.03	.27	3.84***	.09	.04	.21	2.58*	**.07**	**.04**	**.16**	**1.98**
	Handwriting					.02	.01	.11	1.59	.02	.01	.12	1.74	**.02**	**.01**	**.10**	**1.57**
	Reading									.03	.02	.12	1.53	**-.01**	**.03**	**-.04**	**-.45**
	EGPS-WC													**.73**	**.26**	**.28**	**2.82*****
	R^2	.10				.19				.19				**.23**			
	R^2 Change					.08***				.009				**.03*****			
	F	5.81***				7.66***				6.94***				**7.28*****			
Paragraph Level	SEN	-2.53	.58	-.29	-4.37***	-.70	.54	-.08	-1.30	-.35	.53	-.04	-.66	**-.27**	**.50**	**-.03**	**-.53**
	EAL	-.09	.51	-.01	-.18	-.41	.43	-.05	-.94	-.57	.42	-.07	-1.35	**-.96**	**.41**	**-.13**	**-2.34**
	PP	-.51	.50	-.07	-1.04	-.22	.43	-.03	-.52	-.12	.41	-.02	-.28	**.13**	**.40**	**.02**	**.32**
	Spelling					.31	.04	.54	8.47***	.23	.04	.39	5.60***	**.19**	**.04**	**.33**	**4.81*****
	Handwriting					.01	.01	.04	.64	.01	.01	.06	1.08	**.01**	**.01**	**.04**	**.75**
	Reading									.10	.03	.28	4.06***	**.02**	**.03**	**.05**	**.59**
	EGPS-WC													**1.36**	**.29**	**.38**	**4.72*****
	R^2	.09				.34				.39				**.45**			
	R^2 Change					.25***				.05***				**.06*****			
	F	7.14***				21.37***				21.91***				**23.92*****			

Note: * $p < .01$, ** $p < .007$ (Bonferroni's adjustment) *** $p < .001$
Bold indicates the final model at each level

		Step 1				Step 2				Step 3				Step 4			
		B	Std error	Beta	t	B	Std error	Beta	t	B	Std error	Beta	t	B	Std error	Beta	t
Overall Writing Quality	SEN	-4.32	.85	-.33	-5.06***	-1.39	.78	-.11	-1.79	-.91	.76	-.07	-1.20	**-.77**	**.71**	**-.06**	**-1.09**
	EAL	-.06	.75	-.01	-.08	-.52	.63	-.05	-.82	-.73	.61	-.06	-1.20	**-1.42**	**.57**	**-.13**	**-2.48**
	PP	-.53	.73	-.05	-.73	-.13	.61	-.01	-.21	.01	.60	.00	.02	**.45**	**.56**	**.04**	**.80**
	Spelling					.46	.05	.53	8.67***	.35	.06	.40	5.87***	**.28**	**.06**	**.32**	**5.02*****
	Handwriting					.04	.02	.12	2.16	.04	.02	.14	2.60*	**.04**	**.02**	**.12**	**2.33**
	Reading									.14	.04	.25	3.79***	**-.01**	**.04**	**-.02**	**-.31**
	EGPS-WC													**2.44**	**.41**	**.46**	**6.00*****
	R²	.12				.39				.43				**.52**			
	R² Change					.27***				.04***				**.09*****			
	F	9.14***				26.08***				25.54***				**30.81*****			

Note: **Bonferroni adjustment** ** *p* <.007, *** *p* < .001
Bold indicates the final model at each level

100

9.4.3 Path Analysis

The results were then further tested using path analysis to specify relationships among the independent variables in predicting overall writing competency as measured by the WIAT-II^{UK} WES composite score. In doing so, it was hoped to establish the contribution of EGPS-WC amongst the proximal and distal factors within writing. It was reasoned from current literature that reading and grammar and punctuation knowledge would be distal factors, and as such exogenous variables within the model. Based on current literature, it was also hypothesised that handwriting and spelling would be proximal factors, endogenous variables, through which distal factors indirectly influence writing competency.

Model 1

The first model drew upon the established cognitive models of writing in the literature whereby the distal factors of reading and EGPS-WC would work indirectly through both proximal measures of spelling and handwriting. Spelling and handwriting would each then work directly on writing competency. However, this had poor model fit ($\chi^2(3) = 56.96$, $p < .001$, RMR = 6.13, GFI = .01, CFI = .87).

Model 2

A more parsimonious variant of model 1 was then tested where, based on Dockrell (2019), spelling would work indirectly through handwriting. The distal factors of reading and EGPS-WC would work indirectly through both proximal measures of spelling and handwriting, but spelling would also work indirectly through handwriting. While providing better model fit than model 1, it remained to demonstrate poor model fit ($\chi^2(2) = 44.82$, $p < .001$, RMR = 2.29, GFI = .93, CFI = .90).

Model 3

Based on the findings of Kim (2013), the next model tested the possibility of EGPS-WC having both a direct effect on writing quality and an indirect effect through the proximal measures. Retaining the indirect effects of reading and EGPS-WC working through spelling and handwriting, and the indirect effect of spelling through handwriting, this model offered good model fit ($\chi^2(1) = .005$, $p = .94$, RMR = .03, GFI = 1 CFI = 1, r=.50).

Model 4

In seeking a more parsimonious model which can explain the data with a minimum number of predictors (Field, 2009), we then hypothesised that based on the insignificant direct

effects of reading and EGPS-WC on handwriting, that a model where reading and EGPS-WC knowledge would work indirectly through spelling would offer better fit. This model also offered good model fit, ($\chi^2(3) = .2.755$, $p = .43$, RMR = .2.92, GFI = .995 CFI = 1, $r=.50$). This final model explained 50 per cent of the variance in the WIAT-II^{UK} composite writing score. As can be seen in Figure 8, standardised betas offer comparable data for the contributions of each variable.
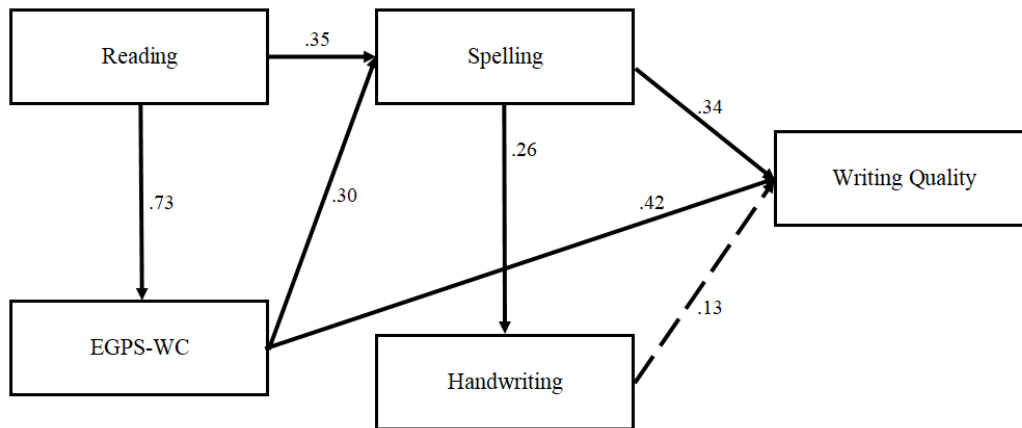
Figure 8 - Standardised regression coefficients and significance levels for proximal and distal factors and WIAT-II^{UK} Written Expression Subtest. Solid lines represent statistically significant paths $p < .001$ and dashed lines represent statistically significant paths $p < .01$.

9.5 RQ3a. To what extent does the EGPS Test account for performance in micro-structural factors of writing products?

The next research question sought to understand how much variance the EGPS Test Paper 1 (EGPS-WC) accounts for within the micro-structural factors of writing performance. For a more detailed analysis of writing, CBM measures were explored in both narrative and expository genres. Zero order correlations between micro-structural factors of writing (e.g., total number of words, number of punctuation marks, number of T-units, mean words per T-unit, proportion of correct word sequences, proportion of correct spelling) as assessed through CBM-W were conducted. The factor loadings of each of the CBM-W measures upon micro-structural factors of writing performance were then explored. Next, zero order correlations between EGPS-WC and micro-structural factors as identified by the principal component analysis were carried out. Finally, multiple regressions were then explored to examine the independent contribution of EGPS-WC within each of the identified micro-structural factors.

102

9.5.1 Associations between microstructural writing measures

The correlation results for the EGPS-WC and CBM-W measures can be found in Table 15. Correlations for narrative writing are found below the diagonal and expository are found above the diagonal line.

As Table 15 shows, there were significant correlations between all CBM narrative measures and EGPS-WC, except for number mean words per T-unit. Alternatively, there were significant correlations between all CBM expository measures and EGPS-WC, except for the number of T-units. For both narrative and expository texts, proportion of CWS correlations with the EGPS-WC scores were the largest.

Table 15 - Bivariate correlations between CBM narrative performance (below diagonal) and expository performance (above diagonal) and EGPS Scores.

	1	2	3	4	5	6	7
1. EGPS Writing Convention Skills		.18*	.33**	.00	.25**	.42**	.34**
2. Total words produced	.26**		.55**	.79**	.05	.18	.17
3. No. punctuation marks	.46**	.60**		.55**	-.07	.30**	.26**
4. No. T-units	.21**	.81**	.59**		-.48**	.01	.07
5. Mean words per T-unit	-.03	.04	-.07	-.46**		.21*	.13
6. Proportion CWS	.54**	.27**	.43**	.17	.11		.88**
7. Proportion Correct Spelling	.40**	.26**	.35**	.19	.07	.84**	

N = 210, Bonferroni's adjustment: *p = .007, **p < .001

9.5.2 Exploratory Factor Analysis

To determine the main components of microstructural factors of writing performance, an EFA with a PCA factor extraction method using varimax rotation was run on the 12 micro-structural measures (6 narrative and 6 expository).

Table 16 - CBM Factor Analysis: Four Factor Solution

| Component | Extraction Sums of Squared Loadings | | | |
	Total	% of Variance	% Cumulative	Total
1	4.55	37.93	37.93	4.55
2	2.61	21.78	59.71	2.61
3	1.22	10.16	69.87	1.22
4	1.15	9.60	79.47	1.15
5	.69	5.74	85.21	
6	.65	5.40	90.61	
7	.53	4.45	95.05	
8	.25	2.11	97.16	
9	.15	1.28	98.44	
10	.10	.80	99.25	
11	.05	.39	99.64	
12	.04	.36	100.00	

Extraction Method: Principal Component Analysis.

As shown in Table 16, the PCA resulted in four components with eigenvalues >1, accounting for 79% of the variance. The four components represented four constructs of writing (see Table 17).

Productivity measures of total words, total punctuation, Total T-units from both genres loaded upon the first factor, accounting for 38% of the variance. Accuracy measures %CWS and % Spelling from both genres loaded upon the second factor, accounting for 22% of the variance. The expository measure of complexity, mean words per T-unit, accounting for 10% of the variance, loaded upon the third factor, while the narrative measure of complexity, mean words per T-unit, accounting for 10% of the variance loaded upon the fourth factor. Loadings were broadly similar between narrative and expository texts.

Table 17 - PCA findings for the CBM-W tasks

	Component Scores			
	Productivity	Accuracy	Complexity E	Complexity N
CBM N Total Words	**.86**	.12	.22	-.09
CBM E Total Words	**.86**	.07	-.15	.20
CBM N Total Punctuation	**.67**	.35	.04	-.20
CBM E Total Punctuation	**.62**	.29	-.31	.02
CMB N Total T-Units	**.77**	.08	.06	-.56
CBM E Total T-Units	**.71**	-.03	-.54	.08
CBM N % CWS	.20	**.85**	.09	.07
CBM E % CWS	.11	**.90**	.08	.02
CBM N % Spelling	.14	**.87**	.01	.00
CBM E % Spelling	.09	**.88**	.00	.01
CBM N Mean Words per T-Unit	-.02	.08	.13	**.92**
CBM E Mean Words per T-Unit	.04	.14	**.89**	.15

Note. Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. Rotation converged in 18 iterations.
E = Expository; N = Narrative; **Bold** indicates >.6

9.5.3 Associations between EGPS, Micro-structural Measures, and Proximal and Distal Factors

Table 18 shows the correlations between the measure of EGPS Writing Convention Skills, the proximal and distal measures, and the four components of microstructural variance: productivity, accuracy, expository complexity, and narrative complexity. Bonferroni's adjustment ($p \leq .006$) was used to control for multiple correlations.

EGPS-WC, handwriting, spelling and reading comprehension were associated to *productivity*, with all measures demonstrating weak correlations. EGPS-WC, spelling and reading comprehension were also associated with *accuracy*. Spelling and EGPS-WC offered strong correlations with accuracy, with reading demonstrating a more moderate relationship. Handwriting demonstrated a relationship with accuracy, but this was not significant. EGPS-WC and reading comprehension were associated with *expository complexity;* however no measures demonstrated significant associations with *narrative complexity*.

Table 18 - Zero order correlations between EGPS, proximal and distal factors, and microstructural factors

	1	2	3	4	5	6	7	8
1. EGPS Writing Convention Skills	-	.15	$.55^{**}$	$.73^{**}$	$.25^{**}$	$.46^{**}$	$.19^{*}$	-.07
2. DASH - Alphabet Raw Score		-	$.26^{**}$	.08	$.22^{**}$	.14	-.03	-.07
3. HAST-2 Spelling Raw score			-	$.57^{**}$	$.24^{**}$	$.66^{**}$	.11	.01
4. Reading Raw Score				-	$.24^{**}$	$.37^{**}$	$.22^{**}$	-.01
5. Productivity FS					-	-	-	-
6. Accuracy FS						-	-	-
7. Expository Complexity FS							-	-
8. Narrative Complexity FS								-

N = 210, Bonferroni's adjustment: $*p = .006$, $**p < .001$
FS = Factor score

9.5.4 Multiple Regressions

Multiple linear regression analysis was used to develop a model for predicting competence at the micro-structural level of writing from performance on the EGPS Test and proximal and distal factors. Separate multiple regressions were conducted to examine the independent contributions of each of these measures on writing productivity and accuracy factors. Owing to the low correlations for narrative and expository complexity factor scores with proximal and distal factors, no further regressions were run using these factors of micro-structural writing.

A stepwise approach was used where age, SEN, EAL and PP were entered at the first step, followed by the proximal measures of spelling and handwriting in the second step. Next the distal factor of reading was added as the third step into the model. Based on the findings of the earlier factor analysis, the EGPS-WC score was used as a distal predictor and was subsequently added as the final step. A summary of the stepwise regressions is reported in Table 19.

9.5.4.1 Predictors for Productivity

The final model for predicting writing productivity was statistically significant, $F(8, 201) = 4.00$, $p < .001$; $R^2 = .12$, $R^2_{Adjusted} = .09$. Only handwriting was found to be a significant predictor of writing productivity (Step 4), although this did not reach significance after applying Bonferroni's adjustment.

9.5.4.2 Predictors for Accuracy

The final model for predicting writing accuracy was statistically significant, $F(8, 201)$ = 25.95, $p < .001$; $R^2 = .47$, $R^2_{Adjusted} = .46$. Writing accuracy was explained by spelling and EGPS-WC. The spelling beta weight (.61) was higher than the EGPS (.19) beta weight.

Table 19 - Regressions Examining Predictors of CBM Writing Productivity and Accuracy Factors

		Step 1				Step 2				Step 3				Step 4			
		B	Std error	Beta	t	B	Std error	Beta	t	B	Std error	Beta	t	B	Std error	Beta	t
Productivity	SEN	-.34	.17	-.14	-2.01	-.09	.18	-.04	-.49	-.03	.18	-.01	-.15	**-.02**	**.18**	**-.01**	**-.11**
	EAL	.22	.15	.10	1.48	.19	.14	.09	1.34	.17	.14	.08	1.16	**.14**	**.15**	**.06**	**.95**
	PP	.15	.14	.07	1.01	.16	.14	.08	1.16	.18	.14	.09	1.30	**.20**	**.14**	**.09**	**1.41**
	Spelling					.03	.01	.18	2.39	.01	.01	.09	1.06	**.01**	**.01**	**.07**	**.85**
	Handwriting					.01	.00	.17	2.46	.01	.00	.19	2.68*	**.01**	**.00**	**.18**	**2.59***
	Reading									.02	.01	.17	2.06	**.01**	**.01**	**.11**	**1.09**
	EGPS													**.10**	**.10**	**.10**	**.93**
	R^2	.04				.08				.12				**.12**			
	R^2 Change					.07**				.02				**.004**			
	F	2.49				4.53***				4.54***				**4.00***			
Accuracy	SEN	-.75	.16	-.31	-4.75***	-.20	.14	-.08	-1.45	-.22	.14	-.09	-1.54	**-.21**	**.14**	**-.08**	**-1.48**
	EAL	.31	.14	.14	2.22	.20	.11	.09	1.82	.21	.11	.10	1.87	**.16**	**.11**	**.07**	**1.38**
	PP	-.23	.14	-.11	-1.68	-.13	.11	-.06	-1.17	-.13	.11	-.06	-1.21	**-.10**	**.11**	**-.05**	**-.91**
	Spelling					.10	.01	.62	10.79***	.10	.01	.65	9.69***	**.10**	**.01**	**.61**	**9.12***
	Handwriting					.00	.00	-.04	-.75	.00	.00	-.04	-.82	**.00**	**.00**	**-.05**	**-1.01**
	Reading									.00	.01	-.05	-.71	**-.02**	**.01**	**-.16**	**-2.00**
	EGPS													**.19**	**.08**	**.19**	**2.38****
	R^2	.14				.46				.46				**.47**			
	R^2 Change					.32***				.001				**.02***			
	F	11.39***				34.38***				28.67***				**25.95***			

Note: * $p < .01$, ** $p < .007$ (Bonferroni's adjustment) *** $p < .001$
Bold indicates the final model at each level

9.6 RQ3b. To what extent can the EGPS Test predict micro-structural performance in expository and narrative genres in writing once proximal and distal factors have been accounted for?

In order to further explore the unique contributions of EGPS-WC for each narrative and expository genre separately, similar principal component analyses were run for each genre. In the initial exploration of PCA for each of the genres, only 2 factors (productivity and accuracy) demonstrated eigenvalues >1 for both genres. However, a third factor (complexity) demonstrated an eigenvalue of .95 (just below the required criterion of >1) which explained 16% of the variance in both genres, with the scree plots offering an inconclusive point of inflection. In line with Jolliffe (1972), who reports that Kaiser's eigenvalue criterion of >1 is too strict for sample sizes greater than 200, factors with eigenvalues < .7 were retained in the analysis.

For each genre of writing, multiple linear regression analysis was used to develop a model for predicting competence at the micro-structural level of from EGPS-WC performance and proximal and distal factors. Separate multiple regressions were conducted to examine the independent contributions of each of these measures on both narrative and expository productivity and accuracy factors, and also in the case of on the expository genre, complexity factor. A stepwise approach was used in-line with previous CBM regression analyses (see Research Question 2a).

9.6.1 CBM Narrative PCA

As shown in Table 20, the PCA resulted in three components with eigenvalues >.7, accounting for 89% of the variance. The three components represented three constructs of writing (see Table 21).

Table 20 - CBM Narrative - Factor Analysis: Three Factor Solution

Component	Total	% of Variance	% Cumulative	Total
		Extraction Sums of Squared Loadings		
1	2.82	47.05	47.05	2.82
2	1.60	26.62	73.68	1.60
3	.95	15.80	89.48	.95
4	.43	7.19	96.67	
5	.15	2.52	99.19	
6	.05	.81	100.00	

Extraction Method: Principal Component Analysis.

Productivity measures of total words, total punctuation, Total T-units loaded upon the first factor, accounting for 47% of the variance. Accuracy measures %CWS and % Spelling loaded upon the second factor, accounting for 27% of the variance. The complexity measure, mean words per T-unit, accounted for 16% of the variance when loaded upon the third factor.

Table 21 - PCA CBM- Narrative

	Component Scores		
	Productivity	Accuracy	Complexity
CBM N Total Words	**.94**	.10	.10
CBM N Total Punctuation	**.76**	.34	-.02
CMB N Total T-Units	**.88**	.05	-.42
CBM N % CWS	.17	**.94**	.07
CBM N % Spelling	.14	**.94**	.01
CBM N Mean Words per T-Unit	-.07	.06	**.99**

Note. Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. Rotation converged in 5 iterations.
N = Narrative; **Bold** indicates >.6

9.6.1.1 Associations between EGPS, Narrative Micro-structural Measures, and Proximal and Distal Factors

Table 22 shows the correlations between the children's measures of EGPS Writing Convention Skills, the proximal and distal measures, and the three components of narrative microstructural variance: productivity, accuracy and complexity. Bonferroni's adjustment ($p \leq$ .007) was used to control for multiple correlations. EGPS, handwriting, spelling and reading comprehension were associated to *productivity*, with all measures demonstrating weak correlations. EGPS, spelling and reading comprehension were also associated with *accuracy*. Spelling and EGPS offered strong correlations with accuracy, with reading demonstrating a more moderate relationship. No measures were associated with *complexity*. Owing to the low correlations for complexity factor scores with proximal and distal factors for narrative genre, no regression analyses were run for *narrative complexity*. A summary of stepwise regressions is reported in Table 23.

Table 22 - Zero order correlations between EGPS, proximal and distal factors, and Narrative microstructural factors

	1	2	3	4	5	6	7
1. EGPS Writing Convention Skills	-	.15	.55**	.73**	.26**	.47**	-.02
2. Handwriting		-	.26**	.08	.25**	.10	-.03
3. Spelling			-	.57**	.23**	.64**	.03
4. Reading Comprehension				-	.25**	.38**	.04
5. Narrative Productivity FS					-	-	-
6. Narrative Accuracy FS						-	-
7. Narrative Complexity FS							-

Note. N = 210, Bonferroni's adjustment: *p = .007, **p < .001
FS = Factor score

9.6.1.2 Predictors for Narrative Productivity

The final model for predicting writing productivity was statistically significant, $F(8, 201) = 4.42$, $p < .001$; $R^2 = .13$, $R^2_{Adjusted} = .10$. Only handwriting was found to be a significant predictor of narrative productivity.

9.6.1.3 Predictors for Narrative Accuracy

The final model for predicting writing accuracy was statistically significant, $F(8, 201) = 21.67$, $p < .001$; $R^2 = .46$, $R^2_{Adjusted} = .44$. Narrative accuracy was explained by spelling as measured by the HAST-2 and EGPS-WC. The spelling beta weight (.58) was higher than EGPS-WC (.23) beta weight.

Table 23 - Regressions Examining Predictors of CBM Narrative Productivity and Accuracy Factors

		Step 1				Step 2				Step 3				Step 4			
		B	Std error	Beta	t	B	Std error	Beta	t	B	Std error	Beta	t	B	Std error	Beta	t
Productivity	SEN	-.26	.17	-.11	-1.56	.00	.18	.00	.02	.07	.18	.03	.41	**.08**	**.18**	**.03**	**.46**
	EAL	.19	.15	.09	1.29	.17	.14	.08	1.17	.14	.14	.06	.96	**.10**	**.14**	**.05**	**.68**
	PP	.10	.15	.05	.72	.12	.14	.06	.85	.14	.14	.07	1.02	**.16**	**.14**	**.08**	**1.17**
	Spelling					.03	.01	.17	2.34	.01	.01	.07	.85	**.01**	**.01**	**.05**	**.59**
	Handwriting					.01	.00	.21	2.96**	.01	.00	.22	3.22***	**.01**	**.00**	**.22**	**3.12****
	Reading									.02	.01	.20	2.41	**.01**	**.01**	**.12**	**1.18**
	EGPS													**.13**	**.10**	**.13**	**1.25**
	R^2	.02				.10				.13				**.13**			
	R^2 Change					.08***				.03				**.007**			
	F	1.60				4.58***				4.88***				**4.42***			
Accuracy	SEN	-.72	.16	-.30	-4.54***	-.20	.14	-.08	-1.41	-.21	.14	-.08	-1.44	**-.19**	**.14**	**-.08**	**-1.37**
	EAL	.35	.14	.16	2.53	.25	.11	.12	2.18	.25	.11	.12	2.20	**.19**	**.11**	**.09**	**1.62**
	PP	-.25	.14	-.12	-1.81	-.15	.11	-.07	-1.33	-.15	.11	-.07	-1.35	**-.11**	**.11**	**-.05**	**-.99**
	Spelling					.10	.01	.60	10.29***	.10	.01	.62	9.08***	**.09**	**.01**	**.58**	**8.48****
	Handwriting					.00	.00	-.07	-1.26	.00	.00	-.07	-1.29	**.00**	**.00**	**-.08**	**-1.53**
	Reading									.00	.01	-.02	-.36	**-.02**	**.01**	**-.16**	**-1.97**
	EGPS													**.23**	**.08**	**.23**	**2.80****
	R^2	.14				.44				.44				**.46**			
	R^2 Change					.29***				.00				**.02***			
	F	11.46***				31.69***				26.31***				**24.44***			

Note: * p <.01, ** p < .007 (Bonferroni's adjustment) *** p < .001
Bold indicates the final model at each level

9.6.2 CBM Expository PCA

As shown in Table 24, the PCA resulted in three components with eigenvalues >.7, accounting for 92% of the variance. The three components represented three constructs of writing (see Table 25).

Table 24 - CBM Expository - Factor Analysis: Three Factor Solution

Component	Total	% of Variance	% Cumulative	Total
		Extraction Sums of Squared Loadings		
1	2.52	42.04	42.04	2.52
2	1.88	31.30	73.34	1.88
3	.95	15.84	89.18	.95
4	.49	8.16	97.34	
5	.11	1.89	99.24	
6	.05	.76	100.00	

Extraction Method: Principal Component Analysis.

Productivity measures of total words, total punctuation, Total T-units loaded upon the first factor, accounting for 42% of the variance. Accuracy measures %CWS and % Spelling loaded upon the second factor, accounting for 31% of the variance. The complexity measure, mean words per T-unit, accounted for 16% of the variance when loaded upon the third factor.

Table 25 - PCA CBM- Expository

	Productivity	Accuracy	Complexity
	Component Scores		
CBM E Total Words	**.94**	.05	.11
CBM E Total Punctuation	**.76**	.25	-.04
CMB E Total T-Units	**.86**	-.03	-.45
CBM E % CWS	.11	**.96**	.12
CBM E % Spelling	.10	**.96**	.02
CBM E Mean Words per T-Unit	-.07	.10	**.99**

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. Rotation converged in 5 iterations.
E = Expository; **Bold** >.6

9.6.2.1 Associations between EGPS, Expository Micro-structural Measures, and Proximal and Distal Factors

Table 26 shows the correlations between the children's measures of EGPS Writing Convention Skills, the proximal and distal measures, and the three components of expository microstructural variance: productivity, accuracy and complexity. Bonferroni's adjustment (p ≤

.007) was used to control for multiple correlations. EGPS, handwriting, spelling and reading comprehension were associated to *productivity*, with all measures demonstrating weak correlations. EGPS-WC, spelling and reading comprehension were also associated with *accuracy*. Spelling offered strong correlations with accuracy, with EGPS-WC and reading demonstrating a more moderate relationship. EGPS-WC, spelling and reading comprehension all offered weak correlations with *complexity*.

Table 26 - Zero order correlations between EGPS, proximal and distal factors, and Expository microstructural factors

	1	2	3	4	5	6	7
1. EGPS Writing Convention Skills	-	.15	.55**	.73**	.19**	.37**	.24**
2. Handwriting		-	.26***	.08	.17*	.15	-.06
3. Spelling			-	.57**	.23**	.58**	.18
4. Reading Comprehension				-	.18	.31**	.27**
5. Expository Productivity FS					-	-	-
6. Expository Accuracy FS						-	-
7. Expository Complexity FS							-

N = 210, Bonferroni's adjustment: *p = .007, **$p < .001$
FS = Factor score

9.6.2.2 Predictors for Expository Productivity

The final model for predicting expository productivity was statistically significant, $F(8, 201) = 2.78$, $p = .013$; $R^2 = .09$, $R^2_{Adjusted} = .06$. However, no significant predictors were identified in the final model.

9.6.2.3 Predictors for Expository Accuracy

The final model for predicting writing accuracy was statistically significant, $F(8, 201) = 16.18$, $p < .001$; $R^2 = .36$, $R^2_{Adjusted} = .34$. Expository accuracy was uniquely explained by spelling with a high beta weight (.56).

9.6.2.4 Predictors for Expository Complexity

The final model for predicting writing productivity was statistically significant, $F(8, 201) = 3.10$, $p = .007$; $R^2 = .10$, $R^2_{Adjusted} = .07$. No significant predictors were identified in the final model.

Table 27 - Regressions Examining Predictors of CBM Expository Productivity, Accuracy and Complexity Factors

		Step 1				Step 2				Step 3				Step 4			
		B	Std error	Beta	t	B	Std error	Beta	t	B	Std error	Beta	t	B	Std error	Beta	t
Productivity	SEN	-.37	.17	-.15	-2.21	-.16	.18	-.06	-.87	-.13	.18	-.05	-.71	**-.13**	**.18**	**-.05**	**-.69**
	EAL	.24	.15	.11	1.62	.21	.14	.10	1.47	.20	.15	.09	1.38	**.19**	**.15**	**.09**	**1.26**
	PP	.14	.14	.06	.94	.16	.14	.07	1.10	.16	.14	.08	1.16	**.17**	**.14**	**.08**	**1.20**
	Spelling					.03	.01	.17	2.29	.02	.01	.13	1.53	**.02**	**.01**	**.12**	**1.41**
	Handwriting					.01	.00	.11	1.56	.01	.00	.12	1.65	**.01**	**.00**	**.11**	**1.60**
	Reading									.01	.01	.08	.90	**.00**	**.01**	**.05**	**.46**
	EGPS													**.05**	**.11**	**.05**	**.44**
	R^2	.04				.08				.09				**.09**			
	R^2 Change					.04*				.004				**.001**			
	F	2.92***				3.71**				3.23**				**2.78***			
Accuracy	SEN	-.68	.16	-.28	-4.18***	-.18	.15	-.07	-1.17	-.20	.15	-.08	-1.30	**-.19**	**.15**	**-.08**	**-1.25**
	EAL	.22	.14	.10	1.55	.13	.12	.06	1.05	.14	.12	.06	1.13	**.10**	**.12**	**.05**	**.82**
	PP	-.16	.14	-.08	-1.18	-.08	.12	-.04	-.65	-.08	.12	-.04	-.71	**-.06**	**.12**	**-.03**	**-.52**
	Spelling					.09	.01	.54	8.63***	.09	.01	.58	7.92***	**.09**	**.01**	**.56**	**7.49***
	Handwriting					.00	.00	.00	-.04	.00	.00	-.01	-.13	**.00**	**.00**	**-.01**	**-.25**
	Reading									-.01	.01	-.06	-.89	**-.01**	**.01**	**-.14**	**-1.57**
	EGPS													**.13**	**.09**	**.13**	**1.43**
	R^2	.10				.35				.35				**.36**			
	R^2 Change					.25***				.003				**.007**			
	F	7.91***				21.99***				18.44***				**16.18***			
Complexity	SEN	-.32	.17	-.13	-1.91	-.21	.18	-.09	-1.17	-.13	.18	-.05	-.70	**-.12**	**.18**	**-.05**	**-.66**
	EAL	-.03	.15	-.01	-.21	-.07	.15	-.03	-.46	-.11	.14	-.05	-.73	**-.14**	**.15**	**-.06**	**-.94**
	PP	.07	.15	.03	.46	.10	.14	.05	.72	.13	.14	.06	.91	**.15**	**.14**	**.07**	**1.05**
	Spelling					.03	.01	.19	2.50	.01	.01	.07	.76	**.01**	**.01**	**.05**	**.52**
	Handwriting					-.01	.00	-.12	-1.75	-.01	.00	-.10	-1.47	**-.01**	**.00**	**-.11**	**-1.56**
	Reading									.03	.01	.24	2.89**	**.02**	**.01**	**.17**	**1.65**
	EGPS													**.12**	**.10**	**.12**	**1.11**
	R^2	.01				.05				.09				**.10**			
	R^2 Change					.04				.04**				**.005**			
	F	1.24				2.33				3.41**				**3.10***			

Note: * p <.01, ** p < .007 (Bonferroni's adjustment) *** p < .001
Bold indicates the final model at each level

9.7 RQ4. Can the EGPS Test Identify struggling writers?

9.7.1 Writing Performance of Typical and Struggling Writers

Struggling writers were identified as scoring below 1 SD from the mean (< 85) on the WIAT-II[UK] composite task. Table 28 provides means (SDs) for the population Z scores for typical writers (TW) and struggling writers (SW) within all assessment measures. These results confirm the groupings of TW and SW as identified by a score of < 85 on the composite WIAT-II[UK] Written Expression Sub-test. As can be seen in the table, for most measures mean Z scores for SW were negative, although this was not the case for narrative mean words per T-unit, or expository number of T-units. However, it is also important to note the small sample of struggling writers ($n = 21$). Of this struggling writers sample, n=10 were also identified as part of the SEN group.

Table 28 – Descriptive Statistics for EGPS, Writing, Proximal and Distal Measures (N=210)

	Typical Writers (n=189)		Struggling Writers (n=21)	
	Mean	(SD)	Mean	(SD)
EGPS 2018 Test				
EGPS Raw Combined Total	.15	(.92)	-1.31	(.67)
Paper 1: Grammar and Punctuation	.12	(.95)	-1.12	(.73)
Paper 2: Spelling	.15	(.92)	-1.34	(.65)
Writing Measures				
WIAT-II^{UK} Standard Score	.20	(.83)	-1.83	(.34)
WIAT-II^{UK} Composite Raw Score	.20	(.84)	-1.80	(.36)
WIAT-II^{UK} Word Fluency Raw Score	.09	(.97)	-.84	(.86)
WIAT-II^{UK} Sentence Combination Raw Score	.14	(.94)	-1.28	(.56)
WIAT-II^{UK} Paragraph Task Raw Score	.17	(.89)	-1.54	(.52)
CBM-N Total words produced	.07	(.97)	-.60	(1.06
No. punctuation marks	.08	(1.01)	-.69	(.51)
No. T-units	.06	(.99)	-.58	(.88)
Mean words per T-unit	-.03	(.96)	.25	(1.29)
Proportion CWS	.14	(.85)	-1.22	(1.40)
Proportion Correct Spelling	.08	(.96)	-.70	(1.14)
CBM-E Total words produced	.06	(.99)	-.50	(1.01)
No. punctuation marks	.05	(1.02)	-.47	(.61)
No. T-units	.02	(.99)	-.18	(1.07)
Mean words per T-unit	.05	(1.00)	-.44	(.89)
Proportion CWS	.10	(.94)	-.93	(1.09)
Proportion Correct Spelling	.07	(.98)	-.59	(1.03)
Proximal Factors				
DASH Handwriting – Copy Task Raw Score	.06	(1.00)	-.51	(.86)
DASH Handwriting – Alphabet Task Raw Score	.04	(.99)	-.35	(1.06)
HAST-II Spelling Raw Score	.14	(.91)	-1.25	(.90)
Distal Factors				
NFER Reading Raw Score	.08	(.99)	-.71	(.78)

9.7.2 ROC Analysis

We next wanted to understand if struggling writers could be distinguished by their performance on the combined EGPS 2018 Test. To test this, a ROC analysis was computed (Dockrell et al., 2019) to establish if the combined EGPS 2018 Test raw score could offer greater discriminate validity for SWs than other proximal, distal, or accuracy measures (CBM-N & E %CWS). Table 29 presents the area under the curve and 95 per cent confidence intervals.

Table 29 – ROC analysis for EGPS Measures

				95% Confidence	
				Lower	Upper
EGPS Measure	Area	Std. Error[a]	Sig.	Bound	Bound
EGPS Combined Raw Score	.89	.04	< .001	.82	.97
HAST-2 Spelling - Raw Score	.87	.03	< .001	.81	.94
DASH - Alphabet - Raw Score	.63	.07	.063	.49	.76
NGRT Reading - Raw Score	.74	.05	< .001	.63	.84
CBM N % CWS	.79	.06	< .001	.69	.90
CBM E % CWS	.80	.05	< .001	.70	.90

Note: E = Expository N = Narrative

Both EGPS Test and HAST-2 Spelling measures demonstrated good (A > .85) sensitivity and specificity that is required to be regarded a reliable discriminator of struggling writers (Swets, Dawes, & Monahan, 2000). The findings demonstrate that a child scoring poorly on the EGPS Test is more likely than 89% of the controls to be identified as a struggling writer. Similarly, a child scoring poorly on the HAST-2 Spelling test is more likely than 87% of the controls to be identified as a struggling writer. Reading and %CWS accuracy measures offered fair sensitivity in identifying SWs, and handwriting offered weak sensitivity and specificity. For children in Year 6, the EGPS Test (combined raw score) provided the best measure to discriminate SWs in our sample compared to the other given measures.

CHAPTER 10 : Discussion

10.1 Introduction

The English Grammar, Punctuation and Spelling Test (EGPS) was introduced in 2013 as a statutory assessment for all Year 6 pupils in England, replacing the statutory assessment of National Curriculum Writing. The test was designed to ascertain if children's writing convention skills – grammar, punctuation, and spelling – are of an expected national standard. In some cases, the EGPS Test has been used as a proxy for writing. However, in-line with a long-standing history of debate regarding the importance of grammar and its testing within the curriculum, criticisms have been raised that the EGPS lacks convergent validity with writing (Education Select Committee, 2016). Historic discussions around the importance of grammar have been found to lack theoretical grounding and, of the limited research which has examined this relationship, little has been able to demonstrate strong empirical rigour (Wyse, 2006) that is favourable for the case of grammar. In contrast, more recent empirical studies into writing convention skills (Canniford, 2019; Daffern et al., 2017) dispute this assumed and non-empirically asserted lack of convergence, rather suggesting that grammar and punctuation jointly contribute to children's writing competencies.

This thesis explored empirically whether a relationship exists between the EGPS Test and children's written products. In doing so, this study sought to better understand the validity of the EGPS Test as an indicator of writing competencies in a sample of 210 children from primary schools across England. Writing competence is not a unitary construct, thus it requires exploration in a variety of ways. In this comprehensive study that extends previous work measuring the component factors in writing, writing competency was operationalised through a variety of macro- and micro-structural measures, exploring the quality, productivity, accuracy, and complexity of children's written products in both narrative and expository genres. To better understand the contributions that grammar and punctuation may offer beyond the established role of spelling within the cognitive writing process, path analyses were conducted to specify the relationships among the proximal and distal factors of writing in predicting overall writing competency. Finally, ROC analyses were used to better understand the discriminant validity of the EGPS Test in identifying struggling writers.

10.2 EGPS Test and its constituent domains

The EGPS Test assesses a range of skills (grammar, punctuation, vocabulary, and spelling) which may be seen as separable. The distinction is clearly drawn between spelling and the remaining components of the test through two distinct papers. While Paper 2 explores only spelling, Paper 1 explores two elements: grammar and punctuation (vocabulary is also included, although this is word-level grammar – see Chapter 5). Thus, an important starting point was to consider whether these domains should have been considered separately in subsequent analyses.

The first research question was designed to identify how many factors within writing the EGPS Test domains were able to explain. The EGPS Paper 1 was first explored to establish if this paper consists of multiple dissociable factors, or whether the two areas tapped one component as identified as w*riting convention skills*. It was hypothesised that two factors would present through data reduction. However, using principal component analysis, each of the seven content domains within Paper 1 loaded upon one single factor, which was theorised as *EGPS-Writing Convention Skills (EGPS-WC).* This identification of commonalities between writing convention skills is consistent with existing literature (Abbott et al., 2010; Berninger, Vaughan, et al., 2002; Daffern et al., 2017), forming part of the skills that underpin written text generation.

Paper 2 of the EGPS Test is a test of spelling. Further exploration of the convergent validity of Paper 2 identified that Paper 2 was highly correlated with a standardised measure of spelling. This is the first study to our knowledge to provide empirical backing regarding the convergence of Paper 2 Spelling and a standardised norm-referenced spelling test.

10.3 The EGPS Test and Writing Quality

A first aim was to establish whether a relationship existed between the combined EGPS Test and children's writing outcomes on a standardised measure of writing. Given the mixed reports regarding transferability of grammar into children's writing, the convergent and criterion validity between the EGPS Test and children's composite scores on the WIAT-II[UK] Written Expression Subtest were explored using Pearson's correlations. Overall, the combined EGPS Test demonstrated convergent validity with writing quality by correlating strongly with the WES for both text-level and composite writing quality, moderately with sentence quality, and weakly with word fluency. This offers further support of Canniford (2019), who identified

the convergence of the EGPS Test and composite writing quality and suggests that the EGPS Test can be used as a proxy for assessment of writing quality.

Writing is a complex and multifaceted skill which can be measured in a multitude of ways at multiple levels. Extant literature regarding the relationship between components skills and writing typically focuses on writing productivity and quality either at sentence- and/or text-levels. This thesis considered these different levels, and the test used to capture writing quality (WIAT-II ^{UK} WES) also allowed for consideration of a composite measure of writing quality, which was inclusive of the multiple levels of writing. Therefore, to make theoretical comparison to extant theory, we explored the role of EGPS-WC in relation to word-, sentence-, and text-levels, as well as composite writing quality using the WIAT-II^{UK} WES. Owing to the established contributions of spelling upon children's writing quality (Berninger, Nielsen, et al., 2008; Dockrell et al., 2015; Graham et al., 1997; Juel, 1988; Mackie & Dockrell, 2004; Olinghouse, 2008; Parker et al., 1991), the subsequent research questions aimed to establish the role of EGPS-WC above and beyond the contribution of spelling. It was predicted that children who performed better on the EGPS Paper 1 would produce higher quality writing products at sentence- and text-levels, as well as on a composite writing test. Based on extant literature, it was expected that handwriting and spelling would be the greatest predictors of word-level fluency (Berninger, 1999; Berninger, Abbott, et al., 2006; Graham et al., 1997; Kim et al., 2011) as these transcriptional skills have been found to predict productivity and compositional fluency.

Associations between the EGPS-WC factor score, writing measures, and proximal and distal factors were then explored. EGPS-WC offered associations with all levels of writing, with strongest relationships demonstrated at the text-level and on the composite writing task. This further confirms the EGPS Test Paper 1, which taps grammar and punctuation, can also be used as a reasonable proxy for assessment of writing quality. However, it should be noted that the combined EGPS Test offers higher associations than Paper 1 alone. This is to be expected as the combined EGPS Test captures multiple component skills which have been found to predict writing quality. By testing multiple component skills which contribute to the translation process (Berninger, Winn, et al., 2006; Hayes, 2012) writing quality could be predicted more accurately.

Regression analyses confirmed the importance of EGPS-WC, independent of spelling, at multiple levels of writing. EGPS-WC was the strongest predictor of writing quality at both sentence- and text-levels, as well as on the composite measure of writing. In line with extant literature, spelling was also a strong predictor of writing quality on the paragraph task and the composite task (Juel, 1988; Parker et al., 1991).

For sentence quality, EGPS-WC was the unique predictor. The significance of the impact that EGPS-WC knowledge may have upon sentence quality is an important discovery in better addressing the under specification of the factors that contribute to text generation processes. Previous research on sentence-level writing has neglected to assess the contribution of grammar and punctuation skills. The lack of spelling as a predictor of sentence quality is also of interest. It may be that by the end of primary education, spelling no longer mediates text generation and that a wider range of distal factors such as grammar and punctuation skills begin to make greater contributions to the writing processes. Not only do these current findings confirm the findings of Canniford (2019), but they also further contribute to the wider body of knowledge that establishes the importance of sentence combining, a crucial element of the sentence quality measure in this study, as an indicator of writing quality (R. Andrews, Torgerson, Beverton, Freeman, et al., 2004). Essentially, children who have greater knowledge of writing convention skills are more likely to combine sentences more competently. As surmised by Andrews (2006: 52), 'sentence-combining suggests a pedagogy of applied knowledge-at its best, applied in situations of contextualised learning'. It appears to be the case that sentence-combining ability acts as a diagnostic of applied knowledge in writing (Canniford, 2019). It is believed that this is the first study to make such an inference regarding the relationship between the EGPS-WC and sentence combination.

For word-level fluency, handwriting was the unique predictor, in line with extant literature (Berninger, 1999; Berninger, Abbott, et al., 2006; Berninger, Vaughan, et al., 1997; Graham et al., 1997; Kim et al., 2011; R. Wagner et al., 2011). While it was predicted that spelling would predict word fluency, this is congruent with the findings of potential decreasing constraint of spelling as found in a study of Korean children as they develop through primary school (Kim & Park, 2019).

For text-level (macro-structural) quality, EGPS-WC explained the greatest variance in the writing. In contrast to the existing body of literature regarding macro-structural writing,

grammar and punctuation knowledge was a stronger predictor than spelling in predicting writing quality. Yet, this confirms and extends to upper primary aged children the unique findings of Arfé et al. (2016) which found grammatical skills to be the greatest predictor of writing quality in Italian middle primary years. While previous studies have explored the relationship between grammar and writing quality, these studies have been concerned with discourse oral language skills (Kim et al., 2014a) or syntactic knowledge (Berninger et al., 2011). Where Daffern (2017) found that spelling, punctuation, and grammar jointly influenced compositional writing for children in Years 3 and 5, the present findings further contribute to the literature by identifying the contribution of writing convention skills to writing quality for Year 6 children within an English context.

Finally, composite analytic writing quality was also found to be predicted by EGPS-WC and Spelling, although EGPS-WC remined the strongest predictor. When jointly considering all levels of writing in this way, the role of EGPS-WC was accentuated by a greater beta weight, further confirming the importance of grammar and punctuation in children's multi-level writing quality.

10.4 Path Analysis

Following the methodology of Dockrell (2019), a theoretical distinction was drawn between the proximal and distal factors that impact upon writing quality. Path analysis was used to empirically test models in relation to the contribution of these proximal and distal factors. In doing so, the direct and indirect contributions of each of the selected factors upon writing quality were explored. It was predicted that EGPS-WC, a combination of the theorised distal measures of grammar and punctuation, would work indirectly through handwriting and spelling.

Within the final model, the proximal measure of spelling offered both a direct and indirect contribution, through handwriting, to writing quality. By contrast, handwriting only contributed directly to writing quality. Both of these findings confirm the recent findings of Dockrell (2019), reaffirming the importance of transcription skills in upper primary education. However, where Dockrell found that handwriting in the upper years of primary contributed to sentence writing, this study extends these findings by offering that handwriting contributes to composite writing quality in line with extant literature (R. A. Alves & Limpo, 2015; Berninger, Vaughan, et al., 1997).

While we predicted EGPS-WC to work indirectly through transcription skills, EGPS-WC contributed both directly and indirectly through spelling to composite writing quality. It is plausible that the direct link identified between EGPS-WC and writing quality is rooted in children's abilities to combine sentences. This finding builds on an increasing body of knowledge regarding the role of grammar in sentence combination (Andrews et al., 2006; Berninger et al., 2011; Arfé and Pizzocaro, 2016; Dockrell et al., 2019). Furthermore, the indirect link between EGPS-WC and writing quality, which has been identified to work through spelling, could be explained by the morphological elements of the EGPS Test. Morphology, inclusive of inflectional and derivational morphemes within the EGPS Tests, can be seen as word-level grammar, and increased morphological knowledge in primary aged children has been found to impact upon spelling (Bryant & Nunes, 2006). To our knowledge, this is the first study to identify the direct and indirect contributions of grammar and punctuation to children's writing quality.

Even though reading is strongly correlated with writing quality, reading did not contribute to writing quality directly in our final model as suggested by Walter (2021). Rather it only contributed to writing quality indirectly through both EGPS-WC and spelling. This extends the findings of Schoonen (2019) where declarative linguistic knowledge is the likely source of common variance between reading and writing, and it could be reasoned that the EGPS Test is a useful measure in tapping declarative linguistic knowledge. While beyond the scope of this study, this is an interesting development in the understanding of the relationship between reading and writing which should be explored through further research.

Fayol (1997) posited that punctuation is integral to the process of transforming thought into a linear dimension. However, it could be argued that this theoretical framing of thought being transformed into a linear dimension should be expanded to include grammar and punctuation as part of translation processes (Flower & Hayes, 1981). EGPS-WC has been shown to impact upon the micro-structure of written texts in sentence-level writing through sentence quality and narrative accuracy. Most significantly, EGPS-WC was identified as the most significant predictor of writing quality at text-level. We believe that this is the first study to identify this relationship in Year 6 writers.

Models of writing are only recently beginning to consider the role of language-specific factors in their understanding of how writing develops. However, this study demonstrates that

EGPS-WC offers a strong case for inclusion of grammar and punctuation within our understanding of the text generation process.

10.4 The EGPS Test and Micro-structural features of writing

The next research question sought to establish the role of EGPS-WC in relation to micro-structural features of writing. From the literature, three dimensions have been found to capture microstructure of writing products within both narrative and expository writing: productivity, accuracy, and complexity (Arfé et al., 2016; Puranik et al., 2008; R. Wagner et al., 2011). Drawing on Wagner's (2011) specific model of development of written language, Curriculum Based Measures were used to explore these dimensions.

Exploratory factor analysis, using a principal component analysis method, confirmed the a priori distinction between use of the selected measures to tap these dimensions within both narrative and expository writing samples. The factors identified were *productivity*, *accuracy*, *narrative complexity*, *and expository complexity*. Interestingly, writing *complexity* was identified as two unique factors within the data, differentiated by genre, whereas *productivit*y and *accuracy* were identified as unified constructs across both genres. This study has confirmed the ability of CBM-W measures to tap the dimensions of productivity and accuracy as identified by Dockrell et al. (2015), yet it goes further to demonstrate CBM-W measure may also tap the dimension of complexity within genres.

Most of the variance in expository writing was explained by the *accuracy* factor. Accuracy in expository writing was uniquely predicted by spelling. Most of the variance in narrative writing was also explained by the *accuracy* factor. However, for narrative writing, spelling was the strongest predictor, followed by EGPS-WC.

The impact of spelling upon writing accuracy is in line with existing literature (Arfé et al., 2016; Berninger, 2000; Juel, 1988; Limpo & Alves, 2013) which explains the importance of spelling at the micro-structural level with participants in middle and upper primary years. Of interest, EGPS-WC did not predict accuracy of expository writing, but it did contribute to accuracy of narrative writing. This also confirms the findings of Mackie et al. (2013), which found that receptive grammar predicted writing accuracy, but extends this by identifying the contribution of EGPS-WC. It is possible that children with greater grammar and punctuation skills are more likely to produce more accurate single draft narrative texts which are less prone to errors. As narrative writing is produced through informal dialogue and everyday language

125

structures (Ravid & Katzenberger, 1999), this may lead to less constraints on working memory, freeing children to make more accurate use of writing conventions. In a study of children with specific language impairment, Mackie et al (2013) identified that children made a greater number of accuracy errors when attempting sentences of greater linguistic complexity. Negro et al (2005) further confirmed that accuracy errors rooted in subject-verb agreement and noun omission occurred as children attempt complex sentences, citing processing issues as a potential factor. Without a measure of working memory within this study, it is difficult to synthesise these findings with our findings regarding the EGPS Test. However, as expository writing has been identified as more linguistically challenging than narrative writing (Beers & Nagy, 2011; Koutsoftas & Gray, 2012), limitations in vocabulary may also account for the differential contributions of EGPS-WC between genres, whereby expository writing is constrained by vocabulary (Berman & Nir-Sagiv, 2007). It is possible that the role of working memory may also constrain any benefits of EGPS-WC skills, and it is possible that the complexity and brevity of EGPS-WC skills required for competent writers is genre dependent. However, both hypotheses are beyond the scope of this study and should be explored through further research.

We predicted that handwriting would account for the greatest amount of variance in writing productivity and this study further confirms the existing body of literature (Berninger, 1999; Berninger, Abbott, et al., 2006; Berninger, Vaughan, et al., 1997; Graham et al., 1997; Kim et al., 2011; R. Wagner et al., 2011), with handwriting identified as the unique predictor for productivity. However, no significant predictors of productivity were identified when controlling for expository writing. As identified above, expository writing is more linguistically challenging and may present greater difficulties owing to the range of vocabulary and sentence structures required within this specific genre. Subsequently, spelling constraints presented by challenges due to increasing lexical and sematic diversity required within expository writing may become a limiting factor uniquely within expository genre. This should be substantiated through further research.

For writing complexity, no significant predictors were identified overall, or within each of the genres. While we predicted that higher scores on EGPS-WC would be associated with longer mean length of T-units, this was not the case. In support of our findings, despite differences in methods, Kim et al. (2014a) reported that discourse oral language skills (grammar and vocabulary) contributed to quality but not productivity or complexity. In

contrast, whilst using the same analytical methods as this study, Arfe et al. (2016) found that oral and written grammatical skills contributed to productivity, accuracy, and complexity in Italian middle primary writing composition. However, this was again operationalised through receptive oral grammar, rather than declarative visually oriented grammar as utilised in this study. It is theoretically plausible that EGPS-WC could predict complexity of writing, but that this may be limited by age. The two studies offered above explored children in lower and middle primary school. However, a body of research in USA schools exploring narrative and expository complexity has confirmed that mean length of T-units and clause density are limited age to such an extent that intra-grade differences are not typically significant until mid-secondary age (Hunt, 1965; Klecan-Aker & Hedrick, 1985; Nippold, Ward-Lonergan, & Fanning, 2005; Scott, 1988). The contribution of grammar and punctuation upon complexity is not fully understood, and further research should explore its contribution, inclusive of a greater number of distal factors.

10.5 The EGPS Test and Struggling Writers

Finally, this thesis sought to understand if the combined EGPS Test was a sensitive measure of discriminating struggling writers (SW) and their age-matched peers beyond the other proximal and distal factors identified within this study. SWs and typical writers (TW) were distinguished using the WIAT-II[UK], a norm referenced writing measure. Children with a standard score of <85 were identified as SWs. We predicted that the EGPS Test would offer strong sensitivity and specificity in identifying struggling writers, above those of the proximal and distal measures used within this study, and that spelling, reading, and handwriting would also offer high levels of sensitivity and specificity.

As with prior studies, this research confirms existing frameworks regarding the development of writing which explain that SWs are weaker than that age-matched peers in both proximal skills (weaker spelling and handwriting) and distal skills (weaker reading comprehension) (Dockrell et al., 2019; Dockrell et al., 2015; Sumner, Connelly, & Barnett, 2014). This is also the case for EGPS scores on both papers, where SWs are statistically significantly more likely to perform worse than their age-matched peers. Of note, this study differs methodologically from previous research in that it measures reading comprehension as opposed to word reading (Dockrell et al., 2019). It extends the body of literature by identifying differences in achievement in grammar and punctuation between SWs and their age-matched peers.

Contrary to our predictions, only EGPS (combined) and the HAST-II measure of spelling offered a high-level of diagnostic accuracy, with performance on the EGPS Test offering stronger sensitivity and specificity than spelling ability. This is likely to be because the EGPS Test also encompasses spelling through Paper 2. This may account for the higher diagnostic accuracy of the EGPS Test, which contributes both directly, and indirectly through spelling, to writing quality

Reading did not offer a high level of diagnostic accuracy as predicted. There are two potential explanations for this. The first may lay in the choice of measure. In previous studies, single word reading, rather than reading comprehension, was used to predict writing quality. Alternatively, reading comprehension may not demonstrate high diagnostic accuracy as, being a distal factor of writing, it only contributes to writing quality indirectly through spelling and EGPS-WC.

As far as we are aware, this study is the first of its kind to identify the discriminative validity of the EGPS Test in identifying struggling writers. However, these results should be considered in light of the small sample size of SWs.

10.7 Theoretical Implications

10.7.1 EGPS Writing Conventions Skills and Text Generation

The current study offers an important insight to the purposefulness of the EGPS Test, as well as the implications of grammar and punctuation skills on our understanding of the writing process. Existing models of writing (Berninger, 2000; Hayes, 2012; Kim et al., 2011) seek to specify the component skills in the writing process. As seen in Chapter 3, writing proficiency in beginning writers is dependent on translation – the combined processes of transcription and text generation (Abbott & Berninger, 1993; Berninger, 2000; Kim et al., 2011). As children approach the end of primary schooling, greater automaticity of transcription skills is expected. At this point, text generation skills become more important in translating ideas into written products as words, phrases, and sentences (Abbott & Berninger, 1993; Berninger et al., 2011). Further frameworks have then explored the direct and indirect factors that contribute to translation processes (Dockrell et al., 2019; Kim & Schatschneider, 2017). The greatest critique levelled at the current cognitive models of writing is the under-specification of the component skills that underpin text generation processes. As can be seen in Kim's (2017) Direct and Indirect Effects model (see Figure 9) grammar is regarded as a

foundational oral language skill which contributes to text generation at discourse level. The contribution of punctuation is omitted.
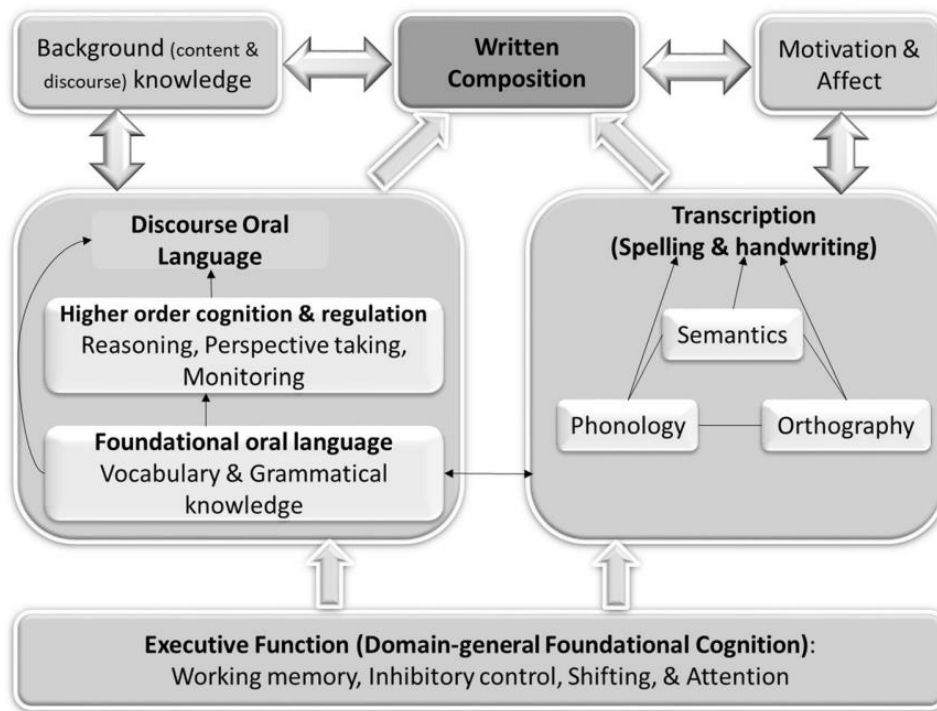

Figure 9 - Direct and Indirect Effects Model of Writing (Kim 2018)

This study has demonstrated the contribution of two writing convention skills to writing quality: grammar and punctuation. Where previous studies have considered grammar and vocabulary as elements of discourse oral language (Kim & Schatschneider, 2017), this study has demonstrated the contribution of these writing convention skills at sentence- and text/discourse-levels. Our path model extended the findings of Dockrell et al. (2019) in identifying the direct and indirect contributions of spelling, as well as the direct contribution of handwriting to composite writing quality. However, it further extends these findings by identifying the direct and indirect contribution of EGPS-WC to composite writing quality. An examination of the EGPS Test has allowed for the exploration of the role of grammar and punctuation skills in text generation processes at word-, sentence- and text/discourse-levels. Figure 10 proposes a model of the direct and indirect factors that contribute to the writing process, inclusive of writing convention skills.
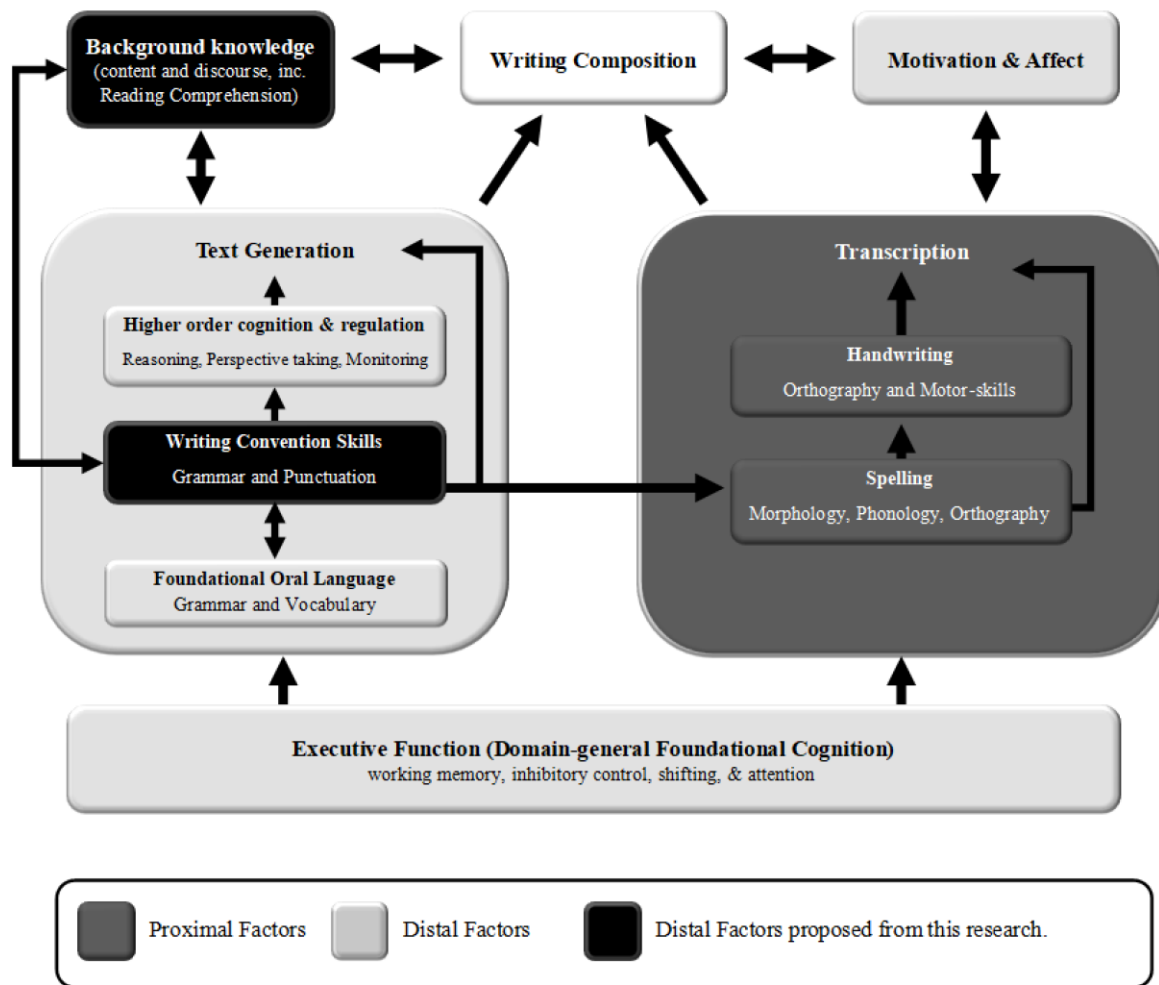
Figure 10 – A Developed Model of Proximal and Distal Factors

This model illustrates how writing convention skills contribute to writing quality as a distal factor involved in the text generation process. Note that this model differs from Kim's (2018) by referring to text generation as a combination of foundational oral language, writing convention skills and higher order cognition and regulation, as opposed to discourse oral language. This model also specifies the indirect contribution of writing convention skills to spelling. Finally, this model specifies the indirect contributions of the distal measure of reading comprehension, which contributes to writing indirectly through writing convention skills. This model extends previous cognitive models by specifying the role of writing convention skills in the writing process. In doing so, it offers greater specificity around the skills which underpin text generation.

10.7.2 Writing Convention Skills: Macro-structure

The findings of this thesis point towards the possibility that children with greater EGPS-WC knowledge as reflected in the test, may be less constrained in the process of translating text, enabling them to focus on vocabulary and idea coordination, thus resulting in greater proficiency at the macro-structural level. These findings highlight the importance of automaticity of writing convention skills in a similar way that is required of the proximal skills within transcription. Automaticity with writing conventions skills has been described as freeing writers with 'a repertoire of possibilities' (Myhill, Lines, & Watson, 2012) which contribute to the translation process through syntactic generation (Bourdin & Fayol, 1994; Hayes, 2000). Greater development of grammar and punctuation skills may result in the freeing up of cognitive resources to focus on macro-structural features which best meet the audience and purpose of writing, thus contributing to overall writing quality.

10.7.3 Writing Convention Skills: Micro-structure

These data also indicate that EGPS-WC offer more than simply a wider store of explicit grammar choices to be accessed by working memory. Rather, children with greater grammar and punctuation skills produce more accurate narrative writing. Two possible explanations are considered. Firstly, this could be due to an increasing awareness of the correctness of standard English required of competent writing. In attaining such awareness, children become more fluent in their writing. Alternatively, this may also be influenced by children's developmental stages in relation to the revision of writing. Younger writers are less likely to revise their work (McCutchen, 2006), meaning that the accuracy of single draft products as produced in this study are more likely to be impacted by the children's grammatical fluency and automaticity. Therefore, the role of the distal factors of grammar and punctuation should be considered alongside another potential distal factor, revision skills, to better understand the individual contributions of the writing convention skills.

What is not clear is why grammar and punctuation skills do not similarly predict accuracy in expository writing as they do in narrative writing. It is possible that this differentiation in genre is linked to children's age and maturity in writing (Berman & Verhoeven, 2002). It is argued that expository writing is more cognitively demanding, leading to shorter and more error prone texts (Scott & Windsor, 2000). It is possible that expository texts may require a greater body of writing convention skills than is assessed by the EGPS Test at age 11 in order to produce the high levels of accuracy required for the greater sophistication

of sentence structure (Reilly, Zamora, & McGivern, 2005). Alternatively, it may not be that EGPS-WC is unable to tap expository-specific content, but rather there may be an additional factor within narrative writing that may explain the difference. Narrative writing has been associated with academic performance (Boudreau, 2008). Without a comparative measure that explores children's wider cognitive ability (e.g. as seen in Mackie, 2012), it cannot be ruled out that children who are successful in narrative would simply be better at taking tests. While this was beyond the scope of this study, the role of non-verbal ability should be considered to better understand the relationship between EGPS-WC and expository writing.

While this study cannot make firm conclusion regarding the interactions between EGPS-WC and writing in both genres, to our knowledge this is the first study to identify the impact of grammar and punctuation upon micro-structural factors of writing and specifically its impact upon narrative writing quality. The results of this study demonstrate that the impact of EGPS-WC can be different when consideration is taken regarding the micro-structural and macro-structural characteristics of texts, extending the findings of Arfé et al. (2016). For children in Year 6, EGPS-WC constrained writing quality at the macro-structural level more than spelling. Inversely, spelling constrained writing accuracy at the micro-structural level significantly more than EGPS-WC. We believe that this differential effect of grammar and punctuation skills for children in the middle years has not previously been reported at this age or point in development.

10.8 Implications for Professional Practice

This study has highlighted multiple benefits of the EGPS Test. Firstly, the EGPS Test may be a reliable and valid way of assessing writing quality in Key Stage 2 at the end of primary school, offering convergent and criterion validity. This directly contradicts suggestions which cast doubt on the value of the EGPS within the suite of statutory assessments administered in English primary schools (Education Select Committee, 2016; Mansell, 2017; Rosen, 2015)

Secondly, these data also confirm that struggling writers have significant difficulties at word-, sentence-, and text-levels. The EGPS Test may have merit as an early warning for children likely to be struggling writers from which appropriate intervention can be provided. Existing literature on the instructional practices involving grammatical skills have not yet been found to improve writing competencies (Graham, McKeown, et al., 2012), but this study draws

a distinction on the importance of grammar and punctuation skills and this should be a renewed focus for intervention practices.

One of the by-products of this research is the identification of the convergent validity of EGPS Paper 2 – Spelling, and the HAST-2 standardised spelling assessment. This is of practical significance for teachers as the test offers high levels of validity, with an administration time of less than half of the HAST-2. This means that the EGPS Paper 2 can offer a valid measure of spelling attainment in a reasonably short, whole-class assessment.

To date, we believe that this is the most robust study of the EGPS Test. This thesis has identified the potential benefits of this test of writing convention skills. In a climate of uncertainty regarding the assessment in schools recovering from a global pandemic, this study identifies that the EGPS is a purposeful assessment for teachers, practitioners, and researchers. As such, these data could have significant implications on future policy decisions regarding the Key Stage 2 English Grammar, Punctuation and Spelling Test. Should debate arise regarding writing convention skills and the future worth of the EGPS Tests, this thesis will offer empirical evidence regarding the convergent validity between the EGPS Test and writing outcomes and the diagnostic ability of the EGPS Test, in addition to the increased understanding of the role of grammar and punctuation in relation to writing quality and narrative accuracy.

With all of these findings in mind, there are potentially significant implications at multiple levels within schools and teacher training institutions. Should headteachers be required to make decisions regarding the purposefulness of the EGPS Test if it does become non-statutory, as suggested by the Education Select Committee (2016), this thesis will provide greater insight to the benefits that the EGPS Test can offer. Schools and teacher training institutions should consider the importance of writing convention knowledge in their training and continuing professional development. If writing conventions are an important component of the writing process, then teachers must have a strong command of grammar, punctuation, and spelling. However, this training must be considered in light of the raised concerns and fears that teachers have raised over grammar subject matter (Safford, 2016).10.9 Limitations of the Empirical Research

This thesis was written within the context of COVID-19. The limitations placed on schools and researchers meant that methodological choices within this study were limited based on health and safety measures. Dockrell's (2019) study on proximal and distal factors of

writing identified the importance of a range of distal factors which contribute to writing. Oral language, non-verbal ability, working memory, as well as grammatical understanding are all measures that may have contributed to a more robust model of explaining the contribution of the skills tapped by the EGPS Test. However, administration of these measures requires one-to-one proximity; in the context of COVID-19, this was not possible. The cognitive models and frameworks also demonstrate the contribution of revising skills in the writing process. This was not considered in the design of this research and should be considered through further research.

Additionally, administration of the tasks within this study had to be carried out by class teachers who had access to the 'bubbles' in which the children were being educated, resulting in potential conflicts with administration. This was most apparent with the DASH Copy task, which is a short, timed exercise. A very small number of children achieved an unrealistically high score on this measure, which indicates potential issues with accuracy of the timing of the task. This specific task was excluded from the study and no other measures demonstrated inconsistencies. Even though teacher-administrators were given training and scripts, it is important to acknowledge the risk of potential administration errors, given the context.

We cannot be confident that the attainment in the EGPS test for the children in this sample is representative of the wider population. As no data are available for the age/point of time when the EGPS Test was administered, a reliable comparison cannot be made. Compared with the national data from 2018 (when children would be 7 months older), the sample in this study are potentially slightly weaker than the national average. The sample within this study also contained significantly higher participants with English as an Additional Language (EAL) compared to the national average. Therefore, it is difficult to know if the effects of EAL on the data would be generalisable to the wider population. Furthermore, while one school noted that the majority of participants within their setting identified as EAL, the reality was that most of their children were *Fluent* on EAL proficiency measures (Bell Foundation, 2021). This study could have considered accounting for EAL proficiency measures to provide a more accurate picture about the effect of EAL on EGPS and writing convention skills. Similarly, more information about the specific profiles of children with SEN would have been useful in identifying difficulties experienced in the writing process or whether the difficulties experienced were literacy based.

Finally, it should also be considered that the relationship between accuracy factor scores as assessed by CBM accuracy measures may have accentuated the results. Where children produce an incorrect spelling, they are penalised accordingly. Where spelling errors are replicated, participants are penalised multiple times, or where only spelling errors occur, accuracy measures are still affected, likely accentuating the influence of spelling. Further research should explore the micro-structural factors of writing using CBM measure which exclude spelling as criterion for accuracy in line with Dockrell (2019), which excluded spelling from CBM measures as to distinguish the contribution of transcription and text generation skills.

This thesis made use of cross-sectional design. While this method is relatively quick and inexpensive, there are limitations around causality. Where regression analyses have been used in this thesis, it should be noted that these are only exploratory and cannot offer causality.

10.10 Suggestions for the Further Research

A range of potential factors in the writing process have not been considered in relation to grammar and punctuation. Oral language, non-verbal ability, grammatical understanding, and revising skills should be explored in conjunction with EGPS-WC.

Within this study, there is also a developing theme of the relationship between EGPS-WC and cognitive resources; future research could be targeted to explore this in greater detail. Owing to the limitations of this study, the role of working memory has not been accounted for. However, the role of EGPS-WC as a distal factor in writing should be further explored to understand how automatisation of writing convention skills impacts upon the multiple dimensions of writing when controlling for working memory.

It would also be interesting to better understand the part that EGPS-WC has to play in narrative and expository when controlling for non-verbal ability and genre specific knowledge as seen in recent developments around genre awareness (Driscoll, Paszek, Gorzelsky, Hayes, & Jones, 2020). Understanding whether grammar and punctuation skills impact upon children's ability to revise single-draft texts in both genres is also of practical importance for educators in primary education.

CHAPTER 11 : Conclusion

The link between grammar and punctuation and writing has long been neglected in empirical research and remains underdeveloped. This thesis explored a statutory element of the English primary school curriculum - the English Grammar, Punctuation and Spelling Test - and its relationship to writing outcomes.

A cross-sectional design was used to explore the relationship between the EGPS 2018 Test and measures of quality, productivity, accuracy, and complexity of writing. Macro-structure was explored through word-, sentence-, and text-levels, as well as through a measure of composite writing quality. Measures of micro-structure were also further explored through narrative and expository writing. A statistically significant relationship was established between the EGPS Test and composite writing quality. This relationship demonstrates convergent and criterion validity between the EGPS Test and children's writing. To further understand the contributions of grammar and punctuation to children's writing, a principal component analysis established that each of these skills loaded upon a single factor: writing convention skills. Using factor scores of EGPS writing convention skills, a number of regression analyses were run, and EGPS-WC was found to be a predictor of writing quality at sentence- and text-levels, and on the composite writing task. The diagnostic ability of the EGPS Test was explored through ROC analysis to better understand if the statutory assessment can identify struggling writers. The EGPS Test offered excellent sensitivity and specificity.

This comprehensive study's subsequent confirmation of a statistically significant relationship between the EGPS Test and children's writing outcomes contradicts the most recent recommendations given to Government regarding the longevity of the EGPS Test. Where previously unsubstantiated advice indicates that the EGPS Test cannot be used as a proxy for writing assessment, this study demonstrates that there is a statistically significant relationship between the EGPS Test and children's overall writing quality and narrative accuracy. While the current body of literature suggests that this application of grammatical knowledge has been ineffective (Wyse, 2006), these findings provide novel insight into the relationship between knowledge of writing convention skills and writing products. The findings of this study consolidate and extend those of Canniford (2019), indicating that there is an apparent transferability of writing convention skills into writing products; children who have greater knowledge of writing convention skills as demonstrated by the EGPS Tests are

more likely to produce increasingly competent writing outcomes. Their overall writing quality at sentence- and text-level is better, with greatest gains seen at text-level. As far as we are aware, this is the first empirical study that is able to demonstrate that greater success on the EGPS Test demonstrates an increased likelihood of greater writing quality at the sentence- and paragraph-levels, as well as greater writing accuracy in the narrative genre. However, these findings lay in partial contradiction with Myhill (2016), who in her evidence for the Education Select Committee (2016), stated that the EGPS Test 'does not provide any information about accuracy in writing'.

This study does not explore how to teach grammar, punctuation, or spelling. What it does tell you is that whichever way you teach it, the knowledge gained has a direct relationship with children's writing products. Elbow (1981) argues that 'nothing helps your writing so much as ignoring grammar'. Yet, this thesis argues that *nothing helps your writing so much as knowing writing conventions explicitly so that you can ignore them.* By doing so, you are potentially reducing the parallel burden upon working memory, enabling you to draw upon a 'repertoire of infinite possibilities' (Myhill et al., 2012).

REFERENCES

References

Abbott, R. D., & Berninger, V. W. (1993). Structural equation modeling of relationships among developmental skills and writing skills in primary-and intermediate-grade writers. *Journal of educational psychology, 85*(3), 478.

Abbott, R. D., Berninger, V. W., & Fayol, M. (2010). Longitudinal relationships of levels of language in writing and between writing and reading in grades 1 to 7. *Journal of educational psychology, 102*(2), 281.

Ahmed, Y., Wagner, R. K., & Lopez, D. (2014). Developmental relations between reading and writing at the word, sentence, and text levels: A latent change score analysis. *Journal of educational psychology, 106*(2), 419.

Alderson, P., & Morrow, V. (2011). *The ethics of research with children and young people: A practical handbook*: SAGE publications Ltd.

Alves, R. (2012). The future is bright for writing research. *Past, present, and future contributions of cognitive writing research to cognitive psychology*.

Alves, R. A., & Limpo, T. (2015). Progress in written language bursts, pauses, transcription, and written composition across schooling. *Scientific studies of reading, 19*(5), 374-391.

Andrews, L. (2013). *Language exploration and awareness: A resource book for teachers*: Routledge.

Andrews, R., Torgerson, C., Beverton, S., Freeman, A., Locke, T., Low, G., . . . Zhu, D. (2004). The effect of grammar teaching (sentence combining) in English on 5 to 16 year olds' accuracy and quality in written composition: Review summary. *University of York, UK*.

Andrews, R., Torgerson, C., Beverton, S., Freeman, A., Locke, T., Low, G., . . . Zhu, D. (2006). The effect of grammar teaching on writing development. *British Educational Research Journal, 32*(1), 39-55.

Andrews, R., Torgerson, C., Beverton, S., Locke, T., Low, G., Robinson, A., & Zhu, D. (2004). The effect of grammar teaching (syntax) in English on 5 to 16 year olds' accuracy and quality in written composition. *Research evidence in education library*.

Apel, K., & Apel, L. (2011). Identifying intraindividual differences in students' written language abilities. *Topics in Language Disorders, 31*(1), 54-72.

Appelman, A., & Bolls, P. (2011). Article recall, credibility lower with grammar errors. *Newspaper Research Journal, 32*(2), 50-62.

Arfé, B., Dockrell, J. E., & De Bernardi, B. (2016). The effect of language specific factors on early written composition: The role of spelling, oral language and text generation skills in a shallow orthography. *Reading and Writing, 29*(3), 501-527.

Arfé, B., & Pizzocaro, E. (2016). Sentence generation in children with and without problems of written expression. In *Written and Spoken Language Development across the Lifespan* (pp. 327-344): Springer.

Assessment, G. (2010). New Group Reading Test (NGRT). In *London, UK: GL Assessment*.

Association, A. E. R. (1999). *Standards for educational and psychological testing*: Amer Educational Research Assn.

Babayiğit, S., & Stainthorp, R. (2010). Component processes of early reading, spelling, and narrative writing skills in Turkish: A longitudinal study. *Reading and Writing, 23*(5), 539-568.

Baddeley, A. D., & Hitch, G. J. (2000). Development of working memory: Should the Pascual-Leone and the Baddeley and Hitch models be merged? *Journal of experimental child psychology, 77*(2), 128-137.

Banerjee, J., Yan, X., Chapman, M., & Elliott, H. (2015). Keeping up with the times: Revising and refreshing a rating scale. *Assessing Writing, 26*, 5-19.

Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of educational research, 74*(1), 29-58.

Bara, F., & Gentaz, É. (2010). Apprendre à tracer les lettres: une revue critique. *Psychologie française, 55*(2), 129-144.

Barnett, A., Henderson, S., Scheib, B., & Schulz, J. (2007). The Detailed Assessment of Speed of Handwriting (DASH). Manual.

Bazerman, C. (2016). What do sociocultural studies of writing tell us about learning to write. *Handbook of writing research, 2*, 11-23.

Bazerman, C. (2018). What Does a Model Model? And for Whom? *Educational Psychologist, 53*(4), 301-318.

Beard, R. (2000). *Developing Writing 3-13*: Hodder and Stoughton.

Becker, A. (2006). A review of writing model research based on cognitive processes. *Revision: History, theory, and practice*, 25-49.

Beers, S. F., & Nagy, W. E. (2011). Writing development in four genres from grades three to seven: Syntactic complexity and genre differentiation. *Reading and Writing, 24*(2), 183-202.

Benton, S. L., & Kiewra, K. A. (1986). Measuring the organizational aspects of writing ability. *Journal of Educational Measurement, 23*(4), 377-386.

Bereiter, C., & Scardamalia, M. (2013). *The psychology of written composition*: Routledge.

Berman, R., & Katzenberger, I. (2004). Form and function in introducing narrative and expository texts: A developmental perspective. *Discourse Processes, 38*(1), 57-94.

Berman, R., & Nir-Sagiv, B. (2007). Comparing Narrative and Expository Text Construction Across Adolescence: A Developmental Paradox. *Discourse Processes, 43*(2), 79-120. doi:10.1080/01638530709336894

Berman, R., & Verhoeven, L. (2002). Cross-linguistic perspectives on the development of text-production abilities: Speech and writing. *Written Language & Literacy, 5*(1), 1-43.

Berninger, V. W. (1999). Coordinating transcription and text generation in working memory during composing: Automatic and constructive processes. *Learning Disability Quarterly, 22*(2), 99-112.

Berninger, V. W. (2000). Development of language by hand and its connections with language by ear, mouth, and eye. *Topics in Language Disorders*.

Berninger, V. W. (2019). *Reading and writing acquisition: A developmental neuropsychological perspective*: Routledge.

Berninger, V. W., & Abbott, R. D. (2010). Listening comprehension, oral expression, reading comprehension, and written expression: Related yet unique language systems in grades 1, 3, 5, and 7. *Journal of educational psychology, 102*(3), 635.

Berninger, V. W., Abbott, R. D., Abbott, S. P., Graham, S., & Richards, T. (2002). Writing and reading: Connections between language by hand and language by eye. *Journal of learning disabilities, 35*(1), 39-56.

Berninger, V. W., Abbott, R. D., & Alsdorf, B. J. (1997). Lexical-and sentence-level processes in comprehension of written sentences. *Reading and Writing, 9*(2), 135-162.

Berninger, V. W., Abbott, R. D., Jones, J., Wolf, B. J., Gould, L., Anderson-Youngstrom, M., . . . Apel, K. (2006). Early development of language by hand: Composing, reading, listening, and speaking connections; three letter-writing modes; and fast mapping in spelling. *Developmental neuropsychology, 29*(1), 61-92.

Berninger, V. W., Abbott, R. D., Nagy, W., & Carlisle, J. (2010). Growth in phonological, orthographic, and morphological awareness in grades 1 to 6. *Journal of psycholinguistic research, 39*(2), 141-163.

Berninger, V. W., Abbott, R. D., Thomson, J. B., & Raskind, W. H. (2001). Language phenotype for reading and writing disability: A family approach. *Scientific studies of reading, 5*(1), 59-106.

Berninger, V. W., Cartwright, A. C., Yates, C. M., Swanson, H. L., & Abbott, R. D. (1994). Developmental skills related to writing and reading acquisition in the intermediate grades. *Reading and Writing, 6*(2), 161-196.

Berninger, V. W., Fuller, F., & Whitaker, D. (1996). A process model of writing development across the life span. *Educational Psychology Review, 8*(3), 193-218. doi:10.1007/bf01464073

Berninger, V. W., Mizokawa, D. T., & Bragg, R. (1991). Scientific practitioner: Theory-based diagnosis and remediation of writing disabilities. *Journal of School Psychology, 29*(1), 57-79.

Berninger, V. W., Nagy, W., & Beers, S. (2011). Child writers' construction and reconstruction of single sentences and construction of multi-sentence texts: Contributions of syntax and transcription to translation. *Reading and Writing, 24*(2), 151-182.

Berninger, V. W., Nielsen, K. H., Abbott, R. D., Wijsman, E., & Raskind, W. (2008). Writing problems in developmental dyslexia: Under-recognized and under-treated. *Journal of School Psychology, 46*(1), 1-21.

Berninger, V. W., & Swanson, H. (1994). Modifying Hayes and Flowers' model of skilled writing to explain developing writing. *Advances in cognition and educational practice. Children's writing: Toward a process theory of the development of skilled writing, 2*, 1-30.

Berninger, V. W., Vaughan, K., Abbott, R. D., Begay, K., Coleman, K. B., Curtin, G., . . . Graham, S. (2002). Teaching spelling and composition alone and together: Implications for the simple view of writing. *Journal of educational psychology, 94*(2), 291.

Berninger, V. W., Vaughan, K. B., Abbott, R. D., Abbott, S. P., Rogan, L. W., Brooks, A., . . . Graham, S. (1997). Treatment of handwriting problems in beginning writers: Transfer from handwriting to composition. *Journal of educational psychology, 89*(4), 652.

Berninger, V. W., & Whitaker, D. (1993). Theory-based branching diagnosis of writing disabilities. *School Psychology Review, 22*(4), 623-642.

Berninger, V. W., Winn, W., MacArthur, C., Graham, S., & Fitzgerald, J. (2006). Implications of advancements in brain research and technology for writing development, writing instruction, and educational evolution. *Handbook of writing research*, 96-114.

Berninger, V. W., Winn, W. D., Stock, P., Abbott, R. D., Eschen, K., Lin, S.-J. C., . . . Lovitt, D. (2008). Tier 3 specialized writing instruction for students with dyslexia. *Reading and Writing, 21*(1), 95-129.

Berninger, V. W., Yates, C., Cartwright, A., Rutberg, J., Remy, E., & Abbott, R. (1992). Lower-level developmental skills in beginning writing. *Reading and Writing, 4*(3), 257-280.

Bernstein, B. (1971). Class, codes and control (Vol. 1). In: London: Routledge & Kegan Paul.

Bew, P. (2011). *Independent review of key stage 2 testing, assessment and accountability : final report*.

Biggam, J. (2015). *Succeeding with your master's dissertation: a step-by-step handbook*: McGraw-Hill Education (UK).

Blank, M. (2001). Classroom discourse: A key to literacy. In *Speaking, reading, and writing in children with language learning disabilities* (pp. 163-186): Psychology Press.

Boudreau, D. (2008). Narrative abilities: Advances in research and implications for clinical practice. *Topics in Language Disorders, 28*(2), 99-114.

Bourdin, B., & Fayol, M. (1994). Is written language production more difficult than oral language production? A working memory approach. *International journal of psychology, 29*(5), 591-620.

Bourke, L., & Adams, A.-M. (2003). The relationship between working memory and early writing assessed at the word, sentence and text level. *Educational and Child Psychology, 20*(3), 19-36.

Bouwer, R., Béguin, A., Sanders, T., & Van den Bergh, H. (2015). Effect of genre on the generalizability of writing scores. *Language Testing, 32*(1), 83-100.

Braddock, R., Lloyd-Jones, R., & Schoer, L. (1963). *Research in written composition*: JSTOR.

Breakwell, G. M., Hammond, S. E., Fife-Schaw, C. E., & Smith, J. A. (2006). *Research methods in psychology*: Sage Publications, Inc.

Brien, J. (2012). *Teaching Primary English*: Sage.

Britton, B. K. (1994). Understanding expository text: Building mental structures to induce insights.

Brown, J. (1996). 1996: Testing in language programs. Upper Saddle River, NJ: Prentice Hall Regents.

Bryant, P., & Nunes, T. (2006). Morphemes and literacy: a starting point. In *Improving literacy by teaching morphemes* (pp. 21-52): Routledge.

Bryman, A. (2016). *Social research methods*: Oxford university press.

Cameron, D. (2012). *Verbal hygiene*: Routledge.

Canniford, J. T. (2019). *The EGPS Test: Purposeful and valid? An exploration of the purpose and validity of the KS2 EGPS Test.* (Unpublished Insitute Focused Study), University College London, London.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment* (Vol. 17): Sage publications.

Carroll, L. A. (2002). *Rehearsing new roles: How college students develop as writers*: SIU press.

Carter, R. (1990). *Knowledge about language and the curriculum: The LINC reader*: Hodder & Stoughton.

Chalker, S., & Weiner, E. S. (1998). *OXFORD DIC ENGLISH GRAMMAR (P)*: Oxford University Press.

Christensen, C. A. (2004). Relationship between orthographic-motor integration and computer use for the production of creative and well-structured written text. *British Journal of Educational Psychology, 74*(4), 551-564.

Cohen, A. (1994). Assessing language ability in the classroom.

Cohen, L., & Holliday, M. (1996). *Practical statistics for students: An introductory text*: Sage.

Cohen, L., Manion, L., & Morrison, K. (2013). *Research Methods in Education*. Hoboken: Taylor and Francis.

Coirier, P. (1996). Composing argumentative texts: Cognitive and/or textual complexity. *Theories, models and methodology in writing research*, 317-338.

Coker Jr, D. L., & Ritchey, K. D. (2010). Curriculum-based measurement of writing in kindergarten and first grade: An investigation of production and qualitative scores. *Exceptional children, 76*(2), 175-193.

Cole, J., Haley, K., & Muenz, T. (1997). Written expression reviewed. *Research in the Schools, 4*(1), 17-34.

Colman, A. M. (2015). A Dictionary of Psychology. In (4 ed.): Oxford University Press.

Connelly, V., Dockrell, J. E., Walter, K., & Critten, S. (2012). Predicting the Quality of Composition and Written Language Bursts From Oral Language, Spelling, and Handwriting Skills in Children With and Without Specific Language Impairment. *Written Communication, 29*(3), 278-302. doi:doi:10.1177/0741088312451109

Coolican, H. (2017). *Research methods and statistics in psychology*: Psychology Press.

Cragg, L., & Nation, K. (2006). Exploring written narrative in children with poor reading comprehension. *Educational Psychology, 26*(1), 55-72.

Cramer, D. (1998). *Fundamental statistics for social research: step-by-step calculations and computer techniques using SPSS for Windows*: Psychology Press.

Critten, S., Sheriston, L., & Mann, F. (2016). Young children's spelling representations and spelling strategies. *Learning and Instruction, 46*, 34-44.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin, 52*(4), 281.

Crotty, M. (1998). *The Foundations of Social Research: Meaning and Perspective in the Research Process*: SAGE Publications.

Crystal, D. (2012). *Spell it out: the singular story of English spelling*: Profile.

Crystal, D., & McLachlan, E. (2004). *Rediscover grammar*: Longman Harlow.

Culham, R. (2003). *6+ 1 traits of writing: The complete guide grades 3 and up*: Scholastic Inc.

Cullen, K. (1999). *Chambers guide to punctuation*: Chambers.

Daffern, T., Mackenzie, N. M., & Hemmings, B. (2017). Predictors of writing success: How important are spelling, grammar and punctuation? *Australian Journal of Education, 61*(1), 75-87. doi:10.1177/0004944116685319

De Lemos, M. (2002). Closing the gap between research and practice: Foundations for the acquisition of literacy.

Deane, P., & Song, Y. (2014). Psicología Educativa. *Psicología Educativa, 20*(2), 99.

Deno, S. L., Marston, D., & Mirkin, P. (1982). Valid measurement procedures for continuous evaluation of written expression. In: SAGE Publications Sage CA: Los Angeles, CA.

Department for Education, D. (2012). *What is the research evidence on writing?*

Department for Education, D. (2019a). *National curriculum assessments at key stage 2 in England, 2019 (provisional)* Retrieved from National Statistics: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/830285/KS2_Provisional_publication_text_2019.pdf

Department for Education, D. (2019b). *National curriculum assessments at key stage 2 in England, 2019 (revised).* Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/851798/KS2_Revised_publication_text_2019_v3.pdf

Derewianka, B. (2003). Trends and issues in genre-based approaches. *RELC journal, 34*(2), 133-154.

Derewianka, B., & Christie, F. (2008). School Discourse: Learning to write across the years of schooling.

DfE. (2014). *The national curriculum in England: Complete framework for key stages 1 to 4.*

Diederich, P. B. (1974). Measuring growth in English.

Doane, D. P., & Seward, L. E. (2011). Measuring skewness: a forgotten statistic? *Journal of statistics education, 19*(2).

Dockrell, J. E., & Arfè, B. (2020). What does the Research Say About Research/Evidence-based (Core) Writing Instruction in General Education Classrooms? In M. W. Dunn (Ed.), *Writing instruction and intervention for struggling writers: Multi-Tiered Systems of Support*. Newcastle upon Tyne, UK: Cambridge Scholars.

Dockrell, J. E., & Connelly, V. (2015). The role of oral language in -underpinning the text generation -difficulties in children with specific -language impairment. *Journal of Research in Reading, 38*(1), 18-34. doi:10.1111/j.1467-9817.2012.01550.x

Dockrell, J. E., & Connelly, V. (2021). Capturing the Challenges in Assessing Writing: Development and Writing Dimensions. In T. O. Limpo, T. (Ed.), *Executive Functions and Writing*. Oxford: OUP.

Dockrell, J. E., Connelly, V., & Arfè, B. (2019). Struggling writers in elementary school: Capturing drivers of performance. *Learning and Instruction, 60*, 75-84.

Dockrell, J. E., Connelly, V., Walter, K., & Critten, S. (2015). Assessing children's writing products: the role of curriculum based measures. *British Educational Research Journal, 41*(4), 575-595.

Dockrell, J. E., Connelly, V., Walter, K., & Critten, S. (2017). The role of Curriculum Based Measures in assessing writing products. In *Writing development in struggling learners* (pp. 182-197): Brill.

Doyle, L., & Godfrey, R. (2005). Investigating the reliability of the Key Stage 2 test results for assessing individual pupil achievement and progress in England. *London Review of Education, 3*(1), 29-45.

Drake, P., & Heath, L. (2010). *Practitioner research at doctoral level: Developing coherent research methodologies*: Routledge.

Driscoll, D. L., Paszek, J., Gorzelsky, G., Hayes, C. L., & Jones, E. (2020). Genre Knowledge and Writing Development: Results From the Writing Transfer Project. *Written Communication, 37*(1), 69-103. doi:10.1177/0741088319882313

Duff, F. J., Mengoni, S. E., Bailey, A. M., & Snowling, M. J. (2015). Validity and sensitivity of the phonics screening check: implications for practice. *Journal of Research in Reading, 38*(2), 109-123.

Dunsmuir, S., Kyriacou, M., Batuwitage, S., Hinson, E., Ingram, V., & O'Sullivan, S. (2015). An evaluation of the Writing Assessment Measure (WAM) for children's narrative writing. *Assessing Writing, 23*, 1-18.

Education, B. o. (1921). *The Teaching of English in England (Newbolt Report)*. Retrieved from

Ehri, L. C. (1987). Learning to read and spell words. *Journal of Reading behavior, 19*(1), 5-31.

Elbow, P. (2002). Vernacular Englishes in the writing classroom? Probing the culture of literacy. *ALT DIS: Alternative discourses and the academy*, 126-138.

Espin, C., Shin, J., Deno, S. L., Skare, S., Robinson, S., & Benner, B. (2000). Identifying indicators of written expression proficiency for middle school students. *The journal of special education, 34*(3), 140-153.

Espin, C., Weissenburger, J., & Benson, B. (2004). Assessing the writing performance of students in special education. *Exceptionality, 12*(1), 55-66.

Fang, Z., & Wang, Z. (2011). Beyond rubrics: Using functional language analysis to evaluate student writing. *Australian Journal of Language and Literacy, The, 34*(2), 147.

Fathman, A. K., & Whalley, E. (1990). Teacher response to student writing: focus on form versus content. In (pp. 178-190): Cambridge University Press.

Fayol, M. (1997). On acquiring and using punctuation: A study of written French. *Processing interclausal relationships. Studies in the production and comprehension of text*, 157-178.

Fayol, M., Hupet, M., & Largy, P. (1999). The acquisition of subject-verb agreement in written French: From novices to experts' errors. *Reading and Writing, 11*(2), 153-174.

Ferreiro, E., & Pontecorvo, C. (1999). Managing the written text: The beginning of punctuation in children's writing. *Learning and Instruction, 9*(6), 543-564.

Fey, M. E., Catts, H. W., Proctor-Williams, K., Tomblin, J. B., & Zhang, X. (2004). Oral and written story composition skills of children with language impairment.

Field, A. (2009). *Discovering statistics using SPSS*: Sage publications.

Fitzgerald, J., & Shanahan, T. (2000). Reading and writing relations and their development. *Educational Psychologist, 35*(1), 39-50.

Fletcher, C. R., & Chrysler, S. T. (1990). Surface forms, textbases, and situation models: Recognition memory for three types of textual information. *Discourse Processes, 13*(2), 175-190.

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College composition and communication, 32*(4), 365-387.

Foundation, B. (2021). Levels of Proficiency in English. Retrieved from https://www.bell-foundation.org.uk/eal-programme/eal-assessment-framework/levels-of-proficiency-in-english/

Frey, B. (2018). The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation. doi:10.4135/9781506326139

Fuchs, L. S. (2004). The Past, Present, and Future of Curriculum-Based Measurement Research. *School Psychology Review, 33*(2), 188-192. doi:10.1080/02796015.2004.12086241

Gansle, K. A., VanDerHeyden, A. M., Noell, G. H., Resetar, J. L., & Williams, K. L. (2006). The technical adequacy of curriculum-based and rating-based measures of written expression for elementary school students. *School Psychology Review, 35*(3), 435-450.

García, E., Crespo, P., & Bermúdez, I. (2017). Writing an independently composed sentence by Spanish-speaking children with and without poor transcription skills: A writing-level match design. *Journal of learning disabilities, 50*(5), 511-521.

Gillespie, A., & Graham, S. (2014). A meta-analysis of writing interventions for students with learning disabilities. *Exceptional children, 80*(4), 454-473.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological review, 101*(3), 371.

Graham, S. (2018). A revised writer (s)-within-community model of writing. *Educational Psychologist, 53*(4), 258-279.

Graham, S., Berninger, V., Weintraub, N., & Schafer, W. (1998). Development of handwriting speed and legibility in grades 1–9. *The Journal of Educational Research, 92*(1), 42-52.

Graham, S., Berninger, V. W., Abbott, R. D., Abbott, S. P., & Whitaker, D. (1997). Role of mechanics in composing of elementary school students: A new methodological approach. *Journal of educational psychology, 89*(1), 170.

Graham, S., Bollinger, A., Olson, C. B., D'Aoust, C., MacArthur, C., McCutchen, D., & Olinghouse, N. (2012). Teaching elementary school students to be effective writers.

Graham, S., Harris, K., & Hebert, M. (2011). Informing writing: The benefits of formative assessment. A Carnegie Corporation Time to Act report. *L1 RESEARCH ARCHIVES ONLINE*.

Graham, S., & Harris, K. R. (2005). Improving the writing performance of young struggling writers: Theoretical and programmatic research from the center on accelerating student learning. *The journal of special education, 39*(1), 19-33.

Graham, S., Hebert, M., Fishman, E., Ray, A. B., & Rouse, A. G. (2020). Do children classified with specific language impairment have a learning disability in writing? A meta-analysis. *Journal of learning disabilities, 53*(4), 292-310.

Graham, S., Hebert, M., & Harris, K. R. (2011). Throw'em out or make'em better? State and district high-stakes writing assessments. *Focus on Exceptional Children, 44*(1).

Graham, S., Liu, X., Bartlett, B., Ng, C., Harris, K. R., Aitken, A., . . . Talukdar, J. (2018). Reading for writing: A meta-analysis of the impact of reading interventions on writing. *Review of educational research, 88*(2), 243-284.

Graham, S., McKeown, D., Kiuhara, S., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of educational psychology, 104*(4), 879.

Graham, S., Morphy, P., Harris, K. R., Fink-Chorzempa, B., Saddler, B., Moran, S., & Mason, L. (2008). Teaching spelling in the primary grades: A national survey of instructional practices and adaptations. *American Educational Research Journal, 45*(3), 796-825.

Graham, S., & Weintraub, N. (1996). A review of handwriting research: Progress and prospects from 1980 to 1994. *Educational Psychology Review, 8*(1), 7-87.

Gray, D. E. (2014). *Doing research in the real world*. London: SAGE Publications Ltd.

Green, S. B., & Salkind, N. J. (2012). *Using SPSS for Windows and Macintosh: Analyzing and understanding data*: Prentice-Hall.

Grimshaw, A. D. (2003). Genres, registers, and contexts of discourse. In A. C. Graesser, Gernsbacher, M. A. and Goldman, S. R. (Ed.), *Handbook of discourse processes*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Guba, E. (1990). Carrying on the dialog. In. EG Guna (Ed.), The paradigm dialog (pp. 368–378). In: Newbury Park, CA. Sage.

Hanley, J. R. (1997). Reading and spelling impairments in undergraduate students with developmental dyslexia. *Journal of Research in Reading*.

Harmey, S. J. (2021). Co-constructing writing: handing over the reins to young authors. *Education 3-13, 49*(4), 412-421.

Hartwell, P. (1985). Grammar, Grammars, and the Teaching of Grammar. . *College English, 47*(2), 105-127.

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. *The science of writing: theories, methods, individual differences, and applications*, 1-27.

Hayes, J. R. (2000). A New Framework for Understanding Cognition and Perspectives on writing: Research, theory, and practice. *6*.

Hayes, J. R. (2012). Evidence from language bursts, revision, and transcription for translation and its relation to other writing processes. *Translation of thought to written text while composing*, 15-25.

Hayes, J. R., & Berninger, V. (2014). Cognitive processes in writing: A framework. *Writing development in children with hearing loss, dyslexia, or oral language problems: Implications for assessment and instruction*, 3-15.

Hebert, M., Simpson, A., & Graham, S. (2013). Comparing effects of different writing activities on reading comprehension: A meta-analysis. *Reading and Writing, 26*(1), 111-138.

Herrington, A. J., & Curtis, M. (2000). *Persons in Process: Four Stories of Writing and Personal Development in College. Refiguring English Studies Series*: ERIC.

Hillocks Jr, G., & Smith, M. W. (1991). Grammar and usage. *Handbook of research on teaching the English language arts*, 591-603.

Hudson, L., & Ozanne, J. (1988). Alternative ways of seeking knowledge in consumer research. *Journal of consumer research, 14*(4), 508-521.

Hudson, R., & Walmsley, J. (2005). The English patient: English grammar and teaching in the Twentieth Century. *Journal of Linguistics, 41*(3), 593-622.

Hudson, R. F., Lane, H. B., & Mercer, C. D. (2005). Writing prompts: The role of various priming conditions on the compositional fluency of developing writers. *Reading and Writing, 18*(6), 473-495.

Hunt, K. W. (1964). Differences in grammatical structures written at three grade levels, the structures to be analyzed by transformational methods.

Hunt, K. W. (1965). Grammatical Structures Written at Three Grade Levels. NCTE Research Report No. 3.

Hunt, K. W. (1970). Syntactic maturity in schoolchildren and adults. *Monographs of the society for research in child development, 35*(1), iii-67.

Huot, B. (1990). The literature of direct writing assessment: Major concerns and prevailing trends. *Review of educational research, 60*(2), 237-263.

Jessmer, S. L., & Anderson, D. (2001). The effect of politeness and grammar on user perceptions of electronic mail. *North American Journal of Psychology, 3*(2), 331-346.

Jewell, J., & Malecki, C. K. (2005). The utility of CBM written language indices: An investigation of production-dependent, production-independent, and accurate-production scores. *School Psychology Review, 34*(1), 27-44.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of educational psychology, 80*(4), 437.

Juel, C., Griffith, P. L., & Gough, P. B. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of educational psychology, 78*(4), 243.

Katzenberger, I. (2005). The Super-Structure of Written Expository Texts—A Developmental Perspective. In *Perspectives on language and language development* (pp. 327-336): Springer.

Kellogg, R. T. (1996). A model of working memory in writing.

Kellogg, R. T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of writing research, 1*(1).

Kent, R., O'leary, D., Diament, C., & Dietz, A. (1974). Expectation biases in observational evaluation of therapeutic change. *Journal of Consulting and Clinical Psychology, 42*(6), 774.

Kent, S., & Wanzek, J. (2016). The relationship between component skills and writing quality and production across developmental levels: A meta-analysis of the last 25 years. *Review of educational research, 86*(2), 570-601.

Kim, Y.-S., Al Otaiba, S., Folsom, J. S., Greulich, L., & Puranik, C. (2014a). Evaluating the dimensionality of first-grade written composition.

Kim, Y.-S., Al Otaiba, S., Folsom, J. S., Greulich, L., & Puranik, C. (2014b). Evaluating the Dimensionality of First-Grade Written Composition. *Journal of Speech, Language and*

Hearing Research (Online), 57(1), 199-211. doi:http://dx.doi.org/10.1044/1092-4388(2013/12-0152)

Kim, Y.-S., Al Otaiba, S., Puranik, C., Folsom, J. S., Greulich, L., & Wagner, R. K. (2011). Componental skills of beginning writing: An exploratory study. *Learning and Individual Differences, 21*(5), 517-525.

Kim, Y.-S., Al Otaiba, S., Sidler, J. F., & Gruelich, L. (2013). Language, literacy, attentional behaviors, and instructional quality predictors of written composition for first graders. *Early Childhood Research Quarterly, 28*(3), 461-469.

Kim, Y.-S., Al Otaiba, S., & Wanzek, J. (2015). Kindergarten predictors of third grade writing. *Learning and Individual Differences, 37*, 27-37.

Kim, Y.-S., Al Otaiba, S., Wanzek, J., & Gatlin, B. (2015). Towards an understanding of dimensions, predictors, and gender gap in written composition. *Journal of educational psychology, 107*(1), 79. doi:10.1037/a0037210

Kim, Y.-S., Gatlin, B., Al Otaiba, S., & Wanzek, J. (2018). Theorization and an empirical investigation of the component-based and developmental text writing fluency construct. *Journal of learning disabilities, 51*(4), 320-335.

Kim, Y.-S., Park, C. H., & Park, Y. (2013). Is academic language use a separate dimension in beginning writing? Evidence from Korean. *Learning and Individual Differences, 27*, 8-15.

Kim, Y.-S., & Park, S.-H. (2019). Unpacking pathways using the direct and indirect effects model of writing (DIEW) and the contributions of higher order cognitive skills to writing. *Reading and Writing, 32*(5), 1319-1343.

Kim, Y.-S., & Schatschneider, C. (2017). Expanding the developmental models of writing: A direct and indirect effects model of developmental writing (DIEW). *Journal of educational psychology, 109*(1), 35.

Klecan-Aker, J. S., & Hedrick, D. L. (1985). A study of the syntactic language skills of normal school-age children. *Language, Speech, and Hearing Services in Schools, 16*(3), 187-198.

Kolln, M. (2006). Rhetorical grammar: grammatical choices, rhetorical effects/Martha Kolln. In: Boston: Allyn and Bacon.

Koutsoftas, A. D., & Gray, S. (2012). Comparison of narrative and expository writing in students with and without language-learning disabilities. *Language, Speech, and Hearing Services in Schools*.

Kuhn, T. S. (2012). *The structure of scientific revolutions*: University of Chicago press.

Laerd. (2015a). Independent-samples t-test using SPSS Statistics. *Statistical tutorials and software guides*. Retrieved from

Laerd. (2015b). Paired-samples t-test using SPSS Statistics. *Statistical tutorials and software guides*. Retrieved from Retrieved from https://statistics.laerd.com/ website:

Laerd. (2018). *Statistical tutorials and software guides*. In.

Lai, E. R., Wolfe, E. W., & Vickers, D. (2015). Differentiation of illusory and true halo in writing scores. *Educational and psychological measurement, 75*(1), 102-125.

Largy, P., Dédéyan, A., & Hupet, M. (2004). Orthographic revision: A developmental study of how revisers check verbal agreements in written texts. *British Journal of Educational Psychology, 74*(4), 533-550.

Lavrakas, P. J. (2008). Encyclopedia of Survey Research Methods. doi:10.4135/9781412963947

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational research methods, 11*(4), 815-852.

Lefstein, A. (2009). Rhetorical grammar and the grammar of schooling: Teaching "powerful verbs" in the English National Literacy Strategy. *Linguistics and Education, 20*(4), 378-400.

Lembke, E., Deno, S. L., & Hall, K. (2003). Identifying an indicator of growth in early writing proficiency for elementary school students. *Assessment for Effective Intervention, 28*(3-4), 23-35.

Limpo, T., & Alves, R. A. (2013). Modeling writing development: Contribution of transcription and self-regulation to Portuguese students' text generation quality. *Journal of educational psychology, 105*(2), 401.

Limpo, T., Alves, R. A., & Connelly, V. (2017). Examining the transcription-writing link: Effects of handwriting fluency and spelling accuracy on writing performance via planning and translating in middle grades. *Learning and Individual Differences, 53*, 26-36.

Looney, J. W. (2011). Integrating formative and summative assessment: progress toward a seamless system?

Love, K., & Sandiford, C. (2016). Teachers' and students' meta-reflections on writing choices: An Australian case study. *International Journal of Educational Research, 80*, 204-216.

MacArthur, C., Graham, S., & Fitzgerald, J. (2006). Handbook of research on writing. *New York: Guilford.*

Macauley, W. J. (1947). The difficulty of grammar. *British Journal of Educational Psychology, 17*(3), 153-162.

MacDonald, S. (1994). *Professional academic writing in the humanities and social sciences*: SIU Press.

Mackenzie, N., & Hemmings, B. (2014). Predictors of success with writing in the first year of school. *Issues in Educational Research, 24*(1), 41.

Mackenzie, N., Scull, J., & Bowles, T. (2015). Writing over time: An analysis of texts created by Year One students. *The Australian Educational Researcher, 42*(5), 567-593.

Mackie, C., & Dockrell, J. E. (2004). The nature of written language deficits in children with SLI. *Journal of Speech, Language, and Hearing Research.*

Mackie, C., Dockrell, J. E., & Lindsay, G. (2013). An evaluation of the written texts of children with SLI: The contributions of oral language, reading and phonological short-term memory. *Reading and Writing, 26*(6), 865-888.

Mansell, W. (2017). Battle on the adverbials front: Grammar advisers raise worries about Sats tests and teaching. *The Guardian. Available online at: www. theguardian. com/education/2017/may/09/fronted-adverbials-sats-grammar-test-primary.*

Mansell, W., & James, M. (2009). *Assessment in schools: Fit for purpose?* : ESRC Teaching and Learning Research Programme.

Martin, J. R. (1985). *'Process and Text: Two Aspects of Human Semiosis' in J.D. Benson and W.S. Greaves (eds.), Systemic Perspectives on Discourse (Norwood, NJ: Ablex), I: 248-74.* .

McCarthy, L. P. (1987). A stranger in strange lands: A college student writing across the curriculum. *Research in the Teaching of English*, 233-265.

McCutchen, D. (2000). Knowledge, processing, and working memory: Implications for a theory of writing. *Educational Psychologist, 35*(1), 13-23.

McCutchen, D. (2006). Cognitive factors in the development of children's writing. *Handbook of writing research, 8*, 115-130.

McMaster, K., Du, X., Yeo, S., Deno, S., Parker, D., & Ellis, T. (2011). Curriculum-based measures of beginning writing: Technical features of the slope. *Exceptional children, 77*(2), 185-206.

McMaster, K., & Espin, C. (2007). Technical features of curriculum-based measurement in writing: A literature review. *The journal of special education, 41*(2), 68-84.

McMaster, K., Ritchey, K., & Lembke, E. (2011). Curriculum-based measurement for beginning writers: Recent developments and future directions. *Assessment and intervention*.

McMillan, J. H. (2013). *SAGE handbook of research on classroom assessment*: Sage.

Messick, S. (1980). Test validity and the ethics of assessment. *American psychologist, 35*(11), 1012.

Miller, B., & McCardle, P. (2011). Reflections on the need for continued research on writing. *Reading and Writing, 24*(2), 121-132.

Miller, D., & Crocker, L. (1990). Validation methods for direct writing assessment. *Applied Measurement in Education, 3*(3), 285-296.

Myers, M. (1980). *A procedure for writing assessment and holistic scoring*: ERIC.

Myhill, D. (2011). The ordeal of deliberate choice: Metalinguistic development in secondary writers. *Past, present, and future contributions of cognitive writing research to cognitive psychology*, 247-274.

Myhill, D. (2016). *Written evidence submitted by the Centre for Research in Writing at the University of Exeter*. Retrieved from Online: http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/education-committee/primary-assessment/written/40810.html

Myhill, D. (2021). Grammar re-imagined: foregrounding understanding of language choice in writing. *English in Education, 55*(3), 265-278. doi:10.1080/04250494.2021.1885975

Myhill, D., & Jones, S. (2009). How talk becomes text: investigating the concept of oral rehearsal in early years' classrooms. *British Journal of Educational Studies, 57*(3), 265-284.

Myhill, D., Jones, S., & Watson, A. (2013). Grammar matters: How teachers' grammatical knowledge impacts on the teaching of writing. *Teaching and teacher education, 36*, 77-91.

Myhill, D., Lines, H., & Watson, A. (2012). Making meaning with grammar: A repertoire of possibilities. *English in Australia, 47*(3), 29-38.

Myhill, D., & Watson, A. (2014). The role of grammar in the writing curriculum: A review of the literature. *Child Language Teaching and Therapy, 30*(1), 41-62.

Negro, I., Chanquoy, L., Fayol, M., & Louis-Sidney, M. (2005). Subject-verb agreement in children and adults: Serial or hierarchical processing? *Journal of psycholinguistic research, 34*(3), 233-258.

Nelson, N. W., & Van Meter, A. M. (2007). Measuring written language ability in narrative samples. *Reading & Writing Quarterly, 23*(3), 287-309.

Nippold, M. A., Ward-Lonergan, J. M., & Fanning, J. L. (2005). Persuasive writing in children, adolescents, and adults.

Ofsted. (2020). Inspection Dashboard. Retrieved from https://reports.ofsted.gov.uk/

Olinghouse, N. G. (2008). Student-and instruction-level predictors of narrative writing in third-grade students. *Reading and Writing, 21*(1-2), 3-26.

Olinghouse, N. G., & Graham, S. (2009). The relationship between the discourse knowledge and the writing performance of elementary-grade students. *Journal of educational psychology, 101*(1), 37.

Olinghouse, N. G., Graham, S., & Gillespie, A. (2015). The relationship of discourse and topic knowledge to fifth graders' writing performance. *Journal of educational psychology, 107*(2), 391.

Olinghouse, N. G., Santangelo, T., & Wilson, J. (2012). Examining the validity of single-occasion, single-genre, holistically scored writing assessments. In *Measuring Writing: Recent Insights into Theory, Methodology and Practice* (pp. 55-82): Brill.

Olive, T. (2004). Working memory in writing: Empirical evidence from the dual-task technique. *European Psychologist, 9*(1), 32-42.

Olive, T. (2014). Toward a parallel and cascading model of the writing system: a review of research on writing processes coordination. *Journal of writing research, 6*(2).

Olive, T., Alves, R. A., & Castro, S. L. (2009). Cognitive processes in writing during pause and execution periods. *European Journal of Cognitive Psychology, 21*(5), 758-785.

Olive, T., & Kellogg, R. T. (2002). Concurrent activation of high-and low-level production processes in written composition. *Memory & Cognition, 30*(4), 594-600.

Ormston, R., Spencer, L., Barnard, M., & Snape, D. (2014). The foundations of qualitative research. *Qualitative research practice: A guide for social science students and researchers, 2*, 52-55.

Painter, C. (2001). Understanding genre and register: Implications for language teaching. *Analysing English in a global context: A reader*, 167-180.

Paltridge, B. (1997). Genre, frames and writing in research settings.

Panter, A. T., & Sterba, S. K. (2011). *Handbook of ethics in quantitative methodology*: Taylor & Francis.

Paraskevas, C. (2006). Grammar apprenticeship. *English Journal*, 65-70.

Parker, R., Tindal, G., & Hasbrouck, J. (1991). Countable indices of writing quality: Their suitability for screening-eligibility decisions. *Exceptionality: A special education journal, 2*(1), 1-17.

Pearce, W. M. (2003). Does the choice of stimulus affect the complexity of children's oral narratives? *Advances in Speech Language Pathology, 5*(2), 95-103. doi:10.1080/14417040510001669121

Perfetti, C. A. (1997). The psycholinguistics of spelling and reading.

Persky, H. R., Daane, M. C., & Jin, Y. (2003). The Nation's Report Card: Writing, 2002.

Peterson, C., & McCabe, A. (1994). A social interactionist account of developing decontextualized narrative skill. *Developmental Psychology, 30*(6), 937.

Polanyi, L. (1982). Linguistic and social constraints on storytelling. *Journal of pragmatics, 6*(5-6), 509-524.

Pontecorvo, C. (1997). *Writing development: An interdisciplinary view* (Vol. 6): John Benjamins Publishing.

Proctor, C. P., Li, Z., Daley, S. G., Xu, Y., Graham, S., & Hall, T. E. (2020). Shared knowledge between reading and writing among middle school adolescent readers. *the elementary school journal, 120*(3), 507-527.

Pullum, G. K. (2018). *Linguistics: Why it matters*: John Wiley & Sons.

Puranik, C. S., & Al Otaiba, S. (2012). Examining the contribution of handwriting and spelling to written expression in kindergarten children. *Reading and Writing, 25*(7), 1523-1546.

Puranik, C. S., Lombardino, L. J., & Altmann, L. J. (2008). Assessing the microstructure of written language using a retelling paradigm.

Purves, A. C. (1988). *Writing across languages and cultures: Issues in contrastive rhetoric* (Vol. 2): SAGE Publications, Incorporated.

Ravid, D., & Katzenberger, I. (1999). Global text evaluation. *Developing Literacy Across Genres, Modalities, And Languages, 1*, 220-223.

Reed, W. M., Burton, J. K., & Kelly, P. P. (1985). The effects of writing ability and mode of discourse on cognitive capacity engagement. *Research in the Teaching of English*, 283-297.

Reilly, J., Zamora, A., & McGivern, R. F. (2005). Acquiring perspective in English: the development of stance. *Journal of pragmatics, 37*(2), 185-208.

Rezaei, A. R., & Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing Writing, 15*(1), 18-39.

Richards, K. (2003). *Qualitative inquiry in TESOL*: Springer.

Richmond, J., Burn, A., Dougill, P., Raleigh, M., & Traves, P. (2017). *Curriculum and assessment in English 3 to 11: A better plan*: Routledge.

Ritchey, K. D., McMaster, K. L., Otaiba, S., Puranik, C., Kim, Y., Parker, D. C., & Ortiz, M. B. (2016). *Indicators of Fluent Writing in Beginning Writers*.

Robson, C., & McCartan, K. (2016). *Real world research : a resource for users of social research methods in applied settings*. Chichester: Wiley.

Romig, J. E., Therrien, W. J., & Lloyd, J. W. (2017). Meta-analysis of criterion validity for curriculum-based measurement in written language. *The journal of special education, 51*(2), 72-82.

Rust, J. (1996). *Wechsler Objective Language Dimensions (WOLD)*: Psychological Corporation.

Sadler, D. R. (2013). Assuring academic achievement standards: from moderation to calibration. *Assessment in Education: Principles, Policy & Practice, 20*(1), 5-19.

Safford, K. (2016). Teaching Grammar and Testing Grammar in the English Primary School: The Impact on Teachers and their Teaching of the Grammar Element of the Statutory Test in Spelling, Punctuation and Grammar (SPaG). *Changing English-Studies in Culture and Education, 23*(1), 3-21. doi:10.1080/1358684x.2015.1133766

Salas, N., & Caravolas, M. (2019). Dimensionality of early writing in English and Spanish. *Journal of Literacy Research, 51*(3), 272-292.

Santangelo, T., & Graham, S. (2016). A comprehensive meta-analysis of handwriting instruction. *Educational Psychology Review, 28*(2), 225-265.

Savage, R., Kozakewich, M., Genesee, F., Erdos, C., & Haigh, C. (2017). Predicting writing development in dual language instructional contexts: exploring cross-linguistic relationships. *Developmental science, 20*(1), e12406.

Scardamalia, M., & Bereiter, C. (1987). Knowledge telling and knowledge transforming in written composition.

Schindler, R. M., & Bickart, B. (2012). Perceived helpfulness of online consumer reviews: The role of message content and style. *Journal of Consumer Behaviour, 11*(3), 234-243.

Schoonen, R. (2005). Generalizability of writing scores: An application of structural equation modeling. *Language Testing, 22*(1), 1-30.

Schoonen, R. (2012). The validity and generalizability of writing scores: The effect of rater, task and language. In *Measuring writing: Recent insights into theory, methodology and practice* (pp. 1-22): Brill.

Schoonen, R. (2019). Are reading and writing building on the same skills? The relationship between reading and writing in L1 and EFL. *Reading and Writing, 32*(3), 511-535.

Scott, C. (1988). Spoken and written syntax. *Later language development: Ages nine through nineteen, 1*, 49-55.

Scott, C. (2004). Syntactic ability in children and adolescents with language and learning disabilities. *Language development across childhood and adolescence, 3*, 111-133.

Scott, C., & Windsor, J. (2000). General language performance measures in spoken and written narrative and expository discourse of school-age children with language learning disabilities. *Journal of Speech, Language, and Hearing Research, 43*(2), 324-339.

Sénéchal, M., Thomas, E., & Monker, J.-A. (1995). Individual differences in 4-year-old children's acquisition of vocabulary during storybook reading. *Journal of educational psychology, 87*(2), 218.

Shanahan, T., MacArthur, C., Graham, S., & Fitzgerald, J. (2006). Relations among oral language, reading, and writing development. *Handbook of writing research*, 171-183.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52*(3/4), 591-611.

Sheard, M. K., Chambers, B., & Elliott, L. (2012). Effects of technology-enhanced formative assessment on achievement in primary grammar.

Shintani, N. (2018). Receptive Versus Productive Grammar Knowledge. *The TESOL Encyclopedia of English Language Teaching*, 1-6.

Snape, D., & Spencer, L. (2003). The foundations of qualitative research'in Ritchie, J. and Lewis, J (eds) Qualitative Research Practice. In: Sage, London.

Snowling, M. J. (2000). *Dyslexia*: Blackwell publishing.

Society, B. P. (2009). Code of ethics and conduct. In: Author Leicester.

Society, B. P. (2014). *Code of human research ethics*.

STA. (2015). *English grammar, punctuation and spelling test framework National curriculum tests from 2016: For test developers*.

STA. (2016). *2017 Interim teacher assessment frameworks at the end of key stage 2*. London: STA Retrieved from www.gov.uk/government/publications.

STA. (2017a). *2017 key stage 2 English grammar, punctuation and spelling: Administering the English grammar, punctuation and spelling test Paper 1: questions*.

STA. (2017b). *2017 key stage 2 English grammar, punctuation and spelling: Administering the English grammar, punctuation and spelling test Paper 2: spelling*.

STA. (2017c). *Key stage 2 tests: 2017 English grammar, punctuation and spelling test materials*. London, UK Retrieved from https://www.gov.uk/government/publications/key-stage-2-tests-2017-english-grammar-punctuation-and-spelling-test-materials.

Stannard, J., & Huxford, L. (2007). *The literacy game: The story of the National Literacy Strategy*: Routledge.

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using Curriculum-Based Measurement to Improve Student Achievement: Review of Research. *Psychology in the Schools, 42*(8), 795-819. doi:10.1002/pits.20113

Sumner, E. (2013). *Children with dyslexia: Spelling as a constraint on writing development.* Oxford Brookes University,

Sumner, E., Connelly, V., & Barnett, A. L. (2013). Children with dyslexia are slow writers because they pause more often and not because they are slow at handwriting execution. *Reading and Writing, 26*(6), 991-1008.

Sumner, E., Connelly, V., & Barnett, A. L. (2014). The influence of spelling ability on handwriting production: Children with and without dyslexia. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(5), 1441.

Swanson, H. L., & Berninger, V. W. (1996). Individual differences in children's working memory and writing skill. *Journal of experimental child psychology, 63*(2), 358-385.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological science in the public interest, 1*(1), 1-26.

Tolchinsky, L. (2006). The emergence of writing. *Handbook of writing research*, 83-95.

Van den Bergh, H., De Maeyer, S., Van Weijen, D., & Tillema, M. (2012). Generalizability of text quality scores. In *Measuring writing: Recent insights into theory, methodology and practice* (pp. 23-32): Brill.

Van der Lely, H. K., & Christian, V. (2000). Lexical word formation in children with grammatical SLI: a grammar-specific versus an input-processing deficit? *Cognition, 75*(1), 33-63.

Varnhagen, C. K. (2000). Shoot the messenger and disregard the message? Children? S attitudes toward spelling. *Reading Psychology, 21*(2), 115-128.

von Koss Torkildsen, J., Morken, F., Helland, W. A., & Helland, T. (2016). The dynamics of narrative writing in primary grade children: writing process factors predict story quality. *Reading and Writing, 29*(3), 529-554.

Wagner, K., Smith, A., Allen, A., McMaster, K., Poch, A., & Lembke, E. (2019). Exploration of new complexity metrics for curriculum-based measures of writing. *Assessment for Effective Intervention, 44*(4), 256-266.

Wagner, R., Puranik, C., Foorman, B., Foster, E., Wilson, L., Tschinkel, E., & Kantor, P. (2011). Modeling the development of written language. *Reading and Writing, 24*(2), 203-220.

Walter, K., Dockrell, J., & Connelly, V. (2021). A sentence-combining intervention for struggling writers: response to intervention. *Reading and Writing*, 1-26.

Watson, A. (2015). The problem of grammar teaching: A case study of the relationship between a teacher's beliefs and pedagogical practice. *Language and Education, 29*(4), 332-346.

Waugh, D., Warner, C., & Waugh, R. (2019). *Teaching grammar, punctuation and spelling in primary schools*: Learning Matters.

Weaver, C. (1996). *Teaching Grammar in Context*: ERIC.

Wechsler, D. (2005). Wechsler Individual Attainment Test–Second UK Edition (WIAT-II UK). In: London, UK: Psychological Corporation.

Weissenburger, J. W., & Espin, C. A. (2005). Curriculum-based measures of writing across grade levels. *Journal of School Psychology, 43*(2), 153-169.

Whitaker, D., Berninger, V., Johnston, J., & Swanson, H. L. (1994). Intraindividual differences in levels of language in intermediate grade writers: Implications for the translating process. *Learning and Individual Differences, 6*(1), 107-130.

Wijekumar, K., Graham, S., Harris, K. R., Lei, P.-W., Barkel, A., Aitken, A., . . . Houston, J. (2019). The roles of writing knowledge, motivation, strategic behaviors, and skills in predicting elementary students' persuasive writing from source material. *Reading and Writing, 32*(6), 1431-1457.

Wiliam, D. (1993). Validity, dependability and reliability in National Curriculum assessment. *The Curriculum Journal, 4*(3), 335-350.

Wiliam, D. (2001). Reliability, validity, and all that jazz. *Education 3-13, 29*(3), 17-21.

Willett, L., & Gardiner, A. (2009). *Testing spelling–exploring NAPLAN*. Paper presented at the Australian Literacy Educators Association Conference.

Williams, G. J., & Larkin, R. F. (2013). Narrative writing, reading and cognitive processes in middle childhood: What are the links? *Learning and Individual Differences, 28*, 142-150.

Wise, J. C., Sevcik, R. A., Morris, R. D., Lovett, M. W., & Wolf, M. (2007). The relationship among receptive and expressive vocabulary, listening comprehension, pre-reading skills, word identification skills, and reading comprehension by children with reading disabilities.

Wyse, D. (2001). Grammar for writing? A critical review of empirical evidence. *British Journal of Educational Studies, 49*(4), 411-427.

Wyse, D. (2006). Pupils' word choices and the teaching of grammar. *Cambridge Journal of Education, 36*(1), 31-47.

Wyse, D., Jones, R., Bradford, H., & Wolpert, M. A. (2013). *Teaching English, language and literacy*: Routledge.

Wyse, D., & Torgerson, C. (2017). Experimental trials and 'what works?' in education: The case of grammar for writing. *British Educational Research Journal, 43*(6), 1019-1047. doi:10.1002/berj.3315

Zimmerman, B. J., & Risemberg, R. (1997). Becoming a self-regulated writer: A social cognitive perspective. *Contemporary educational psychology*.

Zoccolotti, P., De Luca, M., Marinelli, C. V., & Spinelli, D. (2014). Modeling individual differences in text reading fluency: a different pattern of predictors for typically developing and dyslexic readers. *Frontiers in psychology, 5*, 1374.

Appendices

Appendix 1 – Ethics and Consent

A1.1 Ethical Considerations

A1.1.1 Methods

The choice of psychometric model has ethical implications (Panter & Sterba, 2011). In considering if the methods are appropriate, ethical, and that the benefits of the methodological rationale outweigh the harm, this research project aligns with the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (AERA, APA, & NCME, 1999). All the tasks that participants are being asked to carry out are similar to those which would be expected in a typical grammar or writing lessons within school; therefore, these tasks are not considered to be harmful. In addition, the measures chosen are age appropriate. This reduces any potential distress or anxiety related to difficulty or complexity of the tasks. Finally, the tasks are to be carried out as a whole class to reduce perception and anxiety around testing. No 1:1 testing will take place due to COVID-19 restrictions

A1.1.2 Sampling & Recruitment

There are key ethical issues that must be addressed when engaging in research with children. Consent from gatekeepers (CEOs/Headteachers) was obtained (Robson and McCartan, 2016) in a letter inviting school participation (See Appendix 1.2). A letter was then be sent to potential participants in order to ensure informed consent, alongside an information sheet which will set out all the details of the project for parents and for children. Power analyses were carried out to ensure that the study has sufficient power, but also to ensure that no unnecessary testing is carried out by having an unnecessarily large sample size.

A1.1.3 Gatekeepers & Voluntary Informed Consent

It is also important that permission was granted from parents of participants of the study and, as such, a full explanation of the study was sent to parents in the form of a letter (see Appendix 1.2) and was made available on the school websites. Parents/carers were required to opt-in to this study by signing the consent form (see Appendix 1.2). Good practice dictates that all participants, including children, should offer informed consent when engaging in a research study (Coolican, 2017); therefore, opt-in informed consent written in accessible language

explaining what is involved in participation was required. Children deposited their consent forms, which was also signed by parents, into a box in the playground or classroom to demonstrate voluntary informed consent.

A1.1.4 Safeguarding/child protection

The principal researcher is trained in safeguarding and sought assurances that DBS checks were in place in each of the school. Access to the 'Keeping Children Safe' document was in place at each research school. The researcher ensured that the Designated Safeguarding Lead at each school was known and that local school policies for safeguarding are adhered to. The benefit of the teacher delivering the tests is the familiarity of the teacher-tester for the children, as well as the familiarity of safeguarding procedures by the teacher-tester.

A1.1.5 Disclosures/limits to confidentiality

There were no foreseeable reasons for breaking confidentially, excluding issues of child protection and safeguarding. Although no issues were reported, the Designated Safeguarding Lead in the given school would have been made aware through an Initial Concern Form.

A1.1.6 Insider Research

As one of the schools within this study form part of my own institution (multi-academy trust), it was essential to consider the ethical concerns of being an insider researcher. Whilst engaging in insider research can offer benefits to the researcher in terms of familiarity of context and knowledge of participants (Drake & Heath, 2010), it may also result in bias and assumptions regarding data that is not wholly objective. Insider research can result in participants potentially feeling obligated to take part in the study. Complex social relations between colleagues can be evident in ways that may not be apparent with an outside researcher (Miles et al. 1994). In addition to informed consent, a full debrief and withdrawal will be offered within the process.

A1.1.7 Risks to participants and/or researchers

Owing to COVID-19 restrictions, no researchers attended research sites. Key stakeholders such as gatekeepers and Designated Safeguarding Leads were identified and contact procedures recorded. Researchers ensured that in-school staff supervised the children engaging in the study in-line with individual school policies.

To ensure that children are not fatigued by the research project, data collection was part of the normal school day and was broken up into a number of sessions with regular breaks

factored in. Finally, there was a potential risk of harm through distress and embarrassment for participants who find formal assessments distressing (Alderson & Morrow, 2011). Assessments were referred to as 'activities' to reduce stress. Children were briefed about their option to withdraw from the study at any point. Children will also be briefed that all data is anonymised and that no participant will be identifiable within the study.

A1.1.8 Confidentiality and Anonymity; Data Storage

Participants' data will be safeguarded in accordance with General Data Protection Regulations (2018). Participants were be allocated an ID number, rather than using their names, to ensure anonymity before scoring and data input are carried out. Data was stored on an encrypted laptop with files which will be password-protected. Data will then be uploaded to the University College London Network and removed from the laptop.

A1.1.9 Reporting, dissemination and use of findings

The aggregated and anonymised findings of this study will primarily be used to form the thesis of my EdD. In addition, findings will be shared and used to inform curriculum policy within several primary Multi-Academy Trusts across the country. Participants will receive an overview of the findings through optional workshops throughout the participant schools. Owing to the scarcity of data in this area, publication through an appropriate journal will be sought to widen the body of knowledge.

A1.2 – Letters

A1.2.1 – Invitation to School Gatekeepers

Dear [Insert Headteacher name here]

Thank you for your interest in our exciting new research project. This study is exploring the English Grammar, Punctuation and Spelling Test and its relationship with children's writing outcomes. The results of this project will help us better understand factors that directly relate to writing and it will be used to help school leaders and policy makers make decisions regarding administering the EGPS test, which is likely to become non-statutory in the coming years.

What will we ask the children to do?

As part of the normal school day, we will be inviting Year 6 children to engage in a number of activities which will involve writing, spelling, reading and listening. These activities will be administered either by the class teacher and will be delivered to the class as a whole to minimise disruption. These activities will probably take a morning or an afternoon to complete – whichever is convenient for the school. In addition, we would like children to take the 2018 EGPS test, which is likely to be part of your normal preparations for SATs.

What do we need from your school?

- Assistance in seeking parental consent
  (we will provide the relevant paperwork)
- Some standard administrative data and pupil data in order to profile the groups of children involved
  (e.g. EAL/FSM/SEN)
- Collaboration in arranging the relevant activities and tests
  (which can be administered by teachers or the researcher and will be marked by the researcher)

Data Protection and ethics

All pupil data will be treated with the strictest confidence and stored in accordance with current data protection legislation. All results from the study will be anonymised so that no schools or individual pupils will be identified; confidentiality will be maintained at all times. The research team is from the world-leading University College London and the UCL Institute of Education so we have lots of experience of dealing with data ethically.

Do you have any further questions?

To discuss this project further, simply contact the principle researcher, James Canniford (james.canniford.16@ucl.ac.uk) as soon as possible.

Happy to participate?

If you are happy to offer consent as the gatekeeper for you school, please email the principle researcher (james.canniford.16@ucl.ac.uk) with the following reply:

Dear James,

*As the headteacher/CEO and gatekeeper of **[enter school name]**, I give permission for the Year 6 pupils of **[enter school name]** school to participate in the EGPS research study.*

*Yours, [**Insert name here**]*

Thank you for your interest in our research project.

James Canniford
Principle Researcher – UCL Institute of Education
Director of English – Quality First Education Trust

A1.2.2 – Invitation to Parent/Carers

Dear Parents/Carers,

As COVID-19 has affected our children so significantly this year, carrying out assessments to find the gaps in their learning is essential. Your child's school has teamed up with UCL IOE to ensure that effective assessments help get your child back on track as soon as possible.

As part of our ongoing drive for improvement within schools in this unprecedented year, we are interested in exploring the link between grammar and writing quality. As part of our research, we will be exploring the benefits and impacts of carrying out short diagnostic writing tasks and what links they demonstrate with grammar knowledge. We will be using these assessments to measure the impact of the work that has been done/missed in EGPS (English grammar, punctuation, and spelling) and its impact on children's writing.

All children in Year 6 will be taking part in the tasks that form this study as part of their normal school provision. We are required to seek permission to use your child's data as part of this study. You can find more details regarding the study on the reverse of this letter. All children will carry out the assessments, regardless of their participation in the research.

Once the information has been gathered and shared with teachers, all children's research data in this study will be anonymised and will then be used to see how we can improve our understanding of national assessments in order to support our children's learning more effectively. A presentation of the findings from the research will be shared with the school community via an online parents' meeting in 2021 and as part of a research report for the Institute of Education. All data collected will be full in accordance with the Data Protection Act. **No children will be identifiable in this study.**

Please sign and return the attached permission card if you wish to give consent for your child's data to be used in this study. Children will be encouraged to post the permission into the consent box themselves to confirm that they too are giving permission for their data to be used in the study.

Should you have any further enquiries, please do not hesitate to contact Mr Canniford. Contact details can be found on the information sheet.

Many thanks,

 Mr J Canniford
Department of Psychology & Special Needs
UCL Institute of Education

A1.2.3 – Invitation to School Gatekeepers

<ins>Participant Information Sheet for Year 6 Parents:</ins> UCL Research Ethics Committee

Title of Study: The EGPS Test: An exploration of the relationship between the EGPS Test and writing outcomes.
Department: Psychology and Special Needs, Institute of Education, UCL.
Name and Contact Details of the Principal Researcher: James Canniford – james.canniford16@ucl.ac.uk

1. **What is the project's purpose?**
    The purpose of this project is to understand the gaps in children's EGPS and writing knowledge, and to look at the link between learning about English grammar, punctuation, and spelling (EGPS) and how this can improve writing.

2. **Why has my child been chosen?**
    A number of schools across England have been identified based on their school's achievement in Spelling, Grammar and Punctuation. All children will take part in the activities as part of their normal school day. We are seeking permission to use **the data** from the activities as part of our study.

3. **Does my child have to take part?**
    All children will take part in the activities of this study as part of the normal school day. It is up to you to decide whether you **allow your child's anonymised data** to form part of this study. You can withdraw the data at any time without giving a reason.

4. **What will happen if my child takes part?**
    Children will be taking part in a number of short writing tasks which will take approximately one hour. Each task will involve writing sentence or paragraph of writing about a particular theme. The quality of these written paragraphs will be compared with the children's results in the EGPS test in the SATs Tests. Children's data can be withdrawn from the study up until **01/08/2021**. The tasks that the children take part in will be carried out with your child's class teacher in-line with your school's COVID guidelines.

5. **What are the possible disadvantages and risks of taking part?**
    While there is always the possibility of children becoming distressed when being assessed in any aspect of the curriculum, we will take every measure possible to ensure that the experience is an enjoyable one. Any child who becomes uncomfortable during the course of the research will be invited to withdraw from the activity.

6. **What are the possible benefits of taking part?**
    Whilst there are no immediate benefits for the children participating in the project, it is hoped that this work will improve our understanding of the teaching of grammar and its relationship with the EGPS Test. This will, in turn, help us to improve the quality of teaching in our school across the country.

7. **What if something goes wrong?**
    If you have any complaints regarding the research project, you should contact the principle researcher on the email address at the top of the page. If you feel that your complaint has not been handled to your satisfaction, you can contact the Chair of the UCL Research Ethics Committee – ethics@ucl.ac.uk

8. **Will my child's participation in this project be kept confidential?**
    All the information that we collect about you during the course of the research will be kept strictly confidential. Your child will not be identifiable in any reports or publications, although individual children's data will be shared with teachers for assessment purposes.

9. **Limits to confidentiality**
    Confidentiality will be respected unless there are compelling and legitimate reasons for this to be breached. If this was the case, we would inform you of any decisions that might limit your child's confidentiality.

10. **What will happen to the results of the research project?**
    The results of this study will be published as part of a doctoral thesis at UCL Institute of Education. The data from the study will be stored on UCL's Secure Database for 10 years.

11. **Data Protection Privacy Notice**
    The data controller for this project will be University College London (UCL). The UCL Data Protection Office provides oversight of UCL activities involving the processing of personal data and can be contacted at data-protection@ucl.ac.uk. UCL's Data Protection Officer is Lee Shailer and he can also be contacted at data-protection@ucl.ac.uk. Any personal data will be processed for the purposes outlined in this notice. The legal basis that would be used to process your child's personal data will be the provision of your consent. You can provide your consent for the use of personal data in this project by completing the consent form that has been provided to you. Any personal data will be processed so long as it is required for the research project. Where we are able to anonymise or pseudonymise the personal data you provide we will undertake this and will endeavour to minimise the processing of personal data wherever possible. If you are concerned about how your child's personal data is being processed, please contact UCL in the first instance at data-protection@ucl.ac.uk. If you remain unsatisfied, you may wish to contact the Information Commissioner's Office (ICO). Contact details, and details of data subject rights, are available on the ICO website at: https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/individuals-rights/

16. **Contact for further information**
    Should you require more information regarding the research project, please do get in contact with the principal researcher, James Canniford, on the email at the top of this information sheet.

Thank you for reading this information sheet and for considering your child's participation in this research study.

EGPS & Writing Research Project – Children's information

A team of researchers at University College London (currently the world's number one university for education!) are interested to learn all about how grammar and writing are linked. **We would like you to help us with our research!**

Why are we doing this research?
This study examines how learning about English grammar, punctuation and spelling (EGPS) can improve writing.

What do I need to do?
Just like in your normal English/grammar lessons, all children will be asked to write a number of sentences and paragraphs of writing in English based on a variety of interesting topics. You will also be asked to do a reading task. Don't worry, it's not like a SATs Test – its multiple choice!

What am I giving permission for?
All children will be taking part in the writing tasks as part of the normal school day so you cannot skip this part – sorry! You are giving permission for your **data** to be used in a real research project. Your data will be anonymous, which means no one will know which information belongs to you (you will even be given a secret codename!). The only people who this information will be shared with is your teachers so that they can help you catch up after the long COVID-19 lockdown gaps in learning.

Please sign and post this consent form in the box in your classroom.

Name:

Class:

 Parent Permission:

Child Permission:

Appendix 2 – EGPS Paper 1

Appendix 3 – EGPS Paper 2

Appendix 4 – EGPS Test Framework

Appendix 5 – Teacher Assessment Framework

Appendix 6 – National Curriculum Spelling Appendix 1

Appendix 7 – Canniford (2019) – Institution Focused Study