Bangor University

DOCTOR OF PHILOSOPHY

An Exploration of the potential markers of literacy difficulties in Welsh-English Bilinguals

Owen, Carla

Award date:
2021

Awarding institution:
Bangor University

Link to publication

AN EXPLORATION OF THE POTENTIAL MARKERS OF LITERACY DIFFICULTIES IN WELSH-ENGLISH BILINGUALS

CARLA MARIE OWEN

PhD BILINGUALISM

BANGOR UNIVERSITY

Contents

DECLARATION

Yr wyf drwy hyn yn datgan mai canlyniad fy ymchwil fy hun yw'r thesis hwn, ac eithrio lle nodir yn wahanol. Caiff ffynonellau eraill eu cydnabod gan droednodiadau yn rhoi cyfeiriadau eglur. Nid yw sylwedd y gwaith hwn wedi cael ei dderbyn o'r blaen ar gyfer unrhyw radd, ac nid yw'n cael ei gyflwyno ar yr un pryd mewn ymgeisiaeth am unrhyw radd oni bai ei fod, fel y cytunwyd gan y Brifysgol, am gymwysterau deuol cymeradwy.

I hereby declare that this thesis is the results of my own investigations, except where otherwise stated. All other sources are acknowledged by bibliographic references. This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree unless, as agreed by the University, for approved dual awards.

ACKNOWLEDGEMENTS

First and foremost I'd like to thank my supervisor Enlli for having faith in me from the very beginning. I will be forever grateful to her for her unwavering enthusiasm and support, particularly during some of the more challenging times in the last few years, which has helped give me the confidence to complete this thesis. I feel lucky to have had such an amazing supervisor. Diolch o'r galon Enlli! Thank you also to my second supervisor Nia for her support, guidance and enthusiastic approach to the project over the years.

Thank you to the ESRC and Bangor University for funding this project and for giving me the opportunity to work in such an interesting field of study.

I'd like to say a huge thank you to the schools who participated in the studies, the parents who provided consent for their children to take part and of course to the children who took time out of their lessons to complete the tasks with such enthusiasm. I enjoyed working with you all and without your support this project wouldn't have been possible.

Thank you to the very knowledgeable staff at the Miles Dyslexia Centre at Bangor University for their expertise, advice and willingness to help. Thank you also to Marketa for allowing me to use the ELDEL when in its earlier stages.

Thank you to all my friends and colleagues at Bangor University for your support and advice and to Susan for being such a supportive chairperson.

A big thank you goes to Dave for his expertise and endless patience with bringing 'Y Goriad' to life, it wasn't an easy process but we got there! Thank you also to Steve for his brilliant AV knowledge and to Huw for lending his voice and his good humour during the process, I really enjoyed working with you both.

Finally, I'd like to say a huge thank you to my family and friends for their amazing support during the completion of this thesis. Your patience, understanding and love during this time has meant the world to me, there just aren't enough words to express my gratitude, thank you so much.

ABSTRACT

Children who receive their education in Wales are exposed to two very different languages: English, which has an opaque orthographic system, and Welsh, which has a transparent one. Children attending English-medium schools are introduced to basic Welsh vocabulary and phrases from the outset, but progress to studying Welsh as an L2 subject after age 7 and continue to do so up until the age of 16. Children in Welsh-medium schools are immersed in Welsh until the end of the Foundation Phase (age 7) and continue to receive their education primarily through the medium of Welsh, with English being taught as a subject only. At the same time, the majority of assessment tools are currently only available in English and are standardised on monolingual norms. These tools have been developed with the nuances of English in mind, and may not always be appropriate for use with bilinguals or appropriate to be adapted to fit other languages. In the Welsh-medium context, this challenges the notion of equality, particularly in relation to access to services for Welsh-speaking children who may require assessment and/or intervention in Welsh, and this can result in children not receiving the support they need until a later stage in their education.

Given that Welsh has a transparent orthography, and children attending Welsh-medium schools do not produce high volumes of writing in English, any 'traditional' markers of literacy difficulties, as manifested in English, may not appear in their Welsh. The studies presented in this thesis sought to explore potential markers of literacy difficulties in Welsh, with the aim of providing useful findings that could be used to support the future development of screening tests for Welsh-English bilingual children. Study 1 provides an error analysis of children's written text on a 3-minute writing task in English and Welsh. Findings revealed interesting correlations between Welsh and English children's performance on tests of reading ability and the errors made, particularly errors relating to phonological awareness. The study provided information about the types of errors made by bilingual Welsh-English children that warrant further investigation beyond this thesis.

Studies 2 and 3 explored areas of the Welsh language that require higher levels of phonological processing, namely mutation and plural morphology. Due to the absence of appropriate measures, novel tasks were created for these studies that measured children's knowledge and application of Welsh plurals and mutation in the form of computer and paper based tasks related to writing, reading and oral ability. The studies yield interesting findings with regards to the relationship between reading ability, general phonological awareness abilities, and children's scores on the mutation and plurals tasks and the ability of these tasks to predict levels of reading. Results suggest that pupils' knowledge of Welsh mutation and plural morphology may be indicative of literacy abilities in the Welsh language, and may be useful structures to explore as potential language-specific items for the purpose of assessment in Welsh. However, further research is needed to determine which specific elements of these grammatical structures are the most effective markers.

ACRONYMS

2L1	English/Welsh Bilingual
ALN	Additional Learning Needs
AoO	Age of Onset
AWRT	All Wales Reading Test
AWRTE	All Wales Reading Test English
AWRTW	All Wales Reading Test Welsh
BPVS	British Picture Vocabulary Scales
CDI	Communicative Development Inventories
CELF	Clinical Evaluation of Language Fundamentals
CTOPP	Comprehensive Test of Phonological Processing
DLD	Developmental Language Disorder
FP	Foundation Phase
FSM	Free School Meals
GPC	Grapheme-Phoneme Conversion
IDP	Individual Development Plan
L1	First Language
L2	Second Language
LI	Language Impairment
LNF	Literacy and Numeracy Framework
LoE	Length of Exposure
LRE	Letter Recognition English
LRW	Letter Recognition Welsh
NARA	Neale Analysis of Reading Ability
NRT	National Reading Test
NRTE	National Reading Test English
NRTW	National Reading Test Welsh
OoA	Order of Acquisition
PGC	Prawf Geirfa Cymraeg
PGM	Prawf Glannau Menai
SES	Socio-economic Status
SLI	Specific Language Impairment
SM	Soft Mutation
SRep	Sentence Repetition

TD	Typically Developing
WPS	Words per Sentence

List of Tables

List of figures

INTRODUCTION

In recent years, there has been a growing interest in the development of Welsh-medium education and in the desire to preserve Welsh as a minority language. This has resulted in an increase in the use of Welsh within schools in Wales through, for example, immersion programmes and an increase of incidental Welsh within predominantly English schools. These increasing amounts of exposure to Welsh allow children the opportunity to become competent speakers/users of both languages and to thrive bilingually. Whilst increased interest of this kind is a welcomed development, particularly in light of the recent vision for Wales to host a million Welsh speakers by 2050 (Cymraeg 2050: A million Welsh speakers - Welsh Government, 2017a) paradoxically, this increase in Welsh-medium and bilingual education has not been matched with similar developments within support services (Speech and Language Therapy; Educational Psychology; Social Work), particularly in relation to the development of appropriate assessment resources for the purpose of psychometric testing. This deficiency in appropriate assessment tools poses a challenge for educators of bilingual children. As an experienced teacher of bilingual children myself, having worked in both Wales and Spain, I understand the importance of appropriate assessment tools in order to gain an accurate view of a child's abilities. Although it is understood that children with literacy difficulties will present difficulties in both of their languages, without a suitable measure of their abilities in both languages it is difficult to distinguish between an underlying literacy difficulty such as dyslexia and a child exhibiting delayed or deviant development that is typical of a child learning a second language.

Whilst a lot is known about the markers of literacy difficulties in English, which has led to the development of a variety of tests measuring difficulties such as dyslexia, much less is known about the potential difficulties facing learners who are receiving the majority of their education in other languages such as Welsh. Anecdotal evidence from teachers dealing with children who show reading and writing difficulties that are educated through the medium of Welsh

suggest that pupils may not demonstrate clear indications of any obvious difficulties (as traditionally assessed in English) due to the transparency of the orthography (see Chapter 2). It may not be until later in their education, when more focus is given to the development of their English, that any potential difficulties begin to surface within their reading and writing. This can be problematic in two ways: first, for learners who run the risk of spending many years without appropriate intervention to enable them to make suitable progress in their learning; and second, for educators who have very limited diagnostic tools available to enable them to gain a formal diagnosis of literacy difficulty and thus suitable support to provide the intervention necessary. Whilst the transparency of Welsh may well mask the types of errors that are typical of those with literacy difficulties, what tends to underpin most cases of literacy disorders is a lack of phonological awareness (see Chapter 2). In a language like Welsh, phonological awareness difficulties may lead to specific types of difficulties with structures that require certain levels of awareness to subtle phonological alterations – the Welsh mutation system and the noun plural system being two cases in point.

With regards to the profile of a bilingual learner, little is known about the linguistic profiles and language trajectories of children learning Welsh in general, either as a first (L1) or a second (L2) language, and little is known about the impact of learning both Welsh and English simultaneously or successively on those profiles. What is known is that bilinguals, all over the world, have unique and varied linguistic profiles, and that any tests that aim to provide a true assessment of their linguistic abilities should be developed with the types of target speakers in mind, and that learners should be tested in both of their languages.

The research presented in this thesis aimed to explore some of the issues relating to the assessment of Welsh-English bilinguals in receipt of Welsh-medium education. In particular, the studies presented explored the effectiveness of a variety of tasks in predicting possible literacy difficulties among pupils, with a specific focus in Study 2 and 3 on those structures that require strong phonological awareness abilities and that are known to provide

specific problems to learners according to specialist teachers (e.g., Davies, 2016).

Chapter 1 provides an outline of education provision in Wales, including the role of the Welsh language, the linguistic context of schools and Welsh within the curriculum. Chapter 2 goes on to present a review of the literature pertaining to bilingual assessment, including issues in relation to assessing bilinguals, similarities and differences in the performance of monolingual and bilingual learners and the types of tests currently available and being developed with bilingual speakers in mind. In Chapter 3 the first of the three studies presented in this thesis is introduced and the findings discussed. This study explores the errors Welsh-English children make in their written work when completing a 3-minute writing task. The study reveals some interesting findings that are used to influence the design of Study 2. Chapter 4 provides an introduction and rationale to Studies 2 and 3, including a description of each of the background measures used within each of the studies, the participant sample and findings of the pilot study.

Chapter 5 presents the design and results of Study 2, which explores the relationship between Welsh-English children's performance on a number of novel tasks measuring morpho-phonological ability (Welsh mutation), and their scores on Welsh and English reading tests and Welsh and English phonological awareness tests, as typically used in assessment tools of English literacy. Chapter 6 presents the design and findings of Study 3, which explores the relationship between children's performance on tasks of Welsh plural morphology and their performance on standardised reading and phonological awareness tasks. Both studies yield interesting findings for educators and researchers.

Finally, Chapter 7 provides a general summary and discussion of the findings from Studies 1, 2 and 3 and provides recommendations for future research.

CHAPTER 1

<u>THE LINGUISTIC CONTEXT OF EDUCATION IN WALES</u>

Introduction

The linguistic context of education in Wales is both interesting and complex. One of the oldest languages of Europe, *Cymraeg* 'Welsh' is currently spoken by 891,100 people aged over three in Wales (Welsh Government, 2019), which equates to 29.6% of the population. Whilst at the start of the 20th Century Welsh speakers were often monolingual speakers of Welsh, today, all Welsh speakers (over the age of 3 years) are developing bilinguals with English (and in some cases another language), although the 'type' of bilingual they might be will vary from one individual to the next. Welsh is transmitted in the home either exclusively, leading to L1 competence, or alongside another language in a one-parent-one language household, leading to 2L1 competence in many cases. In other cases, children become Welsh-speakers via exposure to Welsh at school, sometimes as an L2 (e.g. 'late bilinguals' Baker, 2002) or as an L1 (e.g. when placed in an early immersion setting). In either case, education holds an important role in facilitating language development and providing continuous and rich exposure to the language in both the oral and literacy domains. However, for some children, education is delivered through the medium of Welsh whereas for others, it is delivered as a subject only, which varies their exposure to the language. It is important to note that L1 competence in a language does not indicate that the bilingual speaker can be considered within a monolingual norm. A bilingual is not two monolinguals and therefore should not be treated as such in any assessment of their language abilities. Regardless of how a child becomes bilingual, the linguistic medium of the school or where they may be on the bilingualism continuum (assuming a continuum starting at the one end with L1 competence in both languages - the so called 'balanced bilingual' - and ending at the other end with minimal knowledge in the second language), they deserve to be supported equally, both at school and by external service providers where appropriate (Speech

and Language Therapists, Educational Psychologists, etc.). At present, however, there are inconsistences in the services that are available to support children with language difficulties, including services that are available through the medium of Welsh and services that aim to support children's development with the Welsh language. This is not least the case in relation to supporting children's literacy abilities in Welsh. The limited number of Welsh-speaking professionals (for example, targets to increase number of Welsh-speaking teachers stated in Cymraeg:2050, Welsh Government, 2017a), coupled with a lack of understanding of the linguistic profiles of different types of Welsh-English bilinguals due to limited research, and a dearth of language-related tools to help identify issues and support children who are conducting most of their education through the medium of Welsh (as also highlighted in Cymraeg: 2050, Welsh Government, 2017a and the Additional Learning Needs and Education Tribunal Wales Act, 2018 (https://gov.wales/additional-learning-needs-and-education-tribunal-wales-act-explanatory-memorandum)), is a concern. In order to address these issues, it is useful to understand the linguistic context of education in Wales and the types of services that are (or are not) available for the different types of speakers.

This Chapter will provide an overview of the linguistic context of schools in Wales, as they currently stand, the role of Welsh within the curriculum and the current situation with regards to Welsh-speaking children with Additional Learning Needs. The Chapter ends with a summary of some of the issues related to Welsh in education which will be discussed further in the next chapter.

The role of Welsh within schools in Wales

The Welsh Language Act was introduced in Wales in 1993 stating that Welsh should be treated equally to English. This legislation was later developed in 2011 when a legal framework for the status and use of Welsh, known as the Welsh Language Measure[1] (2011), was introduced. This framework, regulated

---

[1] https://www.legislation.gov.uk/mwa/2011/1/contents/enacted

by the Welsh Language Commissioner, imposes Welsh Language Standards on all public services to ensure that the language is not treated any less favourably than English. Welsh also has protected status through the European Charter for Regional and Minority Languages Treaty, adopted in 1992 under the support of the Council of Europe to protect and promote historical, regional and minority languages in Europe. These are all significant steps forward in helping to promote Welsh. More recently, the Welsh Government introduced the *Cymraeg 2050* strategy, an action plan that aims to reach one million Welsh speakers in Wales by the year 2050, increase the proportion of speakers who use the language daily from 10% to 20%, and to ensure that the appropriate infrastructure is in place to allow these two goals to be achieved (Welsh Government, 2017a).

Realising the aims of Cymraeg 2050 is rooted in the education system in Wales. This includes targeting the early years by expanding Welsh-medium Early Years provision, the promotion of Welsh within statutory education and the workplace to ensure that not only is the language being promoted within schools, but also that there are sufficient numbers of teachers who can teach through the medium of Welsh. There is also an aim to develop more tools and resources to help support children with Additional Learning Needs (ALN). Delivering on the aims of the strategy will primarily lie on the shoulders of educators and schools in Wales and so it is essential that they are provided with the support required to help deliver those aims. To this end, this thesis aims to offer some support by providing further research into how children who are educated in Welsh-medium schools, who have literacy difficulties, can be effectively assessed and thus receive the educational support they require in their mother tongue.

Some steps have already been taken to drive this new language initiative. For example, the introduction of the National Literacy and Numeracy Framework (LNF) (Welsh Government, 2013a), a framework designed to improve standards of Numeracy and Literacy in Wales, was established to promote a more cross-curricular approach incorporating literacy and numeracy within

each subject area across the curriculum (the LNF is discussed further in the next section). There has also been an increased emphasis on the use of Welsh in the classroom (e.g. via the implementation of Y Siarter Iaith[2]) and the importance of Welsh in the wider community as highlighted in the new *Professional Standards for Teaching and Leadership* (Welsh Government, 2017b). Due to the current provision available, children's exposure to and acquisition of Welsh varies considerably depending on the type of provision they attend. However, with such a challenging language initiative in place there is a necessity to ensure that appropriate resources are available to help support and develop Welsh-speaking children's skills.   The 'Welsh in Education: Action Plan 2017-2021' (Welsh Government, 2017c) notes:

> The Evaluation of the Welsh-medium Education Strategy: Review of the Use and Quality of Resources Commissioned by the Welsh Government's Welsh in Education Unit14 (2014) found that although our resources commissioning programme makes a valuable contribution, the limited range of resources in a number of areas remains problematic for headteachers and practitioners, and practitioners in the Welsh-medium sector still face an additional workload in preparing Welsh-medium resources to fill gaps (p. 12).

This again highlights the need for better quality and quantity of resources to help support Welsh teachers in delivering education through the medium of Welsh in order to help the language thrive among pupils.

The linguistic medium of schools in Wales

Welsh became part of the National Curriculum following the introduction of the Education reform Act (1988)[3]. By 1990 Welsh was a compulsory subject for all Key Stage 1-3 (KS1: 4-7 years; KS2: 7-11 years; KS3: 11-16 years)

---

[2] See also http://d6vsczyu1rky0.cloudfront.net/31669_b/wp-content/uploads/2018/02/Handbook-Cymraeg-Campus.pdf
 https://gov.wales/cymraeg-education/schools/welsh-language-charter

[3] https://www.legislation.gov.uk/ukpga/1988/40/contents

learners in Wales and in 1999 also became compulsory for learners across Wales in Key Stage 4 (16-18 years) (Welsh in Education: Action Plan 2017-2021).

However, within Wales, the provision of Welsh varies across Welsh-medium, bilingual and English-medium (with Welsh as a subject) providers. What is meant by Welsh-medium provision and bilingual provision can vary quite dramatically, both within and across local authorities within Wales. Due to this variety, it was necessary to categorise schools according to their linguistic medium in order for parents to be able to make informed choices about which school they would like their child to attend. The Education Act 2002 provided a definition of 'Welsh-speaking' schools in Wales, stating that:

> …. a school is Welsh-speaking if more than one half of the following subjects are taught (wholly or partly) in Welsh - (a) religious education, and (b) the subjects other than English and Welsh which are foundation subjects in relation to pupils at the school (Section 105(7))*.

Clearly, this definition did not provide a complete description of the varying 'types' of Welsh education in Wales, but it did provide a starting point in terms of defining what would be categorised as a 'Welsh-speaking' school. Subsequent to this definition, the Welsh Government (2007) published a more thorough classification index, presented in their *Defining schools according to Welsh medium provision* document. According to this approach, Welsh-medium primary schools are categorised as Type A if at least 50% of the pupils are taught mainly though the medium of Welsh and Type B if less than 50% of the pupils receive part of their education through the medium of Welsh or attend Welsh-medium classes. Tables 1 and 2 below provide a description of each school type in Wales according to language use.

Table 1: Language categorisation of Primary Schools in Wales (adapted from 'Defining schools according to Welsh medium provision').

School Category	Curriculum	Language of School
Welsh Medium Primary School	Teaching within the Foundation Phase (4-7 years of age) is through the medium of Welsh only. In KS2 Welsh is the main language of instruction with at least 70% of the curriculum being delivered in Welsh. English not formally introduced until KS2.	Welsh is the dominant language of the school although communication is in both English and Welsh where applicable.
Dual Stream Primary School	Both English language and Welsh language provision is delivered at these schools. Parents/pupils have the option of either Welsh or English as their main language of learning.	English and Welsh are used for communication dependent on the chosen language of provision. High priority is given to creating a Welsh ethos within the school.
Transitional Primary School	Welsh is the main language of teaching during the Foundation Phase. Both English and Welsh are used at KS2 but Welsh is the dominant language and used when teaching between 50-70% of the curriculum.	Welsh is the dominant language of the school but communication is also in English where applicable.
Predominantly English	Foundation Phase learning is delivered in both languages however English is the dominant	Both English and Welsh are used for communication with parents/pupils

Medium Primary School but significant use of Welsh	language of instruction. Welsh is used for teaching between 20-50% of the curriculum.	although emphasis is also placed on creating a Welsh ethos.
Predominantly English Medium Primary School	The Foundation phase and KS2 are delivered in English with the exception of Welsh at KS2 which is taught as a second language subject and some other subjects may also include aspects of Welsh. Welsh medium teaching is less than 20%.	English is the dominant language of communication. Some everyday use of Welsh (incidental Welsh) is also encouraged.

Table 2: Language categorisation in Secondary Schools in Wales[4].

School Category	Curriculum		Language of School
Welsh-Medium Secondary School	Welsh is the language of instruction for all subjects except English. There may also be some introduction of English terms in one or two subjects.		Predominantly Welsh is used for teaching and communication. English is also used for communication where applicable.
Bilingual Secondary School	This category is split into 4 sub divisions.		Language of school will be dependent on the category of the provision however a Welsh ethos will be promoted in all schools and communication will be in both English and Welsh.
Category (AB)	**2A**	Welsh is the main language of instruction with at least 80% of subjects being delivered to all pupils through the medium of Welsh. One or two subjects may be delivered in English also.	
Category (BB)	**2B**	The teaching of at least 80% of all subjects will be through the medium of Welsh but will	

---

[4] Content adapted from '*Defining schools according to Welsh medium provision*' https://gov.wales/sites/default/files/publications/2018-02/defining-schools-according-to-welsh-medium-provision.pdf

	also be available through the medium of English.	
Category 2C (CB)	For 50-79% of subjects teaching will be through the medium Welsh and in English.	
Category 2CH (CH)	Teaching of all subjects to all pupils will be through the medium of Welsh and English.	
Predominantly English Medium Secondary School with significant use of Welsh	Teaching is in both English and Welsh with 20-49% being delivered through the medium of Welsh.	Although a strong emphasis is placed on creating a Welsh ethos, communication will be predominantly through the medium of English or both languages where applicable.
Predominantly English Medium	All pupils are taught mainly through the medium of English with Welsh being taught as a second language. One or two other	English is the dominant language for communication although the use of everyday Welsh is also promoted.

Secondary School	subjects may also be taught through the medium of Welsh as an option or in both languages.	Communication is in both English and Welsh as applicable.

In the same way as the 'types' of language provision vary from one school to the next, the number of pupils attending the different types of linguistic mediums varies quite dramatically. Figure 1 shows the proportion of pupils attending the various types of schools.

Figure 1: Percentage of school pupils in Wales according to School language category
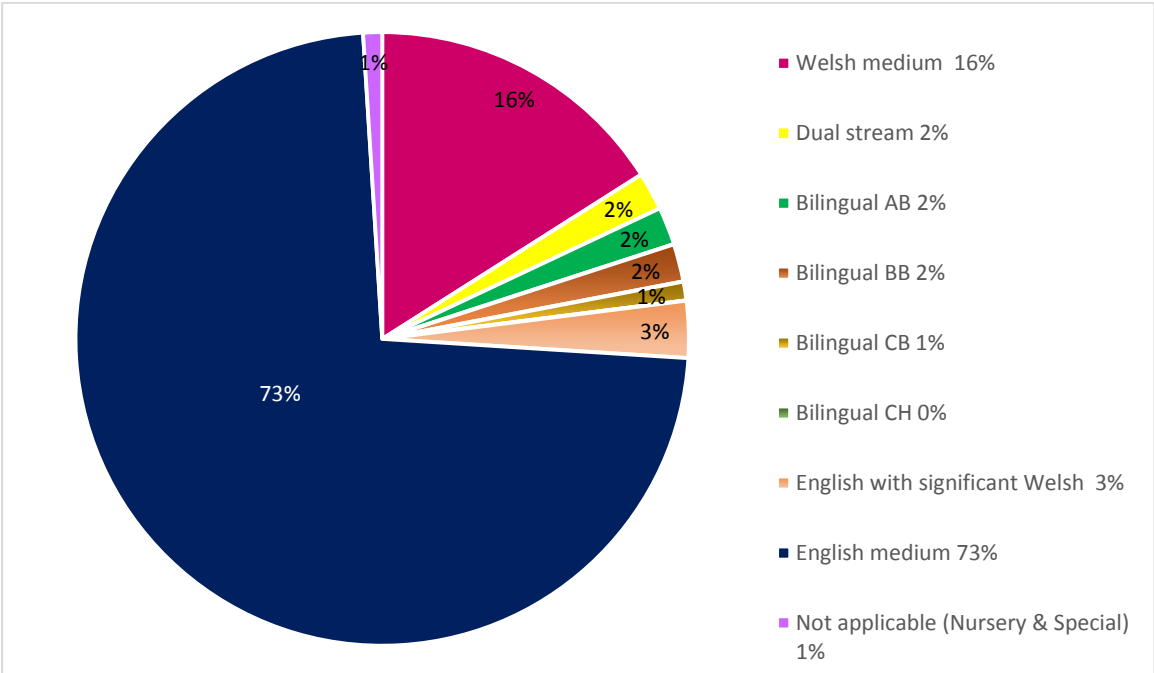
It is clear, from Figure 1, that the majority of pupils in Wales attend English-medium education, learning Welsh as a subject. Nevertheless, it is important to note that almost a quarter of the children in Wales are receiving their education either wholly or partly through the medium of Welsh (as categorised by Welsh-medium, Dual stream, Transitional and Bilingual schooling - categories AB, BB, CB, CH). The 2020 School census results (Welsh Government, 2020) indicate that from a total of 469,176 children attending local authority maintained schools across Wales, 108,638 pupils attended some level of Welsh-medium provision, therefore approximately 23% of the maintained schools' population are children being taught through the medium of Welsh. Moreover, of this figure approximately 5% of the children have

reported Speech, language and communication difficulties, with just over 1% having a diagnosis of Dyslexia. These figures relate to children with a formal diagnosis and so there is the potential for underrepresentation of children who are experiencing difficulties with literacy in schools but have not yet been formally diagnosed with an impairment. Although these figures are low, there is still a responsibility faced by educators and professionals working with children in Welsh-medium schools to ensure pupils receive equal and appropriate support when experiencing any literacy difficulties. This support begins with assessment. However, Welsh-medium assessment tools are scarce within Wales as is research into how best to support bilingual children in schools in Wales. It is therefore essential that resources are developed to help support those children who may find language learning difficult and also for those children within Welsh-medium education who experience difficulties with literacy that may potentially be masked by the transparency of the Welsh language (discussed further in the next chapter). Current steps are being taken by the Welsh Government to address this issue of Welsh language resources for schools but further research is required to develop appropriate assessment tools for Welsh-educated children with literacy difficulties. This thesis will contribute to that research base through exploring potential markers of literacy difficulties in Welsh, the results of which may contribute to the development of a screening test that can be used by educators and other professionals working with children in Wales.

Language and literacy within the Welsh Curriculum

Language and literacy have always played a key role within the curriculum, providing both access to knowledge, and opportunities to express developing competence, and to develop excellent communication skills that set children up for the future. In an attempt to raise standards of education in Wales the National Literacy and Numeracy Framework (LNF) (Welsh Government, 2013a) was introduced in schools in Wales in 2013 for children aged 5-14 years (Key Stage 1-3). The Framework is intended to help develop a clearer,

cross-curricular approach to teaching, ensuring that literacy and numeracy are incorporated into all subjects, moving away from being considered as stand-alone subjects. The LNF is split into 3 strands - Oracy, Reading and Writing - and targets/expectations are set in both English and Welsh. The strands are subdivided as follows[5]:

Oracy

- Speaking – communicate ideas and information to a wide range of audiences and a variety of situations.
- Listening – listen and respond to the viewpoints and ideas of others.
- Collaboration and discussion – contribute to discussions and presentations – discuss the viewpoints/ideas of others to reach agreement.

Reading

- Reading strategies – use a range of appropriate reading strategies to make sense of texts – use strategies to improve the fluency of reading – assess quality and reliability of texts. Responding to what has been read is subdivided into these two aspects:
- Comprehension – gain an understanding of unfamiliar information – identify main ideas, events and supporting details – predict, make inferences, understand layers of meaning – make connections within/across a range of texts/themes – carry out research to develop a full understanding.
- Response and analysis – organise and analyse relevant information – distinguish between facts, theories and opinions – compare a range of views – evaluate the content, presentation and reliability of texts.

Writing

- Meaning, purposes and readers – plan and adapt writing style to suit the audience and purpose – improve writing through independent

---

[5] Taken directly from the LNF

review and redrafting – write to ensure full coverage of a topic – improve the presentation of the writing (by including the use of ICT) – reflect, edit and redraft to improve writing.

- Structure and organisation – use a structure that is appropriate to the purpose and focus of the task – select analyse and present information appropriately – establish a structure to organise writing. Writing accurately is subdivided into these two aspects:

- Language – use language appropriate to writing – use appropriate technical terms and vocabulary/language appropriate to the subject-specific context.

- Grammar, punctuation, spelling, handwriting – specific statements relating to the progression of each aspect depending on the type of writing and the age level.

Skills expectations across both English and Welsh are fairly consistent within the framework, with the exception of particular Welsh specific grammatical rules such as mutation (see below). The intention is that children will develop the same literacy skills regardless of which type of school they attend (Welsh-medium, bilingual or English-medium); however, in Welsh-medium schools only the Welsh aspects of the LNF need to be considered from the Early Years as English is not formally introduced until the end of the Foundation Phase. The Foundation Phase Framework (FP) for 3- to 7-year-olds in Wales includes an area of learning entitled 'Language, Literacy and Communication Skills' which was revised to ensure it was strongly linked to the LNF. This is also echoed within the Welsh and English programmes of study within the National Curriculum, which is also structured using the same strands and sub-strands as those found within the LNF, emphasising the importance placed on the LNF. Within the FP a Welsh language area of learning is included to ensure it is fully incorporated into schools across Wales, even if the main language of instruction is English. Welsh-medium schools do not refer to this area of learning within their curriculum, they only refer to the language, literacy and communication strand which includes particular Welsh language-related statements.

One grammatical feature that is exclusively related to Welsh, and highlighted in various parts within the curriculum in Wales, is mutation. Within the LNF (and also the FP) 'Welsh-medium statements' are listed under particular language strands, particularly Oracy and Writing, which refer to elements that are only applicable to the Welsh language. There are 3 types of mutation in Welsh – soft, nasal and aspirate – all of which cause a morpho-phonological change in the initial consonant of words following particular syntactic 'triggers'. (For more information about mutation see Ball & Müller, 1992; Thomas & Gathercole, 2007; Thomas & Mayr, 2010. See also Chapter 3 of this thesis). Children in Wales are exposed to this grammatical process from a very young age, and formally so from as young as Nursery within the curriculum in Wales, where it is expected that the concept is introduced through rhymes and songs (Oracy strand of LNF and FP). They are also expected to show an awareness of mutation in their writing from Reception - 'Welsh-medium statement': show awareness that some sounds change at the beginning of words, e.g. *dau gi* 'two dogs' < *ci* 'dog', *y gath* 'the cat' < *cath* 'cat' (Writing strand p.26 of Foundation Phase Framework) (see Chapter 3 of this thesis for more examples of how mutation occurs in Welsh). This continues into the National curriculum Welsh programme of study and Welsh L2 programme of study where again mutation is incorporated into both the oracy and writing strands of the curriculum. Given that mutation is one of the more challenging elements of the Welsh language it will be further explored within the current research in order to ascertain its role in relation to identifying children with literacy difficulties. This will be discussed in more detail later in this thesis.

Welsh and Additional Learning Needs (ALN) within the curriculum

The Additional Learning Needs and Education Tribunal Act 2018 has three overarching aims (https://gov.wales/additional-learning-needs-and-education-tribunal-wales-act-explanatory-memorandum):

1. a unified legislative framework to support all children of compulsory school age or below with ALN, and young people with ALN in school or further education;
2. an integrated, collaborative process of assessment, planning and monitoring which facilitates early, timely and effective interventions; and
3. a fair and transparent system for providing information and advice, and for resolving concerns and appeals.

Eleven core objectives have been established to help achieve these aims one of which is 'a bilingual system' that focuses on not only ensuring provision is available to ALN children in both English and Welsh but on a requirement that Welsh must be stated on their Individual Development Plan (IDP). The aim is to have a bilingual ALN system to ensure education is accessible to all. The act also explicitly refers to children with an additional language stating that

> A person does not have a learning difficulty or disability solely because the language (or form of language) in which he or she is or will be taught is different from a language (or form of language) which is or has been used at home. (p. 4)

This highlights the importance of appropriate assessment of children's language and literacy abilities in order to determine whether any difficulties they are experiencing are due to general literacy difficulties or a lack of understanding of the language. The next chapter will discuss this issue in more depth. Also, in order to ensure education is accessible to all it is important to have access to necessary resources and tools to help support children with ALN.

Although there are no clear explicit plans to help support children with literacy difficulties, promising steps appear to be being taken by the Welsh Government to develop resources to help children with ALN such as commissioning research for the development of screening, assessment and diagnostic tools. As the act will be prioritising younger children there is a need for more assessment tools that can be used with younger children to enable

practitioners to formally identify children with additional learning needs from an earlier age in order to offer them the necessary support at an earlier stage in their education. As the research in this thesis includes participants from within the FP it will provide some guidance regarding whether the potential markers explored here would be suitable for use within a screening test for younger children.

The Welsh Government has expressed a commitment to providing resources to support the language development of children in Wales and also the use of research to inform pedagogy. What remains uncertain within the Cymraeg 2050 strategy is *how* these resources will be developed and what kind of support will be provided for children with literacy difficulties who attend a Welsh-medium school, where diagnostic materials are currently unavailable in Welsh. There is an intention to ensure that those with ALN are not disadvantaged as the proportion of Welsh teaching is increased over the coming years. This will be actioned through

> …Work with partners to understand what resources are available and are being used to support learners with ALN in the classroom, and commission research to identify the need for Welsh language screening, assessment and diagnostic tools (Welsh Government, 2017C, p.31,).

This is very encouraging in terms of resource development. The results within this thesis will help provide some insight into what further research will need to be done in order to help develop diagnostic tools for assessing Welsh-speaking children with literacy difficulties. This is very timely given that there is a

> Long-term strategic planning of school places to increase the number of learners in Welsh-medium or bilingual education, including increasing the proportion of Welsh-medium teaching and learning in schools. (p. 18).

If this is achieved there will be more Welsh speakers, leading to more children potentially attending Welsh-medium education therefore increasing the need for better assessment tools to help support Welsh-speaking children.

Within the LNF some consideration is given for Additional Learning Needs (ALN) in relation to those children who have a confirmed diagnosis/statement; however, in order to obtain diagnosis and therefore additional support, children must be assessed using the relevant standardised assessment tools. Assessment tools that are currently available include the *Profion Glannau Menai* (PGM) (Payne, 1998) which assesses children's reading and spelling skills. Although developed using similar structures to those adopted for the *Neal Analysis of Reading Assessment* (NARA) (Neale, 1997) suite of tools, the over-use of the same tool with the same children at various points in time means that the content is becoming well known by pupils and their performance on the task may not be an accurate reflection of their abilities. When children attending Welsh-medium schools are experiencing literacy difficulties, this lack of appropriate assessment tools to help formally identify those literacy difficulties makes it difficult for teachers to provide the necessary support to ensure they are developing their literacy skills. The LNF provides a framework for monitoring children's literacy progress but no clear guidance for supporting Welsh children who are experiencing difficulties. Literacy plays a very important and significant role within the curriculum in Wales and therefore developing assessment tools and resources that can help support any children who have literacy difficulties is essential to enable them to reach their potential.

Development of a new curriculum for Wales

The Welsh Government is currently developing a new curriculum for Wales for 3-16 year olds which is intended to be rolled out in 2022. The new curriculum aims to take a holistic approach to learning and is based on four key principles that are intended to be interwoven within all six areas of learning, all of which are designed to be cross-curricular, not seen as stand-alone subjects. The six areas are:

Expressive Arts
Health and Well-being

Humanities

Languages, Literacy and Communication

Mathematics and Numeracy

Science and Technology

The development and delivery of the new curriculum will be school specific and based on the needs of learners. The new curriculum will be viewed as a guidance document enabling more flexibility and creativity for educators. High importance continues to be placed on Literacy; the LNF will continue to be available following a review to ensure consistency with progression within the new curriculum. Welsh will be expected to be taught to all children 3-16 years of age attending funded non-maintained settings with the aim of offering all children the opportunity to develop bilingually. There is a clear emphasis on incorporating Welsh into everyday learning and the principles of the *Siarter Iaith Gymraeg* (Welsh Language Charter - a framework for schools to help support their Welsh learners and to promote positive attitudes towards the language to increase its use both inside and outside school[6]) are in line with the requirements set out within the new curriculum. This again emphasises the focus in Wales on developing the language and as such there is a requirement to provide relevant support to those who have difficulties accessing the language.

Summary

It is clear that the Welsh language plays a significant role within Wales' education system and that the Welsh Government sees Welsh as an area of growth. Strategies are being developed by the Welsh Government to address the ambition of increasing the number of Welsh speakers with the main focus being on schools in Wales to help deliver this. Literacy plays a crucial part in children's education and, in Wales, the added dimension of bilingualism raises

---

[6] For an evaluation of the Welsh Language Charter, see here: https://gov.wales/evaluation-welsh-language-charter?_ga=2.118296777.113077020.1594285772-407666520.1579193870 retrieved 16.2.21

some added complexities due to the diverse range of bilinguals attending education settings in Wales. What is critical in supporting pupils' development of Welsh is the development of assessment and diagnostic tools to help support children who may be experiencing difficulties. Being able to determine the difference between a child who is experiencing predictable, but non-permanent difficulties with learning to read and write Welsh compared with a child who has a particular type of literacy difficulty is essential to avoid under/over representation of a language difficulty. However, in order to do this, we need to have a clear sense of how literacy 'works' in Welsh (see Chapter 2 for a discussion around orthographic depth) and the types of language specific elements that may well be affected.

The next chapter will address the issues related to assessing bilinguals and the possible solutions to help overcome these issues.

CHAPTER 2

ASSESSING BILINGUALS

As presented in the previous chapter, children in Wales access education through the medium of Welsh and/or English to varying degrees, but they all access both languages within the curriculum. Children attending Welsh-medium schools are not formally introduced to English literacy until they reach the end of the Foundation Phase at age 7, and will continue to be exposed to both languages to various degrees depending on the area of Wales within which the child lives. Teachers in Wales are therefore faced with a variety of bilingual pupils within their classroom, each with varying levels of understanding of Welsh.[7] Some pupils may be L1 Welsh speakers from Welsh-speaking homes who receive a high frequency of exposure to the language. Others may have only one parent who speaks Welsh leading to a reduced input in Welsh that is shared with another language (usually English). Yet others will come from English-speaking families or from families that speak a language other than English or Welsh, and their primary source of exposure to Welsh will be within school. The relatively new term 'new speaker' is used to describe the concept where a learner is learning a new language that they are not exposed to within their home or community. In Wales the term '*dysgwyr*' (learner) is more widely used (Selleck, 2018), but some research has shown the negative attitudes displayed in some communities to children who are learning Welsh as a second language, viewing them as 'less Welsh' than L1 speakers (see Roberts, 2009). For the purpose of the present thesis, L1 and L2 are adopted throughout, in keeping with the terms used in the supporting literature.

This variation in language background poses an interesting challenge for educators in Wales. However, given that 'the teacher is the cornerstone of

---

[7] At the early stages, L1 Welsh pupils attending Welsh-medium schools are likely to have lower levels of English skills than L1 English pupils. However, L1, L2 and 2L1 Welsh-English bilinguals tend to converge on their English language skills around age 7 (Gathercole & Thomas, 2009) and often perform within respective age norms for English monolinguals on vocabulary and reading from age 7 onwards (Rhys & Thomas, 2013).

bilingual provision' (Jones and Lewis, 2014, p.170) it falls to them, within wider consultation with the School's and the Local Education Authority's language strategy plans, to be able to adapt their teaching styles according to the needs of the children. Therein lies the difficulty. In Wales, as in many other regions where education is delivered through a minority language, teaching and assessment resources are often limited in Welsh and access to support for children who exhibit any kind of literacy difficulties is challenging without diagnostic evidence. Given the lack of assessment tools available it is very difficult for teachers to be able to formally diagnose a literacy difficulty, such as dyslexia, and so they must wait until the child has produced ample pieces of work in English, and has developed enough native-like English skills to access the English diagnostic measures. Only then can practitioners provide a formal diagnosis which can lead to additional educational support for the child. At the same time, given that many pupils are only engaged in writing at school in Welsh up until age 7, early indicators of literacy difficulties that may or may not lead to a formal diagnosis is difficult. This may be based on the lack of knowledge of the writing profiles of pupils in Welsh, on the fact that Welsh has a transparent orthography that may mask certain types of underlying difficulties, and on the lack of diagnostic tools that are suitable for bilinguals that focus in part on their Welsh. Bilinguals are different from monolinguals, and different bilinguals are different from one another based both on their language experiences and on the nature of the two languages they know. In order to support teachers in identifying early on the potential literacy difficulties among Welsh-English bilinguals in Welsh-medium contexts, we need to understand more about the needs of bilingual learners, in this case around literacy, and how these needs should be reflected in testing.

This chapter provides a brief overview of the similarities and differences between monolingual and bilingual learners. It highlights some of the language assessment tools currently being used and developed for clinical purposes that are also useful within the educational domain. The chapter then provides a review of the literature pertaining to the current issues related to assessing bilinguals and the potential solutions to overcome those problems. The chapter ends with a brief summary of the findings within the literature

and implications of those findings for educators in Wales (and bilingual educators in general) and within the context of literacy in particular.

Monolingual and bilinguals – similarities and differences

As noted in Chapter 1, the majority of the world's population is multilingual. Yet, much of our thinking about language, particularly in relation to our thinking about ultimate achievements, seem very much rooted in monolingual ideals. A number of studies have compared monolinguals and bilinguals on various aspects of language, most notably looking at their vocabulary development but also on aspects of grammar. These have identified certain differences in their profiles, as outlined below.

Vocabulary

Current research in the field of bilingualism has focused widely on assessment and the notion of how best to assess children who are attending bilingual schools. This research fundamentally compares the language development of both monolinguals and bilinguals and the similarities and differences that are displayed in relation to language acquisition. Of particular focus has been the acquisition of vocabulary. Much of the research looking at bilingual and monolingual vocabulary has highlighted differences in speakers' vocabulary knowledge. These differences are often interpreted within the broader concept of a so-called bilingualism deficit hypothesis (Junker & Stockman, 2002), which refers to the notion that the learning of two languages concurrently from an early age (simultaneous bilingualism) has a negative effect on, or slows down children's development in one or both of their languages. This notion is often (mis)interpreted as a permanent condition of bilingualism, particularly when studies demonstrate differences in vocabulary knowledge that seem to last the whole lifetime. For example, Bialystok and Luk (2012) examined English receptive vocabulary scores of monolingual and bilingual adults between the ages of 17 and 89 years and found that the mean score

differed significantly with monolinguals performing better than bilinguals. This result was found despite all bilinguals in the sample admitting to using both languages regularly, for many years, on a daily basis. Elsewhere, clear differences in rates of vocabulary development have been demonstrated among children in numerous studies (for example Gross, Buac & Kaushanskaya, 2014; Klassert, Gagarina, & Kauschke, 2014; Montanari, Abel, Graßer & Tschudinovski, 2018). Whilst these differences often relate to smaller vocabularies among bilinguals as compared to monolinguals, there are obvious reasons why these differences exist and why the monolingual 'norm' is not a useful marker for comparison. These are outlined below.

(i) Distributed Characteristic (Oller, 2005)

The first explanation relates to the 'distributed characteristic' of the bilingual (Oller, 2005). According to the Distributed Characteristic Hypothesis bilingual children are perceived to be slower learners and have vocabulary deficiencies in each of the languages when compared with monolingual children because 'lexicalised concepts of the bilingual are 'distributed' across two languages such that some concepts are lexicalized in one language but not the other and vice versa' (p.1744). This can give the impression – to parents, teachers and/or the learner - that bilingual learners have a perceived learning difficulty in relation to aspects of language when it may only be a natural result of bilingual learning. This poses an issue for the assessment of bilinguals in relation to monolingual norms because a single-language assessment would not tap into the lexicon of each of the bilingual's two languages (Abudarham, 1997) and therefore would not provide a full picture of the child's abilities. Assessment norming is discussed later in the chapter.

(ii) Complementarity Principle (Grosjean, 2008)

A second explanation for the bilingual vocabulary lag has been proposed in the form of the 'Complementarity Principle' (Grosjean, 2008), which refers to 'the differing needs of the bilingual for the two (or more) languages' (p.22). The principle relates to how often a bilingual may need to call on both languages within different scenarios in life and the notion that this impacts

the level of fluency within the language, including the types of vocabulary items they may utilise in any given situation in any one language. For example, if a particular concept is discussed in contexts where speakers tend to use Language A, then the bilingual will call upon and develop their vocabulary knowledge in Language A to facilitate conversations around that concept. Consequently, terms that relate to that context may never or only slowly be developed in Language B, hence the lag.

However, despite numerous studies revealing differences in rates of vocabulary development across bilinguals and monolinguals, other studies refute these claims, finding comparable results across monolinguals and bilinguals alike. For example, Junker and Stockman's (2002) study of German-English bilingual toddlers, in comparison to their German and English monolingual peers, revealed that learning a language simultaneously did not appear to cause a vocabulary lag. When their conceptual vocabulary and verb diversity in both languages was pooled, results indicated that bilingual children were not 'inferior' when compared with their monolingual counterparts within the study. Pearson, Fernandez and Oller (1993) reported similar trends in their earlier study. In their semi-longitudinal study of simultaneous English-Spanish bilinguals and a monolingual comparison group they found no statistical evidence to support the notion that bilingual children are slower at developing early vocabulary than monolingual children. They concluded that assessment norms for lexical development in bilinguals should take into consideration the children's performance in both languages. Similar patterns have also been found by Thordardottir (2011), Thordardottir *et al.,* (2006), Junker & Stockman (2002), and Core *et al.,* (2013). It seems, therefore, that although on the surface, bilinguals may appear to be lagging behind their monolingual peers in areas such as vocabulary knowledge when results in any one language is compared to monolingual peers, careful assessment of a child's abilities in both languages can reveal that bilingual children's vocabulary knowledge matches or even exceeds that of their monolingual peers, particularly in cases where a bilingual makes frequent use of their two languages. For many, the distributed characteristic and the complementarity principle that guides their engagement with each language will lead to early

lags in vocabulary, but, with increased engagement with both languages in various domains over time, this early lag can be reduced or even eliminated. This emphasises therefore the need for assessment and profiling in both of a bilingual child's language knowledge in order to gain a clear picture of their abilities.

(iii)   Frequency of Language Exposure

A third explanation relates to language exposure. Many studies have found there to be a clear relationship between a child's exposure to a language and their level of performance in that language, particularly in the earlier years (Hoff, 2003; Hart & Risley, 1995). It has also been found that the higher the levels of exposure the better the linguistic output (Paradis & Genesee, 1996). Monolinguals would be assumed to have a greater exposure to a single language than bilinguals, although this is not always the case, for example, De Houwer's study of Dutch-French bilingual and Dutch monolingual children examined the influence of maternal language input from an early age finding no evidence of reduced language input for bilingual children (see De Houwer, 2014). Given that children attending bilingual education will be exposed to two languages to varying degrees it is important to consider and be aware of the influence of language exposure on those children and how this will affect their language development (e.g. Hammer *et al.*, 2014; Thordardotir 2017; Unsworth, 2016). The effects of language exposure has been a key area of focus within bilingual research over the past few years, and the degree to which different levels of exposure impact on children's language acquisition.

Numerous studies have supported the notion that children who have higher exposure to a language experience greater success with particular linguistic forms within that language (e.g. Hart & Risley 1995; Oller & Eilers 2002; Gathercole 2002a, 2002b, 2002c) and that a 'critical mass of exposure' (see e.g. Marchman & Bates 1994; Maratsos 2000) is required to enable an individual to gain a native like command of those linguistic forms (Oller & Eilers 2002; Gathercole 2007). This notion of gaining a critical mass of exposure has been supported in a number of studies of vocabulary (e.g. Cobo-Lewis *et al*., 2002a, 2002b) and certain elements of morphosyntax (e.g.

Gathercole 2002c), although the exact amount of exposure required is yet undetermined within these types of studies. In terms of vocabulary, this critical mass is formulated in terms of item level exposure in the sense that a child will only know a word if they are exposed to it. Once they know a word, they may be able to create new words or understand other words that share commonalities, e.g. words with shared stems.

The importance of taking exposure to language into account is highlighted in a study by Gathercole *et al*. (2013) who looked at vocabulary and grammatical abilities in bilingual Welsh-English children. For the vocabulary element of the research receptive vocabulary tests were administered in Welsh using words gathered from the development of the *Prawf Geirfa Cymraeg* (Gathercole & Thomas, 2007) and in English using the *British Picture Vocabulary Scales* (BPVS) (Dunn *et al*., 1997) with children aged 2-15 years. Both tests were administered to bilingual children whilst only the BPVS was administered to the English monolingual children. Results showed that language exposure played a significant role in vocabulary performance. Those children who had more exposure to a language (English or Welsh) advanced earlier with their vocabularies than those who had less exposure. As children gained more experience with both languages, their performance on the tests became more similar, indicating the importance of continuous exposure to a language on children's performance on vocabulary tests (cf. the concept of 'cumulative exposure over time' as expressed by Unsworth, 2013).

Vocabulary Summary

Bilinguals essentially need to learn two sets of vocabularies and so their rates of vocabulary learning will differ from that of monolinguals, although their conceptual vocabulary will often surpass or at least match that of a monolingual. The important issue to note is that although a bilingual child may appear to have a lower vocabulary knowledge in one language when compared with their monolingual peers it does not necessarily imply a language

deficiency. Testing in both languages is the only way to establish an accurate picture of a bilingual child's vocabulary knowledge.

Grammar

Similar types of differences have been found in relation to bilinguals' acquisition of grammar. As with vocabulary, a critical mass of exposure to a language also relates to grammar. For grammar, this means having enough examples of different types of structures to be able to start formulating the basic building blocks of language underpinning the specific rules and patterns that relate to that language. Performance has been found to level out as children get older and have gained sufficient knowledge and experience with the language in order to be able to abstract out the regularities from within the system (e.g. Bahrick *et al.*, 1994; Oller, Pearson & Cobo-Lewis, 2007; Hammer, Miccio & Rodriguez, 2004; Umbel, Pearson, Fernandez & Oller, 1992; Kovelman, Baker & Petitto, 2008; Oller & Eilers, 2002).

In the Gathercole *et al*. (2013) study of Welsh-English children's grammatical (and vocabulary as highlighted in the previous section) abilities, receptive grammatical tests were created in both languages to assess grammatical abilities using comparable underlying grammatical structures for both languages. Thirteen sets of structures were used within the tests including active sentences, negation, passive (truncated), comparative, superlative, present perfect, future, time conjunctions (*before/after/*until), relative clauses and quantification (see study for more in-depth list). Two versions of the tests were created to ensure children would not be given the same picture or translation-equivalent sentences across the two languages. The sample consisted of 376 children aged 2-15 years, including monolinguals and bilinguals from four different home language groups (monolingual English, bilinguals with only English at home, bilinguals with both Welsh and English at home and bilinguals with only Welsh at home). Results showed that as with vocabulary, the pace at which children develop language is directly linked to

the amount of exposure to each language, as measured in this case by exposure practices in the home.

Binks and Thomas (2019) examined the effects of limited input on bilingual teenagers' acquisition of complex morphology in Welsh, focusing on receptive knowledge of grammatical gender and production of plural morphology. The study compared 168 teenagers aged 12 to 13 and 16 to 17 years from varying Welsh-English backgrounds (L1 Welsh, simultaneous Welsh-English bilinguals referred to as 2L1 in the study, and L2 Welsh) and found that simultaneous bilinguals progressed at a slower rate than their L1 Welsh-speaking peers. With plural morphology, the L1 groups performed at a similar rate to each other whereas there was a slower progression among the L2 Welsh participants compared with their L1 peers on both grammatical gender and plural morphology.

In their study of 732 Welsh-English bilingual children and adults aged 3 to over 60 years, Gathercole, Kennedy and Thomas (2016) examined children's performance on Welsh and English vocabulary, grammar and cognitive measures (McCarthy Scales of children's Abilities, 1972, and Ravens progressive matrices, 1998). Results revealed that exposure to language was generally a significant predictor of performance on vocabulary and grammar tasks in the target language. Interestingly, home language exposure was the highest predictor of performance on the vocabulary measures while home language exposure appeared to be the most influential factor amongst younger children on the grammar tasks. These factors were mediated by socio-economic status measured by parental education and profession, which was also found to play a significant role in performance on the language and cognitive measures but particularly so on the language measures.

Another layer of complexity with regards to grammar is that the grammatical properties of a bilingual's two languages are sometimes similar, sometimes not. In some cases, these differences or similarities work in favour of the bilingual leading to positive cross-linguistic influences, and in other cases not.

Grammar Summary

Level of exposure to a language is clearly a significant factor influencing the ultimate achievement of bilinguals in their two languages. Exposure to language in the home, community and education setting needs to be taken into consideration in order to gain an accurate reflection of the level of exposure. Given than bilinguals share their waking hours across two languages, and not only have to learn two vocabularies, but also two sets of (sometimes complimentary, sometimes competing) grammars, it is no surprise that bilinguals differ from monolinguals when tested for their language abilities. However, it is also the case that different types of bilinguals can differ from one another in various ways, particularly in terms of rate of acquisition (see e.g. Gathercole & Thomas, 2009). This causes an additional level of complexity when aiming to test bilinguals for language ability.

Whilst vocabulary and grammar are clearly prominent within any curriculum, one aspect of language that has an all-pervasive role in education is literacy. The next section outlines some of the differences and similarities presented by bilinguals when compared with monolinguals within literacy based tasks.

Literacy

Literacy forms a significant part of any school-based curriculum and educators who work within bilingual settings are faced with the additional demands of developing pupils' biliteracy.

Rhys and Thomas (2013) explored the differences between Welsh-English bilinguals and English monolinguals' literacy development during Key Stage 2 (7-11 years old). Using receptive vocabulary and reading tasks the study revealed differences in the performance of bilinguals in relation to the English tasks, where L1 English bilinguals and monolinguals performed closer to age norms than the L1 Welsh group and simultaneous bilinguals. In fact, the simultaneous bilinguals and L1 Welsh bilinguals fell short of the monolingual English age matched norms on vocabulary, demonstrating the perils of using

single language tests (in this case English) as a measure of language performance among bilinguals. On the Welsh tasks results revealed higher levels of performance relative to increasing amounts of exposure. L1 Welsh bilinguals consistently performed better than the simultaneous bilinguals and L1 English children on the Welsh tasks within the sample. This finding echoed findings of previous studies by Hakuta and Diaz (1985), Pearson, Fernandez and Oller (1995), Pearson (1998) and Genesee (2001) in their studies on bilinguals' acquisition of vocabulary as well as Gathercole's (2002a, b, c) research of morphological development. Furthermore, Rhys and Thomas' study found significant relationships between vocabulary and reading comprehension, which supports the notion that better vocabulary skills relates to better literacy success. These findings are supported by previous studies by Duursma *et al.* (2007), Oller, Pearson and Cobo-Lewis (2007), Bialystok (2006) and Silverman *et al.* (2015) who also found vocabulary knowledge and reading comprehension to be linked.

These findings highlight the role of language exposure and the importance therefore of enabling bilingual learners to use both languages equally (where possible) within a bilingual education setting to ensure they can thrive in both languages.

Similarly, Oller and Eilers (2002) in their study of Spanish-English children's literacy development in Miami, Florida examined children from monolingual, immersion and two-way elementary bilingual education using a variety of standardized tests evaluating their reading, writing and oral skills. Results showed that the English monolingual children performed best on English language measures although results were comparable with those in the immersion and two-way elementary bilingual education groups. Oller, Pearson and Cobo-Lewis (2007) also reported comparable reading abilities across bilingual and monolinguals, even though lower vocabulary scores were recorded for the bilingual group. These findings are echoed by Rhys and Thomas (2013) in their study of Bilingual Welsh-English children's acquisition of vocabulary and reading.

Paradoxically, lower vocabulary performance within a classroom setting could lead to under-representation of pupils with literacy difficulties, because there may be an expectation that a bilingual child will have a smaller vocabulary than their monolingual peers in the target language (Simonsen, 2017). Assessing bilingual children's language abilities in both language is thus crucial in order to maximise the potential to pick up any irregularities in their linguistic behaviours, despite the fact that bilingual children with language impairment (LI) are likely to present problems in both languages (Caesar & Kohler, 2007, Thordardotir *et al*. 2006 as cited in Peña, Bedore & Kester 2016). Nevertheless, if bilingual children are only assessed in one language it is more difficult to ascertain whether the problems they present are due to a lack of exposure and understanding of the language of assessment, or due to a language impairment or a difficulty in language learning.

Literacy Summary

There are clear differences in the way bilingual children perform on literacy tasks when compared with their monolingual peers. There is a tendency for monolingual children to outperform their bilingual peers on tasks that involve the bilingual child's L2; however, some of these differences may level out as children get older such as vocabulary acquisition. However, if a child has underlying literacy difficulties this levelling out may not occur, which highlights the need for effective bilingual assessment at an early stage to ensure an accurate reflection of a child's abilities is being achieved.

Another layer of complexity with regard to literacy relates to orthographic depth. The next section discusses the influence of orthography on children's language and literacy learning.

Orthography

The written form of a language is constrained by a set of rules and conventions, namely its orthography. There are two main types of orthographies – transparent (or 'shallow') and opaque (or 'deep'). Welsh is regarded as having a transparent orthography, in which the grapheme-phoneme correspondence is relatively consistent. English, on the other hand, is said to have an opaque orthography, in which the grapheme-phoneme correspondence is not as consistent. In terms of how orthographic depth impacts on children's acquisition of reading, children learning to read in a language with an opaque orthography, as is the case for English, rely not only on phonological awareness in terms of being able to decode the word, but also semantic knowledge, where an individual will use recognisable and known parts of a word to piece together the word. The 'dual-route' theory such as Coltheart's Dual Route Cascaded Model (Coltheart *et al.,* 2001) has been used to describe the process of using both a phoneme system and semantic system to read words. A child learning to read in a language with more opaque orthography, such as in English, would typically not be able to rely solely on a grapheme-phoneme conversion (GPC) approach due to the irregularities inherent to its orthography. A child learning to read in a language such as Welsh, on the other hand, with a more transparent orthography exhibiting a higher frequency of one-to-one letter to sound correspondence than in a language like English, would be less reliant on a semantic system for reading words. In this case, the GPC route would be a more effective model in learning to read in a language like Welsh as compared to English. Similar approaches, such as the connectionist triangle model (as mentioned in Young, 2014) claim that all information that a reader collects (whether phonological, orthographic or semantic) work together in order to read a given word and that no single route is faster than another (for reviews see Young, 2014). With this in mind, researchers have explored the influence of orthography on children's literacy skills.

The Orthographic Depth Hypothesis (Katz & Frost, 1992) refers to the relative ease with which a person can learn to read a transparent language when

compared with an opaque language, which can take longer to master. In opaque languages such as English, reading instruction includes a combination of phonics and sight word teaching, whereas a transparent language may be reliant on phonics instruction alone due to the consistent grapheme-phoneme correspondence. Learning to read in transparent languages such as Welsh has been found to be much quicker than learning an opaque language such as English (Hanley, Masterson, Spencer, & Evans, 2004; Spencer & Hanley, 2004). A study by Spencer and Hanley (2003) involving 74 Welsh-speaking children and 88 English-speaking children aged 5 to 6 years examined the relationship between development of reading skills and the influence of orthography. Results found that Welsh children significantly outperformed the English children on the reading of both real words and non-words. Performance on tasks of phonological awareness was also better amongst the Welsh group than the English group. Welsh speaking children performed significantly better on tasks related to phoneme counting, regardless of the language of testing and the number of phonemes/letters in the words. The study suggested that phonemic awareness test scores in a transparent orthography are less affected by knowledge of the spellings of the words than readers of more opaque orthographies.

Ellis and Hooper (2001) reported similar findings. Their study of 40 children (20 Welsh-speaking children and 20 English-speaking children) aged 6 to 7 years investigated the rate of literacy acquisition in a transparent (Welsh) versus an opaque (English) language through measuring children's reading aloud accuracy, latency, comprehension and reading errors. Welsh children were able to read aloud significantly more of their language than English children. In terms of the nature of errors made, English children's errors appeared to focus primarily on real word substitutions and null attempts whereas Welsh children's errors related more to non-word mispronunciations, indicating a reliance on alphabetic decoding. The study concluded that orthographic depth of a language can play a significant part in reading acquisition. Aro and Wimmer (2003) explored the reading performance of German-, Dutch-, Swedish-, French-, Spanish- and Finnish-speaking children in comparison with English-speaking children (all languages were at varying

points on the orthographic depth continuum). Results found that whilst Dutch-, Swedish-, French- and Finnish-speaking children read pseudowords with the same high accuracy as German and Spanish (as was also found in Wimmer & Goswami, 1994), this was not the case with the English children. Their results imply that learning to read in the language with the deepest orthography - English - was the most difficult.

Phonological deficit has been a key characteristic in the diagnosis of children with dyslexia (Ramus, 2004; Snowling, 2000; Sprenger-Charolles & Serniclaes, 2003). However, whether phonological awareness skills can be used equally as effectively as a measure of literacy difficulties among children learning a transparent language has had mixed reviews. Some support the hypothesis that phonological awareness is the best predictor of reading development in both transparent and opaque orthographies in typically developing children. For example, Caravolas *et al.'s* (2013) study presented longitudinal evidence of the growth of children's reading skills in English, Spanish and Czech, finding that although learning to read in English is slower and follows a different trajectory than more transparent languages, the importance of phonological awareness skills (measured by phoneme awareness and letter-sound knowledge) in predicting reading development did not differ according to the transparency of the language. Furnes and Samuelsson (2009) reported similar findings in their study of English-speaking and Scandinavian children, noting that the cognitive and language skills that underlie early reading development are similar regardless of the orthographic depth of a language, (see also Vaessen *et al*., 2010), although items should be made sufficiently difficult in the transparent language (Caravolas *et al*., 2005). Others suggest that relying on phonological awareness alone could lead to misinterpretation of a child's abilities. For example, Furnes and Samuelsson (2010) in their study of Australian and US same sex twin pairs and Scandinavian twin pairs aged between ages 4;10 and 8;9 years examined the predictors of reading and spelling difficulties among children who were learning a language with an opaque orthography (English) and a language with a relatively transparent orthography (Norwegian/Swedish). They found that phonological awareness tasks significantly predicted reading abilities in

English and also in Scandinavian for children in the first years of schooling. However, this did not apply to older children suggesting that this relationship diminishes as children get older. The study also found that the association between phonological awareness and spelling difficulties in transparent and opaque orthographies is similar. On the contrary, Ziegler *et al*. (2010) conducted a study involving 1,265 children at Grade 2 from differing L1 language backgrounds namely Finnish, Hungarian, Dutch, Portuguese and French .The study explored the role of phonological awareness, memory, vocabulary, rapid naming and nonverbal intelligence in reading performance across the five languages. Similar to Furnes and Samuelsson, findings of this study revealed phonological awareness to be the most prominent factor associated with reading ability across all languages; however, the weight of this influence varied according to the orthographic depth of the language, with the impact of phonological awareness being stronger in more transparent languages.

Referred to by Cummins (1979) as the Linguistic Interdependence Hypothesis, children who have any underlying literacy difficulties such as dyslexia will present these difficulties regardless of which language they are learning (see also Paulesu *et al*., 2001). The Script Dependent Hypothesis (Cummins 1979), however, predicts that deficiencies could be specific and relative to the structure of a language. Therefore difficulties may be displayed in one language but not another, or will be displayed in different ways, or to different degrees, in different languages, depending on where they fall on the orthographic depth continuum. Whilst this is a natural result of the varying nature of languages, where this becomes complex, particularly in relation to assessment, is when bilinguals are learning languages that are polar opposites on the depth continuum. Welsh-English bilinguals are a case in point. As noted previously within this section, some children are able to acquire languages with transparent orthographies (as measured by literacy abilities) with relative ease compared with children learning languages with opaque orthographies, and phonological abilities may or may not relate strongly with literacy abilities among those learning languages with transparent orthographies. The concern here is that some languages with highly transparent orthographies could

potentially 'mask' underlying difficulties if assessments are primarily focused on phonological awareness. In line with the Script Dependent Hypothesis, there is a need to explore whether there may be other markers, beyond the traditional phonological awareness tasks that were developed with English in mind, which could support effective assessment of literacy abilities in languages with transparent orthographies. From a classroom teacher perspective, this means knowing how children who have literacy difficulties typically manifest those difficulties on paper in Welsh. In order to do this, there is a need to consider the nuances of the language being assessed and explore the extent to which some of those nuances may affect TD and children with literacy difficulties in different ways. This was the focus of the present set of research studies as presented in this Thesis.

Following on from this need to explore other markers and means for assessing bilinguals the next section of this chapter presents some examples of current assessment tools and discussed the benefits and challenges of using these tools with bilinguals.

Bilingual Assessment Tools

A wide range of assessment tools are typically used to measure literacy abilities across a variety of languages. These include measures of vocabulary, grammar and reading, many of which are based on monolingual norms rather than bilingual or even multilingual norms (Gathercole, Thomas & Hughes, 2008) and are therefore not sufficient as stand-alone tests of a bilingual child's abilities (Thordardottir *et al*., 2006). Whilst reading and spelling tasks yield useful information regarding pupils' literacy abilities, diagnostic assessment measures examine these abilities within the wider context of their language abilities. In the present thesis, a series of language tests, comprising of both novel and standardised tasks, were used to measure various aspects of language abilities and related those abilities to pupils' scores on standardised reading tasks.

Testing a child in both languages, using a variety of methods, is essential in order to develop a holistic view of a child's language abilities. Some of the tests that are currently offered within assessment batteries, and some that are under development, are discussed below.

Non-word repetition tasks.

One type of task that is often included as a measure of language ability is the non-word repetition task. Non-word repetition tasks were originally developed by Gathercole and Baddeley (1989) as a measure of phonological short-term memory. Non-words are words that a child is unlikely to have been exposed to previously as they are, by definition, words that do not exist in any given vocabulary. However, they usually follow the phonotactic constraints of the target language, i.e. 'items that are consistent with the phonetics of the target language, being made up of consonants and vowels that occur in that language' (Chiat *et al*., 2013, p.60). Children's ability to recognise non-words is not based on their prior exposure to the target language but rather their short-term memory of the items used in testing. Non-word repetition tests are widely used by clinicians when exploring Developmental Language Disorder (DLD) and developmental dyslexia with mixed results (e.g. Van Ijzendoorn & Bus, 1994; Van den Broeck & Geudens, 2012). As mentioned in Chiat (2015) there is much empirical evidence to support the use of non-word repetition as a tool for identifying children with DLD. Studies have been conducted in a number of languages, including English (Coady & Evans, 2008; Gathercole, 2006; and Graf Estes *et al*., 2007), Spanish (Girbau & Schwartz, 2007), Italian (Casalini *et al*., 2007), Icelandic (Thordardottir, 2008), French (Thordardottir & Brandeker, 2013) and Swedish (Sahlén *et al*., 1999). All show consistent findings, namely that children with DLD have difficulties reading non-words (however, see Stokes *et al*.'s, 2006 study of Cantonese children for an exception to this).

With regards to bilinguals who have various exposure to each of their languages, a key benefit of non-word repetition tasks would be that they are

not reliant on the amount of exposure a child has had to a language but rather their phonological abilities, rendering the tasks suitable for a variety of types of speakers. That is to say that real word tasks, such as receptive or expressive vocabulary tasks, requires knowledge of a word and all its associated meanings. Since vocabulary is largely learned piecemeal (item-by-item, as they occur in the child's environment) input frequency would be a critical component of real word task performance. On non-word tasks on the other hand, since the critical task component would be focused around phonology and phonological memory, it could be argued that language exposure would have less of an impact on performance. However, studies have reported differences in bilinguals' performance, depending on the language in which the non-words have been created. For example, Messer *et al*. (2010) (as cited in Chiat 2015) in their study involving Turkish-Dutch bilinguals found that the children performed better on the Turkish test than on the Dutch test when compared with their monolingual peers, which suggests differences between the two groups' language experience could have an effect on performance. Consistent findings of language and/or group differences have also been reported within the Hispanic population in the USA. For example, Summers *et al*. (2010) (as cited in Chiat 2015) found that bilingual children performed significantly better in Spanish than English during their study on a group of typically developing Spanish-English children. In both studies, bilinguals seemed to perform least well on the most opaque language in terms of its orthography (Dutch in the first example; English in the second) suggesting that the structure of the language, and the typological distance between a bilingual's two languages, may have an effect on performance outcomes in different languages. However, Windsor *et al.* (2010) in their study with Spanish-English children, found that whilst monolingual children scored higher on an English non-word test when compared with bilinguals, typically developing bilinguals still performed better than bilinguals with Language Impairment, suggesting that non-word repetition tasks may yield useful information about phonological memory that can distinguish typical vs. atypical performance among bilinguals, particularly with regards to DLD.

However, these differences observed across languages may arise not only because of the known differences within languages, but due to methodological and/or other issues that impact the reliability of the results. These issues are discussed below.

First, it is important to consider the content of the tasks. The types of words used may vary quite considerably across languages, across studies and across diagnostic tools in relation to their likeness, prosodic structure, phonotactic sequences, phonological difficulty and whether or not they contain morphemes of the language used. These elements can all have a potential impact on children's performance on non-word repetition tasks (see Gathercole, 2006).

Second, Sharp and Gathercole (2013), in their study involving unique Welsh sounds, noted that children's performance was linked to language knowledge. This is consistent with the findings of Messer *et al*. (2010) and Summers *et al.* (2010) with regards to language experience playing a role in performance on non-word repetition tasks (see also Engel de Abreu, 2011; Sharp & Gathercole, 2013; Summers *et al.,* 2010).

Third, item length can influence performance on non-word repetition tasks (Chiat, 2015). As item length increases, performance tends to decline, which initially influenced the belief that these tasks measure short-term memory capacity. However, given the effects prosodic and segmental structure play a role in influencing performance, non-word repetition tasks became measures of phonological ability rather than just short-term memory, although memory does play a significant role in terms of retaining and processing of the items.

However, under controlled conditions with suitable samples, non-word repetition tasks can play a valuable role in assessing bilinguals and in identifying potential language impairment. This is particularly the case for bilinguals since performance is not reliant on previous vocabulary knowledge or length of exposure (although language knowledge may have an effect). However, the phonotactic structure of the languages spoken by the children and the resulting structure of the task items themselves in relation to their

similarity, prosody, phonology and morphosyntax may well be impacted by previous language knowledge and therefore enhanced by higher levels of exposure to a language.

As mentioned in Perez *et al.* (2013) when drawing conclusions about an individual's ability to differentiate between a real and invented or 'non-word' one must consider the nature of the non-words (James, 1975; Richardson, 1976; Rubenstein *et al*., 1971). When words that are created using only consonants, individuals need only look at the word to know if it is real or not. However, when non-words are similar to real words it is fair to assume that lexical access is required to be able to say the word aloud (James, 1975; Richardson, 1976 as cited in Perez *et al*., 2013).

Non-words by their nature do not contain real words which contrasts with the sentence repetition tasks as these are tasks containing a variety of different sentences of varying length using real words. The processing systems used here are therefore more complex as children not only need to repeat a word but need to be able to understand the sentence in order to reproduce it. Sentence repetition tests are therefore considered next and reviewed according to their usefulness in helping to assess bilinguals and identify language impairment.

Sentence Repetition (SRep) Tests.

Sentence repetition tasks are often included in assessment toolkits as they are quick and relatively easy to use, include a wide range of sentence types, have clear target sentences and can be scored in a variety of ways depending on what is to be the focus of the analysis (Seeff-Gabriel *et al*., 2010).

Sentence repetition tasks involve an individual listening to a prepared sentence and repeating it out loud. One example of a sentence repetition task is the LITMUS-SRep test (Marinis & Armon-Lotem, 2015), which uses 'structural scoring' to identify '..if participants were able to use the specific structure irrespective of whether or not they repeated the sentence verbatim'

(Marinis & Armon-Lotem, 2015, p.118). They are designed using relatively long sentences that avoid the use of passive memory when repeating them. By having longer sentences, individuals need to use their knowledge of the grammatical structure of the language to be able to process and analyse the sentence in terms of its phonological, morphosyntactic and semantic nature in order to reproduce what they have heard from 'activated representations in long-term memory' (Lombardi & Potter, 1992; Potter & Lombardi, 1990, 1998, as cited in Marinis & Armon-Lotum, 2015, p.9892). Consequently, accuracy, in a sentence repetition test, relies on these processes working together. It is worth noting however, when using sentence repetition tasks, that sentences should not be too long as participants may find them too challenging, exceeding the participant's memory capacity and therefore result in a floor effect (Marinis & Armon-Lotum, 2015).

As noted in relation to non-word repetition tasks above, language ability is key to a child's performance on sentence repetition tasks, which makes the task particularly suitable when needing to differentiate between typical development and language impairment. Polišenská's (2011) study is illustrative of this point. She tested typically developing Czech and English children aged 4 to 6 years and looked at the influence of different types of language knowledge on their short-term memory span for sequences of words. Sentence repetition was found to draw particularly well on knowledge of syntax and morphosyntax at the sentence level, but was also deemed useful for all aspects of sentence knowledge (Chiat *et al*., 2013). Should a child's processing abilities and knowledge of grammatical structure have not been adequately developed at the time of testing, they would be unlikely to perform at the normed level suggesting a possible presence of language impairment (Marinis & Armon-Lotem, 2015). In English, morphosyntax difficulties, in particular with regards to function words, inflections, and grammatical morphemes, are taken as indicators of DLD (Leonard, 1998 as cited in Chiat *et al*., 2013) and these difficulties can be assessed using sentence repetition (Chiat & Roy, 2007; Seeff-Gabriel *et al*., 2008, 2010)**.**

The Clinical Evaluation of Language Fundamentals test (CELF; Semel *et al*., 1994; Wiig *et al*., 1992) is one example of a widely-used sentence repetition test that consists of a variety of sentences of differing length, content and complexity. Conti-Ramsden *et al*. (2001) used the sentence recall subtest of the CELF within their studies and found a significant difference in performance between typically developing and DLD children on sentence repetition tests, with 90% of children (age 11 years within the study) with language impairment scoring below the cut-off point. Other studies have also reported similar findings using different forms of sentence recall tests with differences being displayed between DLD, typically developing and other clinical groups (Redmond, 2005; Redmond *et al*., 2011; Riches *et al*., 2010; Willis & Gathercole, 2001). Similar findings have also been reported in Cantonese (Stokes *et al*., 2006). It is worth noting however that Conti-Ramsden *et al.* also found that 15% of TD children also scored below cut-off which suggests that the CELF may not be suitable for use on its own as a measure of DLD, other information would be required to form an accurate conclusion of a child's abilities.

Although the CELF has been found to be successful at distinguishing between DLD and TD children, the research in relation to the use of sentence repetition tasks with bilinguals is still developing. For example a study by Chiat *et al*. (2013) of Turkish-English sequential bilinguals (and age-matched L1 English) using the CELF in English, investigated the effects of Age of Onset (AoO) and Length of Exposure (LoE) on performance on sentence repetition tasks. Children were tested against monolingual norms and findings revealed that AoO and LoE had no effect on their performance on the CELF but sentence length and complexity did, with better performance observed when using shorter and simple sentences. However, their study (2013) with Russian-Hebrew and English-Hebrew children revealed similar findings in that they fell within monolingual norms or even performed better than their monolingual peers on sentence repetition tests. In line with the findings around the use of non-word tests, such findings suggest that AoO and LoE potentially do not affect performance on sentence repetition tasks (although they should be

taken into consideration), which is a key concern when testing bi- or multilinguals.

Since all bilinguals fell within the monolingual norms in these studies, and none had any known language disorders, any child that would fall below the test norm may well do so because of an underlying language impairment. However, falling below the norm would not be sufficient in itself to signal impairment, for a number of reasons, as noted previously in this chapter (see also Chiat *et al*., 2013). It may also be concluded that as all children fell within the monolingual norms, any who did not could be assumed to have DLD. However, it is with caution that we should reach this conclusion (Chiat *et al*., 2013) as this may not be applicable to every language. Chiat *et al*.'s study of Turkish-English bilinguals revealed that the majority of L2 children performed below the monolingual norms. Had this notion been applied to this particular situation, almost all of the L2 children would be viewed as having a language impairment. Whether or not the same patterns would have been found using bilingual rather than monolingual norms is unknown, but given what we know about the differences between typically developing bilinguals and monolinguals (as noted earlier in this chapter), it is likely that monolingual norms are not sensitive enough to identify the nuances between bilingual and atypical development, hence why some typically developing bilinguals fell short of the CELF monolingual norms in Conti Ramsden *et al*. (2001) (see also later in the chapter). Performance on sentence repetition tasks may also be influenced by the design of the test used i.e. the vocabulary and syntax selected. In some cases, targeted sentences include complex vocabulary and morphosyntax that may not be entirely suitable for L2 children. Using items that are too complex may rely too much on language ability and so a test becomes more a test of language ability than language 'deviance' per se. Also, although the CELF is a useful assessment tool for highlighting differences in performance, as noted above, it fails to reveal the root cause of a child's difficulties or the nature of the mistakes they have made due to the quantitative nature of its scoring system. Marinis and Armon-Lotem (2015) highlight the importance of supplementing the SRep test with other language tests as the SRep does not assess all levels of language ability, only lexical,

morphosyntactic and syntactic skills: '[it]does not separate comprehension from production and it does not assess the children's pragmatic skills' (p.106). Other tests should also be used to ascertain a full profile of a child's abilities including the participants' full language history to inform of a child's age of acquisition and length of exposure to a language, since language performance is rarely completely devoid of the influence of context (Carroll, 2015). Parents are the primary educator for a child and so when creating any profile of a child's language abilities it is essential that information is sought from parents regarding children's language abilities at home. Finally, tests that are created in different languages will of course represent the morphosyntactic features of the particular language being tested and will invariably be different. This may therefore affect the performance of particular groups due to the varying nature of the languages being tested.

Sentence Repetition tasks do therefore have a place in the identification of DLD, and are not impacted greatly by length of exposure and age of onset, although bilingual norming is more suitable than monolingual norms for the purpose of identifying differentiation. However, SRep tasks should not be used in isolation as an indicator of DLD. Length of sentences must be considered if we are to gain a clearer picture of a child's abilities and not just their short-term memory capacity. Low performance may be an indicator of the effect of the specific language being used as grammatical structures may be more complex in one language when compared to another. Also, when bilingual children are compared to monolingual norms, poorer performance may be an indicator of the content of the tests used and so simpler vocabulary and morphosyntax may be a solution when testing L2. It may also be due to the absence of bilingual norming.

Parental reports

Parental reports are fundamentally reports provided by parents about a child's language ability. One particular example of this kind of report is Fenson *et al.'s* (1993) MacArthur-Bates Communicative Development Inventory (CDI).

This assessment tool was created for use with very young children based on parental report of their children's vocabulary. The aim of the tool was to be able to identify developmental delay in children below age 2;6 years and has been validated according to its reliability for American English by Fenson *et al.* (1994), for Mexican Spanish by Jackson-Maldonado *et al*. (2003), and for European Spanish by Lopez Ornat *et al*. (2005), among others. The CDI is now available for 61 languages.

Although parental reports can provide valuable information in relation to children's vocabulary knowledge there are limitations to its reliability as the parent of the child providing the information may over or under represent the results. However, parental reports can be very useful for measuring children's level of language exposure in the home which can then be used to analyse varying language groups.

Vocabulary tests

Vocabulary tests are often used as an assessment of children's language abilities as children with LI demonstrate delays in vocabulary acquisition. Vocabulary also correlated with literacy abilities, as noted above. There are a variety of vocabulary tasks available measuring both receptive and productive vocabulary although these have primarily been standardised using monolingual norms and so are not representative of bilinguals e.g. *British Picture Vocabulary Scales* (BPVS) (Dunn *et al*., 1997). In Wales the *Prawf Geirfa Cymraeg:Fersiwn 7-11* was developed by Gathercole and Thomas (2007) to measure bilingual Welsh-English children's receptive vocabulary. The test was developed using a generic age-matched norm, and age x home language (bilingual type) norm, and a school year x age x home language (bilingual type) norm. Although vocabulary tests are an integral part of language assessment it is important that these tests are not used in isolation as this can lead to overrepresentation of literacy problems as bilingual children will often appear to underperform in comparison with their monolingual peers, leading to bilingual children inaccurately being viewed as having literacy

difficulties, as noted previously in this chapter. However, the Gathercole and Thomas (2007) test offers a partial solution to this problem by providing a way of exploring bilinguals' performance in one language (in this case, Welsh) according to other bilinguals that share similar exposure to that language. However, as mentioned in Peña, Bedore and Kester (2016) although children with LI may initially exhibit delays in vocabulary acquisition they have been found to catch up with their TD peers and score within normal range (McGregor, 2009).

Reading comprehension tests

Reading comprehension tests are widely used by language practitioners and within schools to measure children's reading abilities as they provide an indication of children's knowledge and understanding of vocabulary and syntax. Given that reading comprehension provides the practitioner with a view of a child's literacy abilities these tests are used as a formal assessment, often being administered under test conditions. These tests are standardised using monolingual norms and are primarily administered in the child's language of instruction within the educational setting. For example, in Wales, National Reading tests have been devised for use in schools across the country. The tests are available in both English and Welsh, although they are only administered in Welsh to children within the Foundation Phase (3-7 years) in Welsh-medium schools due to the linguistic nature of their education, as outlined in Chapter 1. Children who are L1 English attending Welsh-medium schools are therefore not given the opportunity to demonstrate their literacy abilities in their L1 at that age. Beyond age 7, all children receive the test in both English and Welsh providing a more holistic reflection of their literacy abilities.

It is also important to note that reading comprehension tasks cannot sit as a sole reflection of a child's literacy abilities as they alone cannot be seen as representative of a bilingual learner's abilities. For example, Butvilofsky *et al.*

(2020), in their qualitative study of Spanish-English children examined the importance of biliterate writing assessment in gaining a better understanding of bilingual children's reading ability when compared with monolingual reading assessments. They found that writing assessments can provide an insight into children's reading development which would not be revealed through reading comprehension alone.

Cloze/fill the gap tasks

The cloze task was originally created by Taylor (1953, 1957) as a 'measure of an examinee's ability to comprehend text by removing words in a passage and having examinees replace them using the context' (Trace, 2020, p.235). Due to their ability to measure comprehension at a sentence and passage level they are often used within tests of reading comprehension and are a popular method for use in language assessment. A study by Trace (2020) explored examinee performance on a variety of 15 cloze tests. Findings suggest that reading ability of both L1 and L2 examinees can generally be measured using cloze tasks, although better understanding of what exactly a cloze test measures requires further exploration. In terms of their use in bilingual assessment the use of these tests could be influenced by level of exposure to the language as they require a certain level of reading ability and vocabulary knowledge and therefore would require a particular level of language knowledge to complete them. It would be important therefore that these tests are not used in isolation and that children are tested in both languages wherever possible.

Reaction time tasks

Another aspect of testing which comes into play is response time. Response time measures how quickly an individual is able to process the stimuli presented to them and is measured in milliseconds. By measuring language behaviours we are able to gauge whether monolinguals are able to process

language more quickly and efficiently than bilinguals or vice versa or whether these skills are developed differently (Perez *et al*., 2013). Due to its informative nature, response time has therefore been used by cognitive experimental psychologists as a basis for identifying any underlying issues with mental processes (Meyer *et al*., 1988; Posner, 1978).

Perez *et al.* (2013) draw attention to exploring the interaction between experimental factors such as lexical frequency with subject factors such as reading level as the effects are very interesting. Assumptions that certain processes (depending on what is being tested) are not functioning well may be made during analysis of particular tests. If comparing typical readers with dyslexic readers for example, the effect size between word frequency, a well-known marker in lexical access research (Coltheart *et al*., 1977; Forster, 1976, as cited in Perez *et al*., 2013), and word recognition time can be very small or non-existent in surface dyslexics. This is viewed as suggesting 'a poor functionality in the access to orthographic lexicon in the normal reading activity of dyslexics (Castles & Coltheart, 1993; Coltheart *et al*., 2001)' (Perez *et al*., 2013, p.134).

There are certain aspects of processing language which need to be considered with regards to testing response/processing times. Firstly, if an individual is asked to read text aloud their performance may be effected by the placement of the word i.e. the visual field in which it is placed. The frequency of the words presented may also have an effect on performance and also the lexical status and orthographic structure. Therefore words that are of low frequency and placed within the left visual field (Fiebach *et al*., 2002) with unpredictable pronunciation are not recognised as quickly as high-frequency words, centrally placed and with a predictable orthographic nature.

Measuring response times can be a very useful tool in assessing potential issues children are experiencing with language processing. However, these measurements should not be used in isolation as an indicator of language impairment but more to support other assessments made of bilingual children's performance. Other factors exist which may have an impact on children's performance on these tests include content of the words (whether

they are consonant based or pseudowords) and the population being tested. For example, it has been noted that very young children may not perform very well on such tasks due to not understanding the speed/accuracy instructions (Geisinger, 2003). Also, when asking participants to complete the task with speed and accuracy there may be an internal battle to decide whether to be as quick as possible through risk of being incorrect, or taking one's time to answer correctly which may affect the speed with which they complete the task.

Another issue to address when using reaction time tasks is the orthography of the language. As highlighted earlier in the chapter, some children take longer to read in some languages dependent on where the language lies on the orthographic depth continuum (i.e. languages with a deeper orthography can take longer to learn to read than those with a more shallower orthography). This could therefore affect the speed with which a child could complete a task.

Due to the current thesis being an exploratory study, a number of tasks were used to explore children's language and literacy abilities. A variety of standardised and novel tasks were used which included elements of the assessment methods discussed above. Vocabulary and reading comprehension were assessed through the use of standardised tests, these tests also included a cloze/gap fill element. Non-words, sentence repetition and cloze/gap fill were incorporated into the novel tasks designed specifically for the studies (studies 2 and 3). Reaction time was used within a timed written task (study 1) and parental report was used in seeking information regarding children's language background.

As has already been highlighted within this chapter there are a number of issues which practitioners need to keep in mind when making assessments of bilingual learners. The next section discusses some of these issues in more depth.

Issues for assessment

Some of the issues identified within the research literature discussed above support the need for fair and equitable assessment of bilinguals' language abilities that take into account (i) their abilities in each of their languages, (ii) the specific nuances of the languages they are learning, and (iii) the specific aspects of the individuals' experiences with each of their languages. This makes the design of language tests more complex, but very important to get right. Below, issues relating to testing in both languages, appreciating the uniqueness of each language (issues with translation/adapting existing tasks), and the role of linguistic experience in shaping test norms are discussed.

Testing in both languages

It has been demonstrated that a bilingual child's abilities in one language do not necessarily reflect their abilities in their other language (Gathercole *et al.*, 2013). If so, it would be unfair to make assumptions about a bilingual child's abilities based on assessment in one language alone. As stated in the Royal College of Speech and Language Therapists (RCSLT, 2006, cited in O'Toole, 2013) 'assessment of communication skills should take place in all the languages to which that person is exposed' (O'Toole, 2013 p.82). A child might find one language particularly difficult to learn, but this does not necessarily mean that they have a language impairment. As noted above, failing to test a bilingual child in both languages may, in some cases, lead to over-representation of a problem because the target language may not be their strongest language and therefore children may not perform as well as their monolingual peers. This could then lead to children being perceived as having a language impairment rather than what could be simply a difficulty with acquiring the second language (see for example Winter, 2001 and Bedore and Peña, 2008). Also, as noted by Armon-Lotem and de Jong (2015) 'L2 learners often produce language forms resembling those of

children with SLI' (p.1), however there are limited diagnostic tools available to help distinguish between a child with SLI and a second language learner, hence the importance of testing in both of a child's languages. There is also the issue of under-representation of potential problems that may not surface or be targeted in a given language: 'To gain a full picture of a child's language abilities, the ideal would be to examine performance in both languages of the bilingual child to determine if there is a linguistic problem' (Gathercole, 2013, p.7). This statement is echoed by many researchers in the field who emphasise the importance of assessing bilingual children in both languages in order to ascertain whether length/amount of exposure is the issue rather than developmental deficit (Chiat *et al.,* 2013). If a child has a language impairment it would be evident in both languages (Caesar & Kohler, 2007, Thordardotir *et al.,* 2006), in the same way as bilinguals who perform to a good standard in one language would tend to also perform relatively well in the other (Perez-Tattam *et al.*, 2013). The holistic theory of bilingualism states that a '…bilingual learner's languages are mutually reinforcing… development in one language influences development of the other' (Dworin 2003, as cited in Butvilofsky *et al.,* 2020, p. 4). However, if only one language is being assessed it is not possible to determine whether the difficulty is a result of under-exposure to a language or a language difficulty. It is therefore of vital importance to assess a child's language abilities in both of their languages (e.g. Marinis & Armon-Lotem, 2015; Gathercole, Thomas *et al.*, 2013).

As shown in earlier sections of this chapter there are many examples where bilingual children are seen to underperform in comparison with their monolingual peers but that this can level out as children get older. This would suggest there was no literacy deficit present only a deficiency in exposure to or knowledge of the language of testing, which would influence their performance in tests in a language which is not their L1. Through not being tested in both languages children can be perceived as having a literacy difficulty if they underperform (over-representation), or not be identified as having a literacy difficulty due to the expectation that their vocabulary knowledge for example would be lower due to having less knowledge of or

exposure to the language (under-representation). As highlighted by Butvilofsky *et al.* (2020) '..the experiences and knowledge of bilingual learners can never be fully measured or understood if their abilities are not assessed bilingually' (Butvilofsky *et al.,* 2020, p. 4).

Research is underway to try to develop tests which are suitable for bilinguals within different languages e.g. the Bilingual English Spanish Assessment, (BESA) (Peña *et al.* 2014); however, there are very few diagnostic tools available for assessing bilingual children with literacy difficulties. In Wales for example, there is only one literacy measure available in Welsh namely the Profion Glannau Menai (Payne, 1998), which although provided an excellent starting point for assessment of literacy difficulties in Welsh, by now children who have taken the test have become very familiar with its content and it therefore requires updating.

Normative data

As noted earlier on in this chapter, monolinguals and bilinguals differ in many respects, and these differences continue across the lifespan: 'Children growing up bilingually are not the same as children growing up monolingually and ... even fully fluent bilingual adults perform distinctly from their monolingual peers' (Gathercole, 2013, p.1). For that reason, bilingual children (and adults) cannot be normed effectively on monolingual samples since the way they organise language and their exposure to language differs, leading to differential performance on tests. As discussed, some bilingual children may have smaller vocabularies than their monolingual peers (but potentially higher vocabularies if the two languages are combined) due to differences in exposure and use of the language, which, at times may lead to the incorrect assumption that they are language impaired (Thordardottir *et al*., 2006). Yet, many current measures of language abilities rely on monolingual norms. The following section outlines some of the concerns with using monolingual norms with bilingual speakers, and discusses some of the alternative solutions that have come into play.

Since researchers have questioned the effectiveness of monolingual-normed tests as appropriate tools for the identification of language problems among bilinguals, researchers have set about creating tests that align the norms closer to the type of linguistic experiences of a given bilingual in an attempt to provide a more accurate view of their true language abilities. Some notable examples include Gathercole and Thomas (2007) for Welsh bilinguals; Peña (2014) for Spanish-English bilinguals; Verhoeven and Vermeer (1993) for Dutch monolinguals and bilinguals; Muñoz-Sandoval *et al*. (2005) for bilingual abilities in 18 languages; Mattes (1995) for vocabulary in English combined with Spanish, French, Italian, or Vietnamese; and Paradis and Libben (1987) for aphasic individuals. Dunn *et al.*'s (1997) British Picture Vocabulary Scale (BPVS) has also been developed to include bilingual norms and the Communicative Development Inventories (CDI) are also being developed in bilingual versions (see O'Toole, 2013; Ezeizabarrena *et al.*, 2013)**.** However, in most cases, the bilingual samples used in drawing a bilingual or multilingual norm is quite heterogeneous, including speakers who are on various points of the bilingualism continuum or involve a variety of different bilingual pairs, including those who speak English plus another language (e.g. the new BPVS). However, as noted previously in this chapter, bilinguals are themselves a heterogeneous group of individuals whose exposure to each language varies substantially, both quantitatively and qualitatively. One could therefore question to what extent norms that include L2 and L1 speakers relate to a given L2 or L2 speaker's performance on a task. Whilst it would be impossible to create norms that reflected the exact linguistic experience a given bilingual child has with their two languages, one alternative way of norming for bilinguals that has been developed involves norming according to the nature of language transmission practices in the home (Gathercole *et al*., 2013).

As found in other studies of Welsh-English bilinguals (see e.g. Gathercole & Thomas, 2009; Gathercole *et al*., 2008; Rhys & Thomas, 2013; Thomas *et al*., 2014), the results obtained by Gathercole and Thomas (2007) revealed clear and sustained differences in bilinguals' vocabulary knowledge at all stages of development between 7 years and 11 years of age. In addition to these Home Language norms, the test also provides an Age x School Year norm. This is an

extension on the usual age-reference norm found on other tests. This norm reflected that whilst some children are old in their year other children are young in their year, which may affect the rate of their development and, subsequently, their performance on standardised tasks (e.g. Bell and Daniels, 1990; Martin *et al.,* 2004; Oshima and Domaleski, 2006; Verachtert, 2010). In the PGC, children are not only normed for home language, but also normed for their age in a particular year. However, whilst the normative sample used for the PGC was relatively small (611 children), particularly when distributed into Home Language x Age x School Year groups (see Gathercole & Thomas, 2007, guidebook), the fact that the normative data produced for this test do show statistically significant differences across all the ages tested and across the three home languages demonstrate the importance of such norming practices for bilinguals, and the importance of comparing L2 speakers to other L2 speakers rather than being assessed in comparison with their L1 peers (Gathercole *et al*., 2008).

As highlighted in Boerma and Blom (2017) norm referencing is difficult given the complexity of bilinguals (e.g. variety in levels of exposure). However, as outlined in their study, norm-referencing may not be necessary when using a specific combination of assessment instrument. Their study included 132 5 and 6 year olds, half with previous diagnosis of LI and half typically developing. They used parental questionnaire, non-word repetition task and narrative task and found that late language emergence is a risk factor of LI and that family history may be a less reliable indicator of LI in bilingual than monolingual children. Their study provides some evidence that where norm-referencing is not possible, a suitable combination of assessment measures may provide valuable information about a bilingual child's abilities. This finding is supported by White and Jin's (2011) review of different bilingual assessment approaches, finding the sociocultural approach the most promising in providing an accurate assessment of children's abilities and that using a combination of approaches can help to mitigate any weaknesses found when only using one approach. Further research in this area is needed however to validate these findings.

Age/Order of Acquisition

When assessing bilinguals or creating suitable tests for the assessment of bilingual children another important factor to consider is the age or order of acquisition of a language. Age/order of acquisition (AoA or OoA) refers to the stage at which an individual begins to learn and/or acquire a new language. It has been well established that age/order of acquisition has an effect on lexical processing in the sense that words that have been learnt from an earlier age can be recalled quicker, and more accurately than newer words learnt at a later stage (for reviews, see Johnson & Barry, 2006; Juhasz, 2005, as cited in Perez *et al.*, 2013) as long as there are no other criteria present that could affect performance. This has been found amongst various populations and languages and in a variety of different tasks (as mentioned in Perez *et al*., 2013).

Perez *et al.* (2013) conducted a study on the effect of OoA when processing Welsh and Spanish words. A group of native Spanish adults were taught a series of Welsh words that they had not heard before. Words were introduced under controlled conditions where word length and frequency of presentation was controlled. Results found no effect of the OoA in terms of accuracy but a significant effect was found in relation to response time on lexical decision and semantic categorisation tasks but not on the picture naming tasks. This result differs to that of Izura *et al.* (2011) who found an effect of OoA on picture naming, lexical decision and semantic categorisation.

Issues for Assessment Summary

Bilingual assessment can be influenced by many things, including socio-economic status (Hoff, 2006), length and sources of exposure to a language, age/order of acquisition of both languages and dominant language of the wider community (e.g. Perez-Tattam *et al.*, 2013; Place & Hoff, 2011; Thomas *et al*., 2014). These factors must be taken into account when conducting an

assessment of a bilingual child to ensure a correct profile of the child's abilities is formed. There are a great deal of advantages with using standardised tests for assessment due to the variety of language aspects that can be assessed. However, it is worth bearing in mind some issues, in particular how children feel about the test situation, and also whether the standardised tests are based on monolingual or bilingual norms. Many tests are able to highlight what the child is unable to do but not why they may be unable to do it. It is recommended therefore that a number of different tests covering a wide range of skills are used and investigation into problem areas is carried out, in an attempt to ascertain the root of the problem before a conclusion of language impairment is made.

Implications for educators

As highlighted at the beginning of this chapter educators within bilingual schools are faced with the task of providing suitable support for children from varying language backgrounds. Although there are clear advantages to learning two languages side by side there is an ongoing issue of assessment and being able to successfully provide an accurate profile of a child's abilities in both of their languages. This is particularly important for children who are displaying literacy difficulties. Assessment tools for minority languages such as Welsh are currently limited and so teachers/educators are unable to provide diagnostic evidence of a child's literacy difficulties to enable them to access the necessary support (see Davies, 2016). It is important therefore that research continues to explore the potential markers of literacy difficulties in minority languages such as Welsh through considering the particular nuances of the language that have been deemed particularly problematic for learners.

The current studies

Due to the lack of assessment tools available in the Welsh language, this thesis aims to explore the potential markers of literacy difficulties in Welsh through investigating some of the more complex elements of the Welsh language (namely the mutation and plural systems). There are three studies presented. Study 1 explores the types of errors made by children in their written text to discover which are some of the most common errors made by bilingual Welsh-English children. This information was then used to inform Studies 2 and 3. Studies 2 and 3 explore potential links between literacy behaviours and knowledge of Welsh mutation and plural morphology – two structures that require high levels of phonological awareness abilities, which is known to be problematic in children with literacy difficulties. Study 2 explores children's understanding and application of some of the mutation rules in Welsh using novel tasks and Study 3 explores children's understanding of plural morphology. Performance on these tasks is analysed according to pupils' performance on standardised literacy tasks and a variety of traditional measures of children's phonological awareness, IQ, and memory for digits (short-term memory).

The next Chapter presents the findings of Study 1, an error analysis of children's English and Welsh written text.

CHAPTER 3

STUDY 1

Exploring the use of error analysis to identify markers of literacy difficulties in Welsh-English bilinguals' classroom written texts

This Chapter presents Study 1 – an exploration of the types of errors Welsh-English bilingual children make when writing in Welsh and in English. The chapter begins with an introduction to the study, a description of the sample and measures used, and the results of the error analysis. The chapter ends with a discussion of the findings and key areas for further exploration.

Introduction

As outlined in the previous chapter, one of the key areas that require further exploration in relation to bilingualism is that of assessment and the need for more appropriate, language-specific assessment tools. Given that literacy difficulties are varied and can manifest themselves in a number of different ways, assessments of literacy abilities need to be able to capture various aspects of literacy difficulties in ways that are compatible with the orthographic nuances of the language.

Understanding the difference between bilingual literacy performances that are typical of bilinguals and those that are indicative of a potential underlying problem is difficult, and is largely underexplored in relation to biliteracy in Welsh-English bilinguals. Studies that do exist to date either involve specialist teachers' views, that are largely drawn from anecdotal evidence and personal experiences (Davies, 2016), or involve the use of assessment tasks, adapted 'ad hoc' for Welsh, with a small sample of children (Thomas & Lloyd, 2004). Whilst both types of studies provide useful data, for the purpose of the current study, I wanted to focus on actual examples of children's day-to-day literacy

practices at school to gain a better understanding of the nature of the 'data' educators are regularly exposed to within the classroom environment.

Welsh-English bilingualism: issues for literacy

Welsh has a highly transparent orthography where there is a clear, almost one-to-one correspondence between graphemes and their respective phonemes, particularly in relation to consonants (e.g. *c* - /k/), but also in relation to vowels and diphthongs, albeit with some obvious exceptions. For example, *u* and *y* are realised as /ɨ/ and /ɨ/ in northern dialects (*llun* /ɬɨn/ 'picture', *dyn* /dɨn/ 'man', *byr* /bɨr/ 'short' and *pump* /pɨmp/ 'five') but as /ɪ/ in southern dialects (/ɬɪn/, /dɪn/, /bɪr/) whilst the grapheme *y* is also realised as /ə/ in all dialects (*yn* /ən/ 'in', *ynys* /ənɨs/ 'island'), although the two realizations of *y* are fairly predictable in terms of syllable placement. Similarly, the diphthongs *ai*, *au* and *ae* are all realised as /ai/ or /ɑi/ in southern dialects (*tai* /tai/ 'houses', *cau* /kai/ 'to shut/close', *cae* /kɑi/) but as /ai/, /aɨ/, and /ɑɨ / in northern dialects (/tai/, /kaɨ/ and /kɑɨ/) (see Jones 1993; Thomas & Lloyd, 2004). Despite these exceptions, Welsh is relatively transparent. However, there are certain aspects of the system – particularly those that seem to carry high phonological awareness loads – that do cause problems for children when learning to write, particularly among those who are diagnosed, or are suspected to be with, dyslexia (Davies, 2016).

According to Davies (2016), some of the most common areas that cause problems are exceptions within grapheme-phoneme correspondences around diphthongs that undergo subtle phonological changes when converting singular words to their plural form – e.g., *cae* /kɑɨ/ – *caeau* /kəɨaɨ/ 'fields', but *caead* /kəɨad/ 'lid' – caeadau /kəɨadaɨ/ 'lids'. (See Thomas *et al*., 2014, for an overview of Welsh plural morphology.) A second common area that causes problems is *mutation* – a morpho-phonological process whereby initial consonant sounds undergo phonological change under certain syntactic conditions, and these phonological changes are also represented in print. For

example, nouns with initial *p* /p/ undergo Soft Mutation (SM) into *b* /b/, *t* /t/ -> *d* /d/, *c* /k/ -> *g* /g/, *b* /b/ -> *f* /v/, *d* /d/ - dd /ð/, *ll* /ɬ/ -> *l* /l/, *rh* /r̥/ -> *r* /r/, *m* /m/ -> *f* /v/ and *g* gets deleted /ø/. These sound changes are triggered by a set of lexical items or syntactic contexts – e.g. *dy* 'your', *dau* 'two' and feminine noun gender triggers SM onto following nouns, such as *dy frawd* 'your brother' < *brawd* 'brother', *dau gi* 'two dogs' < *ci* 'dog' *y gath* (feminine) 'the cat' < *cath* 'cat' vs. *y ci* (masculine) 'y ci' < *ci* 'dog'. (For a thorough overview of the mutation system see Ball & Müller, 1992; Thomas & Gathercole, 2007; Thomas & Mayr, 2010.) Given that learners of transparent languages have been shown to rely somewhat on grapheme-phoneme conversions during the process of reading (Ellis & Hooper, 2001), the transparency of the phoneme-grapheme conversion in mutations may lead children, with or without literacy difficulties, to perform fewer errors when reading or spelling in Welsh (Thomas & Lloyd, 2004; Spencer & Hanley, 2003; Hanley *et al.*, 2004). Those with literacy difficulties however may demonstrate more effortful and slower attempts at reading, and produce more errors in writing than their typically-developing age-matched peers (Thomas & Lloyd, 2004; Wimmer 1993; Barca, Burani, Filippo & Zoccolotti 2006). However, given the additional level of accuracy required when writing or spelling words as compared to the spoken form, and the clear associations between anxiety and performance with language (Alexander-Passe, 2006; Burden, 2005; Riddick, Sterling, Farmer & Morgan, 1999), it may well be that early indications of literacy difficulties in Welsh will be more prominent in the written attempts that children produce rather than in their reading abilities (Thomas & Lloyd, 2004). Exploring potential patterns of errors demonstrated in the written texts of Welsh-English bilinguals – in Welsh and in English – is therefore necessary in order to help parents and educators identify (i) the typical errors that are expected of Welsh-English bilinguals, and (ii) whether there are qualitative (in terms of type of errors) and/or quantitative differences (in terms of amount) in the errors performed by bilinguals. This is all the more pertinent in contexts where bilingual children are receiving their education through the medium of a language that has a transparent

orthographic system, which may mask the appearance of literacy difficulties that are usually overtly marked in more opaque languages.

The main aim of Study 1, therefore, was to explore Welsh-English bilinguals' written texts from a timed written task in Welsh and in English, by looking at key patterns of writing behaviour across pupils with differing levels of literacy ability. Three types of bilinguals were involved – children whose dominant language was English (predominantly English at home and attended an English medium school), children who were from predominantly English homes but attended a Welsh medium school and children from predominantly Welsh homes who attended a Welsh medium school. In the absence of standardised tests looking at written/spelling abilities that were comparable across Welsh and English, standardised measures of reading ability – in Welsh and in English – were collected in order to gain some measure of children's general language abilities in literacy-based tasks. This allowed me to explore the extent to which certain patterns were more typical within the timed written texts of lower ability or of higher ability readers.

The research questions were as follows:

(i) Are there differences in the written texts of pupils from different Home x School Language Groups, who score lower and higher on standardised reading tests

(ii) If such differences exist, what is the nature of those differences?

In relation to the first research question, it was predicted that there would be a general reduction in the number of errors produced as the children within the sample became older. It was also predicted that those who scored lower on standardised scores of reading would produce more errors in their written texts than those who scored higher, but that there would be more errors overall in English than in Welsh due to the opaque nature of English. However, when writing in Welsh, any differences would be most prominent in areas that require higher levels of phonological awareness than in those requiring lower levels.

In relation to the second research question, it was predicted that the number and type of errors produced would be related to the child's exposure to a given language as well as to their reading abilities. Of the errors produced, it was predicted that a wider range of errors would be seen within the Welsh dominant bilingual (Welsh Home, Welsh School) group for text written in English and by the Welsh at school bilingual (English Home, Welsh School) group for texts written in Welsh.

Method

Participants

This study was carried out using secondary data from a convenience sample of 73 pupils, between the ages of 6;8 and 12;3 years (30 males, 43 females). These data were collected as part of a larger study that looked at factors influencing literacy abilities and self-esteem in bilingual children attending Welsh-medium and English-medium schools in North Wales (Young, 2014; Young, Rhys, Kennedy & Thomas, 2017). Parents were asked to report which languages were spoken in the home. Children in English-medium schools were only selected where the parents reported only speaking English at home (English dominant group). This provided a group that was as close to monolingual in the dominant language as possible. Children attending Welsh-medium schools were separated into two groups: those where Welsh was used more than 60% of the time in the home and those where English was used more than 60% of the time in the home. Children who heard mainly English at home but attended Welsh-medium schools (Bilingual group) were considered to have the most balanced exposure to the two languages. Children attending Welsh-medium schools who heard predominantly Welsh at home (Welsh dominant Bilingual group) were considered to have dominant Welsh-language exposure. Children whose parents reported using any other languages at home were removed from the sample. Children were selected from a range of schools in mainly rural and semi-rural areas of North Wales. Children were from a range of socio-economic backgrounds. The English-

medium schools used English as the language for all communication with students. The Welsh-medium schools used Welsh only to communicate with children in Key Stage 1. From Key Stage 2, children received specific English language lessons but all other communication with children was in Welsh (see discussion of school types in Wales in Chapter 1). While no formal English literacy instruction was given by Welsh-medium schools until Key Stage 2, the dominance of the English language in Wales meant that the majority of children had some English literacy awareness before that time, either from their parents or from children's media or the society within which they lived. For this reason, all children in Welsh-medium schools were tested in both English and Welsh.

For the purpose of the present set of analysis, only those children for whom data for age and home-school Language were available were included in the initial analyses. This provided a total sample of 63 children (22 male, 41 female). See Table 3 for a breakdown of the participants according to age and language group.

Table 3: Number of participants according to age and language group.

Age Group	English dominant	Welsh at school Bilingual	Welsh dominant Bilingual	Total
6;8 – 10 years	5	15	14	34
10;1-12;3 years	6	18	5	29
Total	11	33	19	63

Key:

English dominant Group– English Home, English School

Welsh at school bilingual Group – English Home, Welsh School

Welsh dominant bilingual Group– Welsh Home, Welsh School

Measures

Children were measured for their literacy abilities using the *Neale Analysis of Reading Accuracy* (NARA-II Revised; Neale, 1997) for English and the *Profion Glannau Menai* (Payne, 1998) for Welsh. Both measures consist of a series of written passages that increase in length and complexity as the child progresses through the task. Children were required to read each passage out loud and the researcher noted down the number and nature of any errors performed. After reading a given passage, the researcher would then ask the child a series of questions about the passage in order to gauge comprehension. Each passage allowed for a set number of errors to be performed. When the number of errors performed equated to or surpassed the number allowed for in a given passage, administration of the test would be discontinued at that point. These reading tests were administered on a one to one basis.

3-minute writing tasks

All children were asked to complete a brief writing task in English only for participants who were from English only homes and in both English and Welsh (on separate occasions) for bilinguals. After consultation with the Miles Dyslexia Centre at Bangor University, it was decided best to use a task based on the one-minute writing task described in the *Dyslexia Screening Test* (Fawcett & Nicolson, 2004). The task was administered in groups under examination conditions. Children were given one of two topics: Topic 1 'What I do in the morning' or Topic 2 'Explaining to an alien how to brush his teeth'. Bilinguals who were given Topic 1 in English would be required to write about Topic 2 in Welsh, and those given Topic 2 in English were required to write about Topic 1 in Welsh. The second topic was chosen to allow children to use similar ideas and terms in each of the writing tasks but without allowing a direct translation of one to the other. Although topic 2 was set in a different context, the basis of the written task was to describe how to brush one's teeth and so describing a process which would be very familiar to the participant.

Half of the children were asked to write in Welsh first, the other half were asked to write in English first to mitigate any practice effects.

Error Analysis

All grammar, spelling, punctuation and any other type of error produced in each of the Welsh and in each of the English written texts were highlighted, and each error was coded and entered into a spreadsheet. The number of words produced in English and in Welsh was also tabulated. A sub-set of the stories (10% of the overall sample – 7 stories in Welsh and 7 stories in English) was given to a second marker – a native Welsh-speaking adult with a background in research – as a measure of inter-rater reliability. Results revealed an agreement rate of 82.18%. In order to explore where the differences lay, a third, native Welsh-speaking marker explored the sample. This exercise revealed that the first marker had highlighted additional errors to the ones identified by the second marker. A separate reliability check was undertaken in terms of the classification of errors into error types. Both the first marker and the second marker agreed (100%) on the classification of the errors identified by the first marker in the sub-sample provided. These error types reflected the typical errors that teachers would highlight in children's written work. The error types identified within the texts fell under the following categories.

Phonological error.

Errors coded as phonological errors included (i) incorrect letter-sound correspondence e.g., using *y*,(which is often realised as /ə/ in Welsh,) for *e* /ə/ (e.g., **thy* for *the*, *w* /u/ for *oo* /u/), English *f* /f/ for Welsh *ff* /f/; (ii) omission of graphemes whose phonemes are not realised in speech - e.g., **monin* /monɨn/ for *morning, *opend* /opənd/ for *opened*; and (iii) applying the silent *w* /u/ where unnecessary (e.g. **wrinse* for *rinse* /rinse/). In some cases, separating a phonological error from a spelling error was difficult. For that reason, errors were recorded in both categories (e.g., **paist* for *paste* –

phonological error recorded for the use of *ai* /eɪ/ instead of *a* /eɪ/ and spelling error recorded for not using the 'magic e' rule).[8]

Letter/word omissions.

Letter/word omission errors included the omission of morphological markers at the end of a word e.g., *get *dres* instead of *get dressed*; the omission of letters in single-syllable prepositions and connectives such as *i* for *in* and *an* for *and*; the omission of grammatical markers such as the predicative particle '*dw i yn codi'(Welsh)* 'I-is-getting up' written as **'dw i codi'*; and definite article omission such as **o gloch* for *o'r gloch (Welsh)* 'of-the-bell' ('o' clock').

Vowel substitutions.

Vowel substitution errors included specific instances where an incorrect vowel grapheme was used to represent a vowel phoneme – e.g. *y* instead of *u* in **brysh* /ə/ instead of *brush*; **olso* /o/ instead of *also*; **healthe* /i/ instead of *healthy*; *mund* /ɨ/ Welsh for 'to go' instead of *mynd*.

Punctuation error.

Punctuation errors included capitalisation errors, missing or inappropriate use of apostrophe or full stop, and the omission of a question mark at the end of a question.

Colloquialism.

Here, colloquialisms were included that mainly consisted of contracted forms that are not permissible in the written form, such as **till* (*until*)*, *gory* /gɔru/ (g*ot to*), **rhoid* (*rhoi* Welsh for 'to give'), etc.

Spelling error.

All words that did not appear in their target form were included in this category. The types of errors that were included in this classification were as follows:

---

[8] The 'magic e rule' is a term used by educators to explain to learners how the addition of the word-final –*e* grapheme alters the sound realisation of vowels in the word stem – e.g., /a/ - /eɪ/ in *past – p*a*ste*.

(i)   incorrect use of digraphs (e.g., lack of homogenous digraphs when adding a suffix – In Welsh *dant* 'tooth' *dannedd* 'teeth' written as *danedd;* application of homogenous digraphs where unnecessary – e.g., *fynnu* for *fyny* 'up');

(ii)   misuse of *c*, *ck* and *k* for /k/;

(iii)   *your* for *you*, *there* for *their, here* for *hear* and vice versa;

(iv)   letter inversion- e.g. *minuets* for *minutes*;

(v)   missing letters e.g. missing silent letters (*no* /no/ for *know*); single letter graphemes instead of heterogeneous digraphs (*redy* instead of *ready*); *-t* instead of *–ed*;

(vi)   word separation - e.g. *down stairs* for *downstairs*, *my self* for *myself*,

(vii)   overt realisation of /ʃ/ in Welsh *si* instead of in English *sh* – e.g. *brwshio* instead of *brwsio* 'to brush'.

Incorrect Verb Tense.

This included a small number of instances where participants had used the incorrect verb tense in the English written task e.g. *pick* for *picked*.

Preposition and plurals.

Incorrect prepositions and incorrect plural forms were targeted but as very few examples existed these were not included in the final analysis.

Code-switching.

In some cases, children borrowed a noun from English into their Welsh texts but without full integration into the language. For example, children would apply a Welsh morphological marker such as *–io* ('-ing') to a plausible Welsh borrowed stem *tecst* 'text' but retained the English stem as in *textio*, or children retained the English word-initial digraph of a borrowed word that is realised differently when pronounced and spelled in Welsh - e.g. *chocled* /tʃ/ for *siocled* /ʃ/.

Mutation errors.

Items coded as Welsh mutation errors included no mutation where a mutation was expected (e.g. *y *cawod* 'the shower' for *y gawod; fy bag* 'my bag' for *fy mag*), over-mutation (words mutated when not necessary to do so – e.g. *y *benwythnos* for *y penwythnos* 'the weekend'), and incorrect mutation (Soft Mutation applied where an Aspirate Mutation was expected – e.g. a *gael <cael* for a *chael*). This category was also split into three sub-categories to represent the type of incorrect mutation made i.e. Aspirate, Soft or Nasal where applicable.

Results

Length of text

As can be seen in Table 4, the average number of words children produced when writing their story texts was fairly similar within a given group of children (40.7 in English vs. 39.19 in Welsh for children in the Welsh at school bilingual group and 24.93 in English vs. 25.93 in Welsh for children in the Welsh dominant bilingual group). Across groups, however, the Welsh at school bilingual children averaged 40.7 words in English, which is similar to the English dominant children who averaged 43.09 words, but much higher than the Welsh dominant bilingual children who averaged 24.93. Likewise, the Welsh at school bilingual children averaged 39.19 words in Welsh compared to the 25.93 words averaged by the Welsh dominant bilingual children. However, in terms of average words per sentence (WPS), performance was similar across-the-board in relation to the English texts (14.82 for the English dominant group, 15.24 for the Welsh at school bilingual group and 15.24 for the Welsh dominant bilingual group) whilst there was a tendency for Welsh at school bilingual children to produce slightly longer sentences than Welsh dominant bilingual children in Welsh (14.02 vs. 12.11). Whilst this pattern may be related to children's ability to punctuate their sentences appropriately, and relied on the subjective ratings made by two independent raters, these results do suggest that there were more data (and therefore more room for

error) produced by the Welsh at school bilingual children than the other two groups, which should be kept in mind whilst making any cross-group comparisons of the results. However, the data also revealed that the range of errors produced by the Welsh at school bilingual and Welsh dominant bilingual children were similar, in English and in Welsh, suggesting that language dominance did not influence the pattern of error in this sample. In order to determine whether the differences between mean scores were statistically significant an ANOVA was run to compare the three groups in relation to their English scores (Number of Words, average number of words and words per sentence (WPS)) and t-tests run between the Welsh at school bilingual and Welsh dominant bilingual groups to compare mean scores on Welsh number of words, average number of words and WPS (as the English dominant group would not have completed the task in Welsh). Results revealed no significant difference between groups on average number of words in English or Welsh nor WPS in English or Welsh. Significant difference did however emerge in relation to the total number of words in English ($p=0.28$, $SE=6.308$) and in Welsh ($p=0.44$, $SE=6.374$) between the Welsh at school bilingual and Welsh dominant bilingual group. This significant difference could however have been impacted by the group sizes and so it would not be possible to draw any definite conclusions from this finding.

Table 4: Measures of text length for English and Welsh according to home-school language groupings.

Measures of text length	English dominant (*n*=11)	Welsh at school bilingual (*n*=33)	Welsh dominant bilingual (*n*=19)
Number of words (English)	474	1099	349
Number of words (Welsh)	-	1058	363
Average number of words (English)	43.09	40.7 (*n*27)	24.93 (*n*14)

Average number of words (Welsh)	-	39.19(*n27*)	25.93 (*n14*)
Average words per sentence (WPS-English)	14.82	15.24	15.24
Average words per sentence (WPS-Welsh)	-	14.02	12.11

In order to reduce potential misinterpretations of the data, in what follows, analyses are conducted in each home-school language group separately. Firstly however, whole sample data is presented to help gain an overall picture of the patterns.

Error data

Table 5 below totals the number of errors produced across the full sample (*n=63*) in each category of error observed.

Table 5: Number of errors observed per error category across the whole sample.

Error Type	Total
Phonological error (English)	236
Omissions (English)	1
Incorrect definite article (English)	0
Vowel substitutions (English)	60
Punctuation error (English)	56
Incorrect connective (English)	3
Colloquialism (English)	1
Spelling error (English)	152
Codeswitching English to Welsh	1
Incorrect preposition (English)	1
Incorrect verb tense (English)	1
Codeswitching	15
Mutation errors	87
Aspirate	12

Soft	34
Nasal	41
Omissions (Welsh)	9
Vowel substitutions (Welsh)	34
Incorrect connective (Welsh)	17
Incorrect definite article (Welsh)	1
Incorrect circumflex (Welsh)	11
Incorrect preposition (Welsh)	3
Colloquialism (Welsh)	17
Phonological error (Welsh)	56
Spelling error (Welsh)	78
Punctuation error (Welsh)	56
Incorrect verb tense (Welsh)	1

As shown in Table 5, the total number of errors observed within the children's written scripts was relatively similar across the two languages (512 in English and 472 in Welsh). However, the errors produced within the English scripts were mostly limited to phonological errors (236), spelling (152), vowel substitutions (60) and punctuation (56). Conversely, the errors in Welsh were more widely dispersed, with the most prominent type consisting of mutation (87) and spelling (78), closely followed by phonological error (56) and punctuation (56).

In order to look at these patterns in relation to children's home-school language background and in terms of their standardised scores for reading, three separate Spearman's correlational analyses were performed looking at performance patterns among English dominant, Welsh at school bilingual, and Welsh dominant bilingual children across the age span involved in the sample. In what follows, all significant correlations are highlighted.

English Dominant Group (English-Home, English-School, monolinguals).

The only two variables that correlated with age among this group were the number of words produced in English (strong positive correlation) and the

number of punctuation errors produced in English (strong negative correlation) (see Table 6 below). As one might predict, the older the child, the more words they produced, and the fewer punctuation errors they made. The only significant correlations relating to English reading abilities was the WPS English (strong positive correlation) – that is, as NARA Comprehension scores increased, so did the WPS English within English dominant children's English stories. As expected, both sets of reading scores – Accuracy and Comprehension – strongly correlated with each other. However, the sample obtained for this group ($n=11$) is small, which may account for the small number of correlations found. Nevertheless, the pattern of performance does suggest that the number of errors produced in English – particularly within the category of phonological errors and spelling – are similar across-the-board and no more prominent among those who have good than those who have poor literacy skills as measured by standardised reading scores. Such data are therefore unlikely to yield much information to help identify potential 'at-risk' children with respect to literacy abilities in English, possibly due to the opaqueness of the orthography.

Table 6: Significant correlations within the English Dominant Group ($n=11$).

Variable	Age		NARA Accuracy Score		NARA Comprehension Score	
	'r'	*'p'*	*'r'*	*'p'*	*'r'*	*'p'*
Number of Words English	.945	.000				
Punctuation Error English	-.575	.032				
NARA Comprehension Standardised Score			.879	.000		
NARA Accuracy Standardised Score					.879	.000
WPS English					.533	.046

Welsh Dominant Bilingual Group (Welsh-Home, Welsh-School).

Table 7 presents the correlations performed on the error data produced by this sample group and their performance on standardised scores of reading abilities.

Table 7: Significant correlations observed within the Welsh Dominant Bilingual Group (*n*=19).

Variable	Age		PGM Accuracy Standardised Score		PGM Comprehension Standardised Score		NARA Accuracy Score		NARA Comprehension Score	
	'r'	'p'	'r'	'p'	'r'	'p'	'r'	'p'	'r'	'p'
Number of Mistakes English	-.639	.007					-.469	.045	-.568	.017
Phonological Error English	-.703	.003					-.570	.017	-.675	.004
Vowel Substitutions English	-.499	.035							-.525	.027
Mutation Errors	.669	.004	-.568	.017						
Nasal Mutations	.761	.001	-.604	.011					.474	.043
Phonological Error welsh							-.579	.015	-.704	.002
Spelling Error Welsh							-.678	.004	-.687	.003
Punctuation Error Welsh	-.493	.036								
PGM Accuracy Standardised Score	-.570	.005			.729	.000				
PGM Comprehension			.729	.000			.452	.030	.504	.016
NARA Accuracy Standardised					.452	.030			.953	.000
NARA Comprehension					.504	.016	.953	.000		

As one might predict, the total number of mistakes performed in English along with the number of phonological errors and vowel substitutions observed, decreased with age, in the same way as punctuation errors decreased with age in Welsh, however phonological error in English displayed the strongest correlation. Contrary to prediction, however, as the participants' age increased so did the number of mutation errors they produced, and particularly so for Nasal Mutation, with a relatively strong significant correlation. Possible reasons for this pattern are presented in the discussion.

Interestingly, scores on the four measures of literacy abilities did not correlate with each other in all instances. Whilst Accuracy and Comprehension scores within each test correlated with each other, PGM Accuracy did not correlate with either NARA Accuracy nor NARA Comprehension whereas PGM Comprehension correlated with both. In addition to that, reading accuracy as measured by the PGM Accuracy score seemed to decrease as Welsh dominant bilingual children became older.

In terms of reading ability, PGM Accuracy scores correlated with one error variable: mutation errors, and Nasal Mutation errors in particular. PGM Comprehension scores did not correlate with any error variables. This may suggest that accuracy scores in Welsh, particularly around mutation, may be more useful as a screening measure for potential literacy difficulties than more general accuracy measures, or comprehension of text in Welsh, and may warrant further investigation.

In contrast, both NARA scores (Accuracy and Comprehension) correlated with a number of variables, including the total number of mistakes made in English, the number of phonological errors made in English and with the number of phonological errors and spelling errors performed in Welsh (with which it had the strongest correlation), suggesting that those who scored highest on the NARA test produced the least amount of these errors. The NARA Comprehension score also correlated negatively with vowel substitutions in English and positively with Nasal Mutations in Welsh.

Together, these patterns may suggest that Accuracy scores, particularly among Welsh dominant bilinguals when writing in Welsh, may yield more language specific data such as practices around the use of mutations, and that the number of

phonological and spelling errors performed by Welsh dominant bilingual children in Welsh may yield useful information to help identify children who are potentially 'at risk' in relation to literacy development in that language.

Welsh at School Bilingual Group (English-home, Welsh-School).

Table 8 presents the correlations performed on the error data produced by this sample group and their performance on standardised scores of reading abilities.

Table 8: Significant correlations within the Welsh at School Bilingual Group (*n*=33).

Variable	Age		PGM Accuracy Standardised Score		PGM Comprehension Standardised Score		NARA Accuracy Score		NARA Comprehension Score	
	'r'	'p'	'r'	'p'	'r'	'p'	'r'	'p'	'r'	'p'
Age			-.800	.000					.353	.026
Number of Words English	.655	.000	-.547	.003			.453	.011	.413	.020
Number of Mistakes English	-.499	.004	.411	.023			-.597	.001	-.486	.007
Phonological Error English	-.545	.002	.408	.024			-.701	.000	-.539	.003
Vowel Substitutions English	-.691	.000	.548	.003			-.600	.001	-.578	.001
Punctuation Error English	-.501	.004								
Number of Words Welsh	.760	.000	-.601	.001			.461	.010	.657	.000
Mutation Errors	.541	.002	-.415	.022					.353	.042
Soft Mutation									.374	.033
Nasal Mutation	.629	.000	-.614	.001						
Colloquialism Welsh	.478	.006	-.344	.050					.455	.011

Number of mistakes Welsh									-.335	.051	-.322	.034
Vowel Substitutions Welsh	-.356	.034	.415	.022								
Phonological Error Welsh	-.600	.000	.402	.026					-.600	.001	-.580	.001
Punctuation Error Welsh	-.647	.000	.506	.006					-.400	.024		
Spelling Error Welsh	-.355	.034	.373	.036								
PGM Accuracy Standardised Scores	-.800	.000			.428	.009						
PGM Comprehension Standardised Scores			.428	.009					.506	.002	.373	.021
NARA Accuracy Standardised Scores					.506	.002					.685	.000
NARA Comprehension Standardised Scores	.353	.026			.373	.021			.685	.000		
Incorrect Preposition Welsh					.347	.048			.340	.048		
WPS English	.454	.009	-.515	.005					.360	.039		
WPS Welsh	.382	.025										

The data produced by this particular group of bilinguals generated a number of correlations. These patterns are summarised below.

First, as predicted, a positive correlation was found between the number of words produced and age and between WPS and age, both for the stories written in English and for stories written in Welsh. That is, the stories produced by this sample of children increased in length as they became older. Likewise, the overall number of errors, phonological errors, vowel substitutions and punctuation errors produced in English and the number of vowel substitutions, phonological errors, punctuation errors, and spelling errors in Welsh decreased as the age of the sample increased. However, contrary to prediction, and in line with the results of the Welsh dominant bilingual group, the number of mutation errors produced, particularly in relation to Nasal Mutation, increased with age, in the same way as the number of colloquialisms produced increased with age (moderate to strong correlations). Potential reasons for these patterns are presented in the discussion.

Second, whilst scores for the PGM Accuracy correlated positively with the total number of errors, the number of phonological errors and the number of vowel substitutions produced in English, scores for the NARA Accuracy and NARA Comprehension correlated negatively with these same variables. Likewise, whilst scores for the PGM Accuracy correlated positively with the number of phonological errors and punctuation errors produced in Welsh, similar correlations involving NARA Accuracy and/or Comprehension scores produced negative correlations. Vowel substitutions in Welsh and spelling errors in Welsh correlated positively with PGM Accuracy scores but did not correlate with any of the other standardised measures of reading abilities. The total number of mistakes produced in Welsh correlated negatively with both NARA Accuracy and NARA Comprehension.

Third, whilst scores for the PGM Accuracy correlated negatively with WPS in English, the number of words produced in Welsh, the number of mutation errors, particularly Nasal Mutation, and colloquialisms produced in Welsh, similar correlations involving NARA Accuracy and/or Comprehension scores produced positive correlations. However, interestingly, for this group, Soft

Mutation errors did correlate positively with NARA Comprehension scores, suggesting that those who score highest on reading comprehension in English produce the most Soft Mutation errors. This will be further discussed later in the chapter.

Fourth, there was a marginal, but positive correlation between scores on PGM Comprehension and NARA Accuracy and the number of preposition errors produced in Welsh.

Finally, as was the case for Welsh dominant bilinguals, whilst Accuracy and Comprehension scores within each test correlated with each other, PGM Accuracy did not correlate with neither NARA Accuracy nor NARA Comprehension whereas PGM Comprehension correlated with both. In addition to that, reading accuracy as measured by the NARA Comprehension score seemed to decrease as children in the Welsh at school bilingual group became older.

Discussion

This small-scale, exploratory study served two main purposes: first, it served to explore the number and types of writing errors children produce in Welsh and in English and, second, it served to explore the extent to which the number and type of errors produced correlated with children's literacy abilities as measured via standardised reading tests. Exploring potential differences across children with varying levels of reading abilities also served to highlight any potential markers of literacy difficulties that could be evident in children's written texts that could warrant further exploration in this thesis. Given that schools in Wales that teach through the medium of Welsh do so with children from varied home language backgrounds, exploring which aspects of each language lead to certain types of errors among these different types of bilinguals – and whether any of these aspects were more prominent among those with poor reading skills than among those with good reading skills – would also provide useful information. However, given that Welsh at school bilingual children (and English dominant children for English) seemed to

produce longer texts in both Welsh and English than Welsh dominant bilingual children, and that Welsh at school bilingual children seemed to produce slightly longer sentences in Welsh than Welsh dominant bilingual children, care should be taken in interpreting the data since it is not clear why such differences were found. For example, these differences may be due to a cohort effect or that they were less used to writing creative texts or less experienced in doing so in both languages than Welsh at school bilingual children. Whatever the reason, it is worth considering that each home-school language group, which consisted of different numbers of children, produced different amounts of data, which may have influenced the number of errors observed. However, whilst the stories produced were, in general, relatively short and unpredictable, it was possible, nonetheless, to assign errors – in a consistent manner – to a series of error types, which revealed some interesting patterns within the data that suggest further avenues for research. These patterns are discussed briefly below.

The research questions for this study were (i) Are there differences in the written texts of pupils from different Home x School Language Groups, who score lower and higher on standardised reading tests? and (ii) If such differences exist, what is the nature of those differences?

First, and as predicted, there was a general increase in the length of the stories children produced and a general reduction in the number of errors they made as the children within the sample became older. A notable exception to this was in relation to mutation, and Nasal Mutation in particular, and, in the case of the Welsh at school bilingual group, to the number of inappropriate colloquialisms that were used. That is, as the children within the sample became older, the more mutation errors – Nasal for Welsh dominant and Welsh at school bilingual groups and the more (inappropriate) colloquial forms they used in their writing (Welsh at school bilingual group only). These types of errors may be reflective of a growing knowledge of the spoken form of the language, whereby Nasal Mutation is often incorrectly omitted and replaced by Soft Mutation (e.g., *yn Gaernarfon < Caernarfon* instead of *yng Nghaernarfon* 'in Caernarfon') and whereby the phonological realisations of

word forms differ across the grapheme-phoneme pairing. For this reason, any tests that are developed to distinguish between literacy disorders and typical developmental delay need to consider language-specific factors that may help or hinder the identification of true difficulties.   Whilst mutation may be complex and multifaceted, given its predominance within the language – particularly in the written form – and its reliance on phonological awareness, understanding more about the types of difficulties children with literacy difficulties may or may not show in relation to mutation could be useful for teachers. Among the Welsh at school bilingual Group, one additional exception was found in relation to PGM Accuracy scores. Whereas NARA Comprehension scores correlated positively with age (i.e. the older the sample, the higher their reading comprehension scores in English), the PGM Accuracy scores correlated negatively (i.e. the older the sample, the lower their reading accuracy scores in Welsh). Whilst the reasons for this pattern is not clear from the current data, it may be the case that as children become older, their increased knowledge of colloquialisms, particularly around mutations, may result in inappropriate 'speech productions' in their texts.

Second, and contrary to our prediction that there would be more errors in English than in Welsh due to the clear difference in orthographic depth across the two languages (Spencer & Hanley, 2003; Hanley *et al.* 2004), the children produced a similar number of errors across the two languages. However, the errors in English were clustered primarily around two specific error types – phonological errors and spelling – and were numerous in number, whereas the errors detected in Welsh were more widespread and less numerous in comparison. Whilst such observations may not seem so significant on the surface, in relation to a language with a transparent orthography, such findings yield useful data that can contribute to on-going efforts to build realistic profiles of bilinguals' literacy competencies. The main error types in Welsh related to mutation, spelling, phonological errors and punctuation, suggesting that although Welsh has a highly transparent orthography, there are still aspects that cause different types of problems for children across-the-board, particularly those that require certain levels of phonological awareness such as mutation, and those that require sound knowledge of subtle spelling-

to-sound correspondences, such as those required in spelling. If the transparent nature of a language's orthography can indeed mask children's underlying difficulties, it may well be that the breadth and number of errors produced within a timed written task can serve as a good first predictor of potential difficulties, echoing claims made elsewhere by Thomas and Lloyd (2004), Wimmer (1993) and Barca, Burani, Filippo and Zoccolotti (2006).

Third, no correlations were found between scores on the standardised measures of reading abilities (NARA Accuracy and NARA Comprehension) and any error types performed by the English dominant group in English. In contrast, a number of correlations were found, both in relation to the texts written in English and to the texts written in Welsh, among the Welsh dominant and Welsh at school bilingual groups. This suggests that among Welsh dominant bilingual and Welsh at school bilingual children, at least, those who are the least skilful at reading, tend to produce more errors of specific types in their written work than those who are more skilful readers. Such errors among the least skilful readers may yield important information for educators and should be explored further. However, not all measures of reading abilities revealed the same correlations, and not all measures revealed correlations that showed the same direction of effect. When dealing with bilingual children, measuring their abilities in both languages is essential (Gathercole, Thomas & Hughes, 2008; see also Chapter 2), but to ensure meaningful measurements within any given language, tests must be able to capture the nuances within that language and measure those aspects that are likely to help distinguish best between errors that are typical of those who have good literacy skills and those who have not. The results of the current study provide some initial indication of what those error types might be, and suggest further studies are necessary to take a closer look at different types of children's knowledge of these structures.

Limitations of Study 1

One of the most challenging elements of study 1 was the categorisation of the error types and making decisions on which category each of the errors fell into. Some error types fell within more than one category, for example, an error in spelling could also be linked to a phonological error. Without being able to question the children directly about the particular error it was not possible to be certain of the precise nature of that error and so a score was recorded for both error types. Another limitation relates to the administration of the tasks within this study. Being that this was secondary data I did not gather the data myself and so was unable to fully analyse the methods used during administration of the tasks.

Summary of Study 1

The results of the current study suggest that a 3-minute writing task can provide some distinguishing features of literacy abilities across children with varying levels of literacy abilities, although, in the present case, it is not known why the texts produced tended to be much shorter among the Welsh dominant bilingual group. The wider application of these results suggest that it is worth exploring more holistically the literacy practices of bilinguals, particularly those who are primarily exposed to a language with transparent orthography, particularly language-specific structures that require high levels of phonological awareness, such as mutation and plural morphology in Welsh.

The following chapters will focus on these specific Welsh language structures (mutation and plural morphology) with an exploration of how children's knowledge and application of these specific structures relate to their reading ability and may yield useful information for educators.

CHAPTER 4

<u>Introduction to Studies 2 and 3</u>

This chapter outlines the development of Studies 2 and 3, their relationship to the literature and to previous findings within this thesis, and the design of the tasks employed. The chapter begins with a rationale for each of the studies, followed by a description of the sample and measures employed within each of the studies. It then presents the outcomes of the Pilot studies that were used to inform the administration and design of all novel tasks (mutation and plurals tasks).

The next chapter, Chapter 5, presents the results of Study 2, followed by Chapter 6, which presents the results of Study 3. The overall findings of the three studies are then discussed in Chapter 7.

Rationale for Studies 2 and 3

Study 1 served mainly as a scoping study to learn some more about the types of errors different types of bilingual children attending Welsh-medium schools in Wales performed in their writing when writing in Welsh and in English. The results highlighted a variety of error types, but revealed that error types in English tended to focus predominantly around four types – phonological error, spelling error, and, to a much lesser degree, vowel substitution and punctuation error – whereas the error types in Welsh were less numerous per type, but distributed across a much wider range of error (see Chapter 3). An additional set of analyses explored patterns of errors relative to pupils' standardised literacy scores in both Welsh and in English, and relative to their home/school language backgrounds. These data revealed some interesting relationships between the types of errors children produced in their writing and their reading abilities and their home/school language, that warrant further exploration. Of particular interest, and something that was also highlighted in Davies (2016), was the role of mutation and its relationship to

children's age, language background, reading ability and phonological awareness skills. Mutation correlated negatively with reading accuracy, which was hypothesised to be due to a growing knowledge of the spoken form of the language, whereby Nasal Mutation is often incorrectly omitted and replaced by Soft Mutation as noted in Chapter 3. As also noted in Chapter 3, this structure is of particular interest because of its overt realisation as a phonological sound/grapheme change, which requires a certain level of phonological awareness skill in order to execute and process. Study 2, therefore, examined children's knowledge of mutation, both in terms of their ability to select the appropriate mutated form (multiple-choice cloze task), whether they heard a mutation error and were able to repeat it (sentence repetition task) and in terms of their ability to identify erroneous mutations (grammaticality judgement task) and present the required form for the triggering context (sentence correction task). Whilst there are a few studies by now that explore children's acquisition of gender mutation in Welsh (e.g., Gathercole, Thomas & Laporte, 2001; 2005; Thomas & Gathercole, 2007; Gathercole & Thomas, 2009; Binks & Thomas, 2019), less is known about children's acquisition of mutation in general (although see Thomas & Gathercole, 2007; and Ball & Müller, 1992 for a review). However, most of the aforementioned studies have focused on oral and/or receptive competence and not so much on the use of or engagement with the system in the written form. Study 2, therefore, took the form of six tasks that involved varied engagement with mutation that required reading, listening and speaking skills.

Likewise, Study 3 examined bilingual children's knowledge of Welsh plural morphology, due to its high level of phonological load, particularly around the sound changes that signal a change from singular to plural forms that require similar changes in the written form. This structure was also featured in Davies (2016) as a structure Specialist Dyslexia Teachers found to be particularly taxing for bilingual, Welsh-speaking children with literacy difficulties. Again, as with mutation, little is known about children's knowledge of plural forms, particularly around their attempts to create plural forms in writing. Information gleaned from Studies 2 and 3 would therefore provide useful and

novel information on what children know about these systems, and whether performance differs across different types of children, and particularly so across those who perform at various levels on standardised literacy tasks.

For both studies, performance on tasks measuring mutation and plural morphology was analysed according to (i) children's standardised reading scores, (ii) children's phonological awareness skills (as measured by a battery of tasks that are typically used in various screening tools – see below), (iii) age, (iv) IQ, and (v) memory for digits. The measures used for each of these variables are described in detail below.

Participant Sample

Language background

A total of 69 (28 female, 41 male) children aged between 7:8 and 11:7 years took part in the study. 52 were classed as L1 Welsh speakers according to a participant questionnaire completed by parents (see Appendix 6). The remainder of the sample (n=17) consisted of 11 L1 English-speaking children, a further four children whose parents considered them to be 2L1 bilinguals and an additional two who had a first language that was neither English nor Welsh. As the focus was primarily on children who were Welsh/English bilinguals, those whose home language was neither English nor Welsh were removed from the sample leaving a final total of 67 which were included in the final analysis.

Language abilities

Ideally, this research would have been conducted with a group of Welsh-English bilinguals who had been diagnosed as dyslexic, and a matched sample controlled for age, language background, SES and non-verbal IQ. However, as experienced also by Thomas and Lloyd (2004), finding Welsh-English bilinguals with the diagnosis was very difficult. Within the sample recruited for the present studies there were two children with a diagnosis of DLD, two

showing signs of dyslexia (although no formal diagnosis had been made) and six who reported other impairments (although these were not necessarily formal diagnoses). One participant did not provide information of dyslexia/other impairment. These impairments were taken into consideration when asking the individuals to take part in certain tasks and in some cases they were not requested to take part. Due to the very low number, the data of the children who had been identified on the questionnaire as having dyslexia/other impairment were not removed from the sample.

Socio-economic status (SES)

The socio-economic status (SES) of the participants within the sample was gathered in the form of parental occupation and whether they were in receipt of Free School Meals (FSM). However, as only two participants within the whole sample (where data was available) were in receipt of FSM, and these two were in two different language groups, therefore not biasing one group more than another (see below), it was unnecessary to eliminate their data. Parental occupation data was used to provide a scale according to predicted salary; however, due to the unreliability of this data, and the relatively large amount of missing values for this category it was decided that FSM would be used as the primary indicator of SES.

Region

In order to recruit children for whom Welsh was their dominant or home language, children were recruited from Welsh medium schools in the Anglesey, Conwy and Denbighshire regions of North Wales where 53%, 37% and 30% of their respective adult populations speak Welsh (Stats Wales, 2018-2019). All children attended full time Welsh-medium education.

Home language background

Whilst the aim was to include only those children for whom Welsh was their L1 and the language spoken to them by each/their parent in the home, as is often the case in bilingual settings, the sample included children with varying degrees of exposure to Welsh (see also Chapter 2). Moreover, past research (e.g. Bellin, 1988) has shown that L1 Welsh children use their receptive knowledge of mutation as gained through their environment when engaged in formal tasks whereas L2 Welsh speakers tend to apply the formal rules they have learnt whilst learning the language in a more structured way. Also, even in Welsh-medium schools in the most Welsh-speaking regions of Wales, classrooms include a mixture of bilinguals with varying degrees of exposure to Welsh. For that reason, it was important to include both cohorts in the studies and to separate the two for the purpose of analysis. The sample was therefore divided into two separate language groups, which are described below. These divisions are based loosely around those applied by Gathercole & Thomas (2007). The categorisation of both language groups were based solely on what parents had selected as their child's first language (Welsh, English or Welsh and English), the home language of the parent was not considered within this categorisation. As children were all attending Welsh medium primary schools their level of exposure to Welsh at school would be relatively equal, this decision to categorise in this way was therefore based solely on parental views of their child's first language as parents would have a more rounded view of their child's language use.

Home language Group 1: Mostly Welsh.

This group consisted of children who had been categorised as L1 Welsh by their parents according to the parental questionnaire. Most children in this category (82.35%) were exposed exclusively to Welsh either in two parent or single parent households. A further 9.8% were exposed mostly to Welsh or bilingually to both languages by one parent and exclusively to Welsh by the other. The remaining 1.96% included instances where Welsh was used

alongside English by one parent only, with the other parent speaking exclusively in English (see figure 2 for breakdown of home language exposure). Where information had been communicated by the parent it appeared that only one child within Group 1 was in receipt of FSM. Figure 2 below provides a breakdown of Group 1 language exposure within the home.
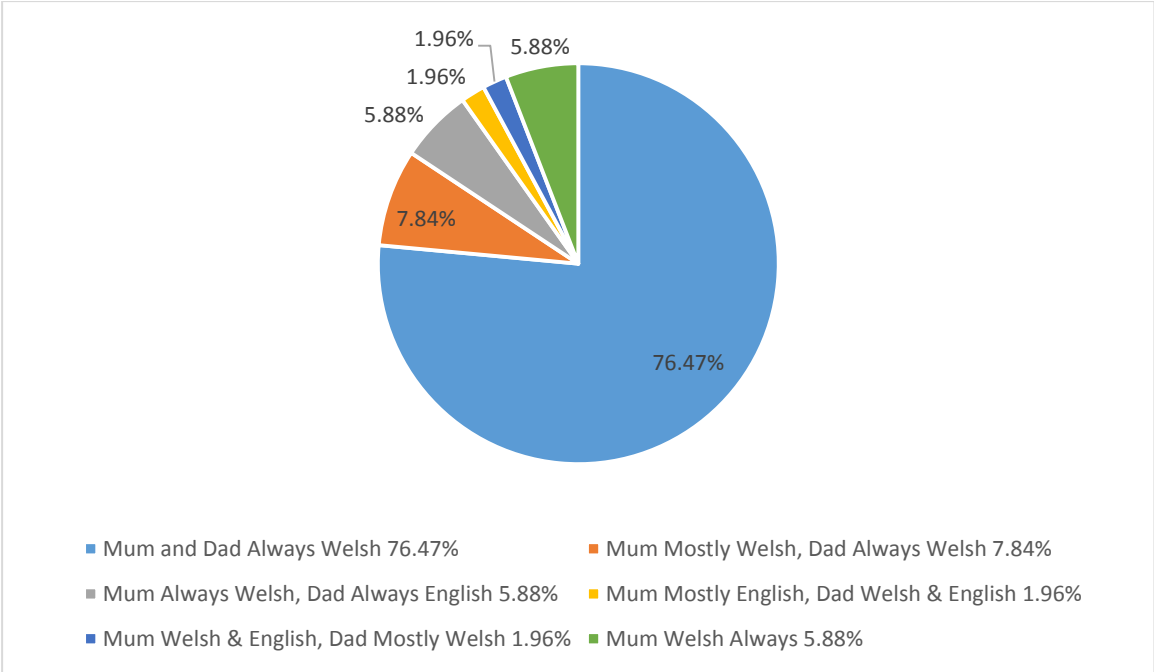
Figure 2 Home language exposure within Group 1 (Mostly Welsh).

Home Language Group 2: Mostly English.

Home language Group 2 consisted of 15 participants who had been categorised by parents as L1 English or 2L1 Welsh/English. The majority of parents of the children within this group always or mostly spoke English at home (60.01%), 13.34% came from homes where one parent mostly spoke Welsh and 26.67% lived in homes where one parent spoke both English and Welsh (see figure 3 for detailed breakdown of language exposure at home). Only one participant was recorded as being in receipt of FSM within this group.
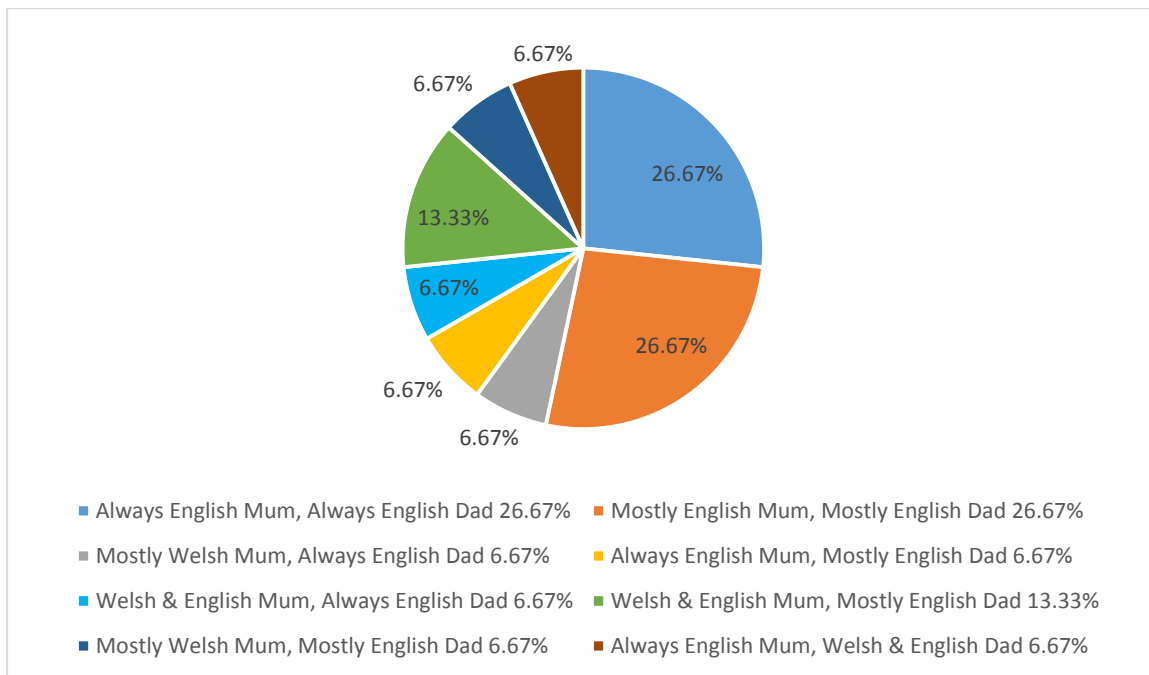
Figure 3 Home language exposure within Group 2 (Mostly English)

Where appropriate, and necessary to answer the research question under investigation, data were analysed by groups in Chapters 5 and 6.

Background Measures

Home language questionnaire

Parental questionnaires were included in the information pack sent to parents regarding the study prior to the commencement of the study (see Appendix 6). The language background questionnaire provided data regarding participants' language background including the parents' first language, the language spoken by the mother and father at home (where applicable), the language the parent regarded as the child's first language, the mother and father's occupation and whether the child was in receipt of free school meals. The completed questionnaires also provided information regarding whether the child had a diagnosis of dyslexia or any other impairment.

Standardised measures of background variables

Before taking part in the experimental conditions of Study 2 and Study 3, each child completed a set of standardised tests to gauge their abilities in relation to the following:

(i)   Reading ability: All Wales Reading Test (Forbes, 1999), National Reading Tests (Welsh Government, 2013b)

(ii)   Short term memory: CTOPP-2 (Wagner, Torgesen, Rashotte & Pearson, 2013)

(iii)   Non-verbal IQ (adapted from Raven, 1998)

(iv)   Phonological Awareness (English and Welsh battery of tests)

The results of these tests were then compared to the children's performance on the experimental tasks, as outlined in Chapters 5 and 6. Each standardised measure is described in more detail below.

(i) Reading ability

Due to the population under study and to the nature of the research questions under investigation, it was essential that all children were measured for their literacy abilities in both languages. In the absence of the ability to distinguish between those with dyslexia and typically developing matched sample peers, having standardised measures of literacy abilities would make it possible to distinguish between those who are good and those who are poor readers. In line with Thordardottir *et al.* (2006) and others' observations that the testing of bi/multilinguals using monolingual normed standardised tests or only testing in one of the individual's languages may lead to under or over-representation of language problems, obtaining standardised measures of children's literacy abilities in Welsh and in English was essential for the present studies. There is a standardised literacy task that is available through the medium of Welsh – *Prawf Glannau Menai* (Payne, 1998) – that was modelled on the NARA standardised assessment of literacy in English. However, as noted in Study 1, the PGM Accuracy scores did not correlate well with NARA

Accuracy or NARA Comprehension scores, which does call into question its validity as a measure of accuracy in Welsh. Also, children have become familiar with its content and so their performance on the tests may not be an accurate reflection of their abilities. For these reasons, it was decided to look for an alternative for Studies 2 and 3. These alternatives came in the form of the All Wales Reading Test and the National Reading Test. Both are detailed below.

All Wales Reading Test (AWRT)

Due to its availability in both Welsh and English, the *All Wales Reading Test* (Forbes, 1999) was used as a measure of reading ability. Unlike the PGM, the AWRT involved a large sampling frame comprising of children across Wales, strengthening its validity as a standardised tool. This test, created relatively recently in 1999 and was developed with the help of schools and pupils across Wales, is largely a reading comprehension task. The test is administered differently across different age groups. Younger children (age 7 – 8) begin the test by circling the correct word to match the image on the paper. This ensures that the children know the meaning of words. This then develops into a cloze exercise where children select the correct word to complete the sentence. Sentences become gradually more difficult as children progress through the test. For older children the test is based solely on a cloze exercise; however, sentences and vocabulary become more complex as the test progresses. This test was administered in groups of no more than 20 when the researcher was supported by another adult and was administered under exam conditions.

National Reading Tests (NRT)

The National Reading Tests (NRT) are standardised tests that were introduced in 2013 (Welsh Government, 2013b) to assist teachers with formal assessment of children's reading abilities in order to inform their learning. The tests are based on the skills found within the National Literacy and Numeracy

Framework (a test is also administered based on numeracy) for Years 2 to 9. There is a test administered in English and a separate test in Welsh. These are unique tests and not direct translations. The content of the test is presented as a piece of text followed by a number of questions related to the text which become more difficult as the test progresses. The items covered within the test for Years 2-3 (age 6-8) include:

- Sentence completion
- Multiple choice
- Find and copy
- Matching
- Sequencing
- True or false

Years 4-6 (age 8-11) include:

- Matching
- Find and copy
- Multiple choice
- Cloze passage
- Underline
- Sequencing
- True or False

Sample questions are provided at the beginning of the test to enable children to practise and understand what they are being asked to do and familiarise themselves with the types of questions that will be asked during the test. The test takes up to an hour to complete and younger children are provided with the option of a break part-way through. Children in Years 2-3 in Welsh-medium schools complete the test in Welsh only; however children in Year 3 may be asked to complete the test in English also, depending on the school. A very small number of children within this study for whom test data were available had not completed the test in English. Reading scores were also not

available for some of the participants and so the sample size for NRT data is smaller than the AWRT in the present set of studies presented in this thesis.

The reading tests were only used to provide a picture of children's reading abilities in order to differentiate between lower and higher ability readers within the sample.

(ii) Short term memory (CTOPP-2).

In order to assess short term memory, participants were asked to complete a memory for digits test. This test was administered to the children in English, on a one to one basis using the Comprehensive Test of Phonological Processing (CTOPP-2). As no standardised test of short term memory in Welsh currently exists, it was decided that an English-based test was sufficient as a guideline of participants' short term memory. Although the majority of participants spoke Welsh within the sample, it was assumed that they had an awareness of numbers 1-9 in English due to their age and stage of education and no impairments regarding understanding of numbers in English had been declared by the parents or teachers. Language was therefore not regarded as a barrier within this particular task. The task involved listening to set sequences of numbers on a CD ranging from 2 numbers to 9 numbers and repeating the numbers exactly as they had been heard. The first 4 items were practice items to ensure the participant was clear on how to complete the task. Participants were only able to listen to each CD track once and were instructed to listen very carefully before repeating the sequence aloud. The task became progressively more difficult as the sequences of digits increased in length. Once participants had scored three incorrect answers in a row the task was discontinued. Raw scores were then calculated and added to a database.

(iii) Non-verbal IQ.

A test of non-verbal IQ was conducted at the beginning of the study. Participants were tested using the Ravens Standard Progressive Matrices (Raven, 1998). This test provides a brief nonverbal screening measure of general ability. Participants were asked to complete a series of diagrammatic puzzles by selecting the correct tile to complete the picture using visual clues within the puzzle. These puzzles became progressively more difficult. Participants were given a subset of the full IQ test (approximately half of the puzzles which are administered in the full version of the test) in order to keep testing time to a minimum and because the aim was to gain a general indication of the participants' IQ. Participants were tested in groups of less than eight when a single researcher was present and no more than 14 when the researcher was supported by another adult. This was in line with the guidelines for administration of the test provided by Ravens (Raven, 1998). They were seated strategically to minimise copying, and were supervised by the researcher to ensure no copying took place. Testing took approximately 20-30 minutes to complete.

(iv) Phonological Awareness.

As noted in Chapter 2, a large body of previous research has identified a strong link between phonological awareness and dyslexia in English-speaking individuals. In order to be able to explore potential links between phonological awareness and children's abilities with Welsh mutation and plural morphology – two systems that rely heavily on good phonological awareness abilities - it was important to measure children's phonological awareness skills. This was achieved through the use of a selection of existing tasks from the CTOPP-2 in English, and a combination of tasks in Welsh from the ELDEL[9] Project Battery (Caravolas, 2011). These different tasks are described below.

CTOPP2 was selected based on discussions with the Miles Dyslexia Centre as a basic measure for phonological awareness. The subtests identified within the

---

[9] This now forms part of the MABEL multilingual assessment battery of early literacy https://www.eldel-mabel.net/

CTOPP2 as a measure of phonological awareness for the age group being tested were Elision, Blending words and Phoneme isolation. As only a basic measure of phonological awareness was required for this research these three tests were sufficient in achieving that measure. Due to children also completing a number of other tasks as part of this research a decision was made not to ask children to complete further, time consuming tasks to gain more in depth phonological ability scores that were not required at this time.

Letter sound recognition.

From a teacher's perspective letter recognition is regularly used in classrooms to gauge children's ability to recognise letters when learning to read. It was decided therefore that children complete a very short task to identify letter sounds before moving on to more challenging phonological awareness tasks. This was to measure their basic phonological abilities before moving to standardised measures of phonological awareness, in this case the CTOPP2 and ELDEL. Other studies such as Caravolas *et al.* (2013) have also included letter sound knowledge as a measure of phonological awareness in children. Children's knowledge of the letter sounds of the alphabet in English and Welsh was measured initially using flashcards (this task was not taken from the ELDEL or CTOPP-2 battery of tests). Children were shown each letter of the alphabet individually and asked to say the *sound* of the letter in English. The same instruction was given in Welsh, asking children to identify the letter sounds in Welsh. During the test phase it was observed that many children did not understand the term 'sound of the letter' and would immediately state the letter name when they were shown the flashcards in English. They were corrected if they made this error in the beginning but their answer was marked as an error if this was done repeatedly following clear instructions. The letter 'Q' was accepted in its name form as it became apparent that most children were not familiar with this letter sound in English (/k/) and knew it only as the name of the letter. All other letters were recorded as errors if not correctly identified as a letter sound. When administering these tasks the language of administration was changed according to the language of the task. As children

were attending Welsh-medium schools the Welsh letter recognition task was administered first. The children were then told to 'switch their brains into English mode because we are now going to do the next task in English' (cf. Grosjean's 1998 principle of 'language mode'). The English letter recognition task was then administered in English.

Comprehensive Test of Phonological Processing 2 (CTOPP-2)

The CTOPP-2 is fundamentally a measure of children's phonological processing skills and is used as a diagnostic tool in the assessment of dyslexia. It seemed logical therefore that this test be used to gain a measure of the participants' phonological awareness in English within the current study. The following tests were administered for the purposes of identifying the participants' phonological processing skills:

1. *Elision.*

    In this test participants were asked to repeat a full word and then asked to say what was remaining when part of that word was removed e.g.

    Researcher: *Say toothbrush*.
    Child: Toothbrush.
    Researcher: *Now say toothbrush without saying tooth*
    Child: brush

    These words become progressively smaller so participants needed to use their phonological skills to decipher what word was remaining when letters were removed from various parts of a word e.g.

    Researcher: Say fixed.
    Child: Fixed.
    Researcher: Now say fixed without saying /k/
    Child: Fist.

There were 34 items in this test, and ceiling was reached once participants had missed three items in a row.

2. *Blending words*

This test involved the use of pre-recordings of particular words and letter sounds on a CD. Participants were asked to listen to parts of a word and then say what word is produced once those different parts are put together. For example, the first item on the test states 'what word do these sounds make? cow-boy /kaʊ bɔɪ/', the answer is 'cowboy'. The parts of the word become gradually smaller and the words become longer as the test continues. There were some limitations to this task, however. The CD items were at times unclear for the children due to the American accent of the speaker and clarity of the CD, also the nasals /m/ and /n/ sound were at times unclear. Participants were able to hear the word up to two times within this test. If, however, it was unclear to them (due to the recording or accent) they were given the option for the researcher to repeat the sounds at the same pace as the recording. The issue regarding the clarity of the /m/ and /n/ was also explained to the participant before beginning the test. There were 33 items within this test and testing stopped after three items in a row were missed/incorrect.

3. *Phoneme isolation*

This test involved participants listening to a word read out by the researcher and identifying sounds within different parts of a word, for example, first sound, middle sound, last sound. For longer words the second, third or fourth sound were also included. The practice items asked the participant to listen carefully to each individual sound within the word before the sounds were joined together to create a whole word. For example,

The word 'frog' has four sounds /f/-/r/-/ɔ/-/g/

What is the third sound in the word frog?

(Answer: /ɔ/)

There were 32 items in this test, and testing was stopped after three items were missed/incorrect in a row. A raw score was recorded for each of the phonological awareness tests.

ELDEL Project Battery (Caravolas, 2011)

Given the importance, as noted above, of ensuring that the children were tested in both languages (see Chapter 2), it was necessary to develop tests aimed at measuring their phonological abilities in Welsh as well as in English in order to gain a more holistic view of their phonological abilities. As no formal tests exist for phonological awareness in Welsh, the ELDEL project (Caravolas *et al.*, 2011), a test being developed within the School of Psychology in partnership with the Miles Dyslexia Centre at Bangor University, was used as a measure of phonological awareness in Welsh. Some of the items were altered slightly for the purpose of the present study, for example, the example non-words provided in the instructions to the non-word tasks (e.g. for the Phoneme deletion task) included non-words that were more phonotactically related to English than to Welsh. These were therefore changed to be more nuanced towards Welsh. For example *Tash* (which could be argued to represent an English slang word for moustache, but also includes the digraph *sh* that does not occur in Welsh) was changed to *Tach* and *Nash* was changed to *Neth*. Within the non-word phoneme deletion task itself a circumflex was added to *dwl* (i.e. *dŵl* /du:l/, since *dwl* is a real word in Welsh meaning 'dull' or 'stupid') and *fwn* (*fŵn*/vu:n/) to provide a more nuanced aspect to the task. An *n* replaced the *m* in the word *trim*[10] as this would be recognised as an English word rather than a non-word. Also, within the phoneme blending task

---

[10] Whilst there does exist a Welsh word 'trin' ('to treat') where the 'I' is elongated /i:/, the aim here was to pronounce the word with a short 'I' /ɪ/ and given that the researcher read out the items, this was not deemed to have caused any issues.

the word *ci* /kɪ/ 'dog' had been repeated within the items and so was removed, the same applied to the word *torri* /torɪ/ 'to break'. Although not a standardised test as yet, together, the administration of the tests provided an insight into children's phonological abilities in Welsh, testing their ability to identify initial and final sounds within non-words, phoneme deletion (initial and final) of non-words and phoneme blending. These tasks are described below.

1. *Phoneme isolation - Initial/final sound*

    An explanation was given to the participant about non-words within the task and they were asked to listen to the word being read out by the researcher and then to state what they thought the first/last sound of the word was. The first set of non-words were in the CVC format (8 words) e.g. initial sound - *sec* /s-ɛ-k/ final sound – *neg* /ŋ-ɛ-g/ and the second set in CCVC or CVCC format (8 words) e.g. initial sound - *crof* /k-r-o-v/ final sound – *dolff* /d-o-l-f/.

2. *Phoneme deletion*

    A set of 10 non-words were read out one at a time and the participant was asked to repeat the word and then state what the word would be if the first sound was removed e.g. *roth*/rɔə/ – *oth*/ɔə/. The task was then repeated with another set of 10 words but participants were asked to remove the final sound and state what word would remain, e.g. *sont* /sɔnt / – *son* /sɔn /. Practise items were used before each task began to ensure participants understood the task.

3. *Phoneme blending*

    This task consisted of a set of 22 words with a maximum of four sounds. The researcher would say each of the sounds aloud at a rate of approximately one sound per second and children were then asked to

say what word they thought the sounds made when they were blended together. E.g. *c-a-th – cath*/kæθ/ 'cat'. All phonological awareness tasks were administered on a one to one basis in the original language of the test. A raw score was recorded for each of the phonological awareness tasks.

Language of Instruction

During data collection the language of instruction for each of the tasks changed according to the language of the particular task being administered. That is, all tasks exploring knowledge of English were presented in English and all tasks exploring knowledge of Welsh were presented in Welsh. This ensured that the children were thinking about the items presented in a given task in the language of that task (Grosjean, 1998). As noted above, all of the participants attended Welsh-medium schools, where they were exposed to Welsh on a daily basis. All children were therefore able to understand any instructions given in Welsh. The same applied to English. As children were exposed to English either at school, at home or in the community, they were familiar enough with English and therefore able to follow the English instructions given. Where possible, visual aids were used to reinforce understanding of the tasks and participants were given ample opportunities to ask questions in either language if they were unclear about the task.

Task administration

The tasks were administered to children per class and were generally administered in the following order:

IQ (Groups)
AWRT Welsh (Where applicable) (Groups)
AWRT English (Where applicable) (Groups)
letter recognition Welsh and English/ Memory for digits (administered during the same session) (Individual)

CTOPP-2 (English phonological awareness tasks) (Individual)

ELDEL (Welsh phonological awareness tasks) (Individual)

Plurals Tasks 1 and 2 (Group)

Mutations Task 1 and Task 3 (a and b) (Pairs)

Mutations Task 2 (Individual)

Mutations Task 4 (a and b) and Plurals Task 3 (Individual)

The order of administration of tasks varied at times due to participant absence on days where a particular task was being administered and therefore they would complete it on a different day whenever possible. The administration of tasks also depended on the classroom/school timetable which could at times vary depending on what was happening within the school on a given day. Children were called by the teacher or the researcher/research assistant to be taken to the testing area (classroom or spare room nearby) either individually, in twos or in groups depending on the type of task to be completed. The testing room was relatively consistent although had to change at times when there was no availability. Times for visiting the schools were pre-arranged with the headteacher or key point of contact at the school. Steps were taken to ensure that children were not taken out of the classroom during periods when it was important that they continued with classroom based work. Each of the testing sessions ranged from 10-35 minutes. Tasks were administered in the main by myself with some support from a research assistant at one school. The research assistant administered some of the phonological awareness tasks due to these being more time consuming and assisted with the supervision of the group administered tasks (i.e. IQ and reading).

The tasks were completed in stages, for example, whenever possible Mutations Task 1 and Task 3 would need to be completed by all participants at the school before moving on to Task 2 and so on.

The NRT scores were obtained from the schools (where possible) and so had not been administered as part of this research. These took place at around the same time as the other tasks being assessed (at least within a month).

The AWRT had been administered by the researcher for 2 schools with the exception of approximately 5 children who had recently completed the tests (3 months prior) within schools and so scores had been passed on to the researcher. One school administered the AWRT themselves as part of their own assessment processes and so scores had been passed on to the researcher (approximately 18 children).

Ethical considerations

An initial letter was sent to a number of schools across North Wales outlining the research and instructing any schools who were interested in taking part to send an email confirming their interest in the project. The recruitment focused primarily on children for whom Welsh was a first language and so schools where 70%+ of the children attending spoke Welsh as a first language were targeted. This information was provided by the local schools consortia GWE.

During the recruitment period an information pack was sent to parents of all 7- to 11-year-olds at those schools who had agreed to take part in the study. The information pack included information for parents about the study, consent forms for audio recording and general consent for taking part in the study (see appendices 1 to 6). Parental questionnaires were also included as part of the information pack to ascertain information regarding each participant's language and socio-economic background (see descriptions earlier in the chapter). Following receipt of this consent, a follow up visit to the school was arranged to speak to the children in groups regarding the study aims, and opportunities were provided for children to ask any questions they had. Children were shown an information sheet that was read to them in the language they felt most comfortable with and they were then asked to read the statements at the back of the information sheets (with assistance where applicable) and draw a smiley face next to the statement to confirm that they agreed and were happy to take part in the study (see appendix 5). Children

also wrote their name and school at the bottom of the sheet to confirm their assent to take part in the study. Children were fully informed of their right to refuse to take part and it was made clear that they could pull out of the study at any time without providing a reason. A temporary research support assistant was recruited to assist with data collection at one of the schools. This individual was an L1 Welsh speaker with research experience and children were introduced to the assistant during the information for children session, so they were fully aware of who they would be working with when completing the tasks. The research assistant was briefed on the need for confidentiality and to store any data collected securely until it was passed to me (usually the same or following day).

Following these visits two children (male) decided that they did not want to take part in the study, and so their names were removed from the data. Throughout data collection children were asked if they were happy to take part in each task before they begun and were also reminded of their right to withdraw. All data were stored securely.

PILOT STUDY

A small pilot study was conducted to test the usability of the new computer-based measures and paper-based plural tasks. This included measuring time on task, clarity of instructions and whether participants were able to complete the task independently. Following completion of the tasks by the pilot participants, the participants were asked about their thoughts on the measures and any difficulties they encountered.

Participants

Participants for the pilot were recruited independently and consisted of three typically developing children from working/middle class backgrounds. The children were from predominantly Welsh-speaking homes aged between 8 and 11. All participants attended Welsh-Medium Primary schools in North Wales.

Parents were given a participant information sheet explaining the research (see Appendix 2) and consent for participation in the research as well as audio-recording consent was sought. Separate assent from the children was also gained. The research was explained to the children in the form of a simplified participant information sheet written in an age appropriate way and was read out by the researcher to the participant before beginning the tasks (see Appendix 5). Participants were given every opportunity to ask questions regarding the tasks and the research in general. They were then asked to complete an assent sheet drawing a smiley face if they agreed with the statements and then writing their name at the end of the form. No data were collected without both consent and assent. The tasks were completed at a place most convenient to the children's parents and in a quiet room near to their parents with no distractions. All tasks were completed in one sitting with children being offered regular breaks. Tasks took a maximum of 1 hour to complete for each child.

Method

The pilot study took place at the University or in the children's home. Parents were close by during testing at both locations.

The study began with the Plurals Tasks 1 and 2 (see Appendix 9) paper-based tasks. These measures were designed to be completed independently by the children following an explanation on how to complete each task. The first two items of each task were used as examples to ensure the participants were clear on the aim of the task. Once it was clear that participants understood the exercise they were given time to complete it independently under supervision.

The second stage of the study involved the use of the computer based tasks based on the Welsh pirate story *Y Goriad* ('The Key') which included a multiple choice real word cloze task, multiple choice non-word cloze task, Grammaticality judgement task and two oral tasks (one sentence repetition

(mutation) task and one plural production task). Full details of these tasks are provided in the next chapter and examples can be seen in appendices 8 and 9.

Observations

Paper-based plural tasks

The pilot test revealed the plural tasks to be very user-friendly and children were able to complete them independently once they had been given clear instructions. This result meant that the tasks could be distributed in small groups, under exam conditions, due to participants being able to complete them independently. However, one adaptation was made to Plural Task 1 following the pilot study. It became apparent that participants were not confident in the plural form of the first word presented – *ffon* 'stick'. This was therefore moved to a later position within the task and an easier word placed at the beginning which all children would have been familiar with – *iâr* 'chicken/hen' – *ieir* 'hens'. This was to ensure participants were not discouraged at the beginning of the task and would therefore be more inclined to continue if they felt more confident from the outset.

Mutation Task 1 – Multiple-choice cloze task: real words

Once instructions had been provided at the beginning of the task, the participants appeared able to complete this task with very little additional guidance. As a result, no changes were made to this task.

Mutation Task 2 – Multiple-choice cloze task: non-words

This task seemed to be fairly challenging due to the task tapping into a completely new experience for the children (i.e. the mutation of non-words). Although a full explanation was given prior to beginning the task, participants were not always clear on what they needed to do because the task was unlike

anything they would have done before. A practice sheet was therefore created for participants to attempt the exercise informally first with the researcher. This involved completing three practice items, with guidance, to ensure they were grasping what they were being asked to do. This practice sheet was then used in the final version of the task.

Mutation Task 3a and b – Grammaticality judgement task

Participants appeared to complete this task and follow the instructions given without difficulty. No changes were made therefore to this task.

Mutation Task 4a and b (Sentence repetition and sentence correction task)

Following clear instructions, participants appeared to complete this task with very little guidance. Although participants were able to use the recording function themselves during the pilot, it was decided that this would be done by the researcher during the final study to avoid any unnecessary issues with regards to recording. It was also decided that to enable children to focus on holding the microphone correctly and listen to the recording, the researcher would click on the treasure chests instead of the participant. This decision was also due to time restrictions within the study to minimise length of time that children would be absent from class.

Plurals Task 3 oral production task

Participants appeared to understand the task clearly following the instructions given. It was highlighted that some participants were not sure of the plural word and would therefore sit in silence. To avoid this during the final study and to avoid unnecessary time out of the classroom, it was decided that participants would be asked to say 'pass' into the microphone if they were not sure of the answer. They were then able to swiftly move on to the next word.

The pilot revealed each task to take between 5 – 20 minutes. This was taken into consideration for the main study to ensure participants spent no longer than 30 minutes at a time on given tasks, as outlined in the participant information sheets given to parents/children.

Implications of the Pilot study findings for Studies 2 and 3

The pilot study findings led to the following implications for Study 2:

- The order of administration of tasks was altered to accommodate time restrictions and availability of participants.

It was important that the *Multiple-choice cloze task: real words (Task 1)* always be administered first to ensure the context of the story was clearly set as this task was incorporated into the main body of the story and therefore if not administered first the other tasks would make very little sense to participants.

- The Multiple-choice cloze task: real words (Task 1) and the Grammaticality judgement task (Task 3) could be administered in pairs as these tasks could be completed fairly independently once instructions had been given.
- Due to its more challenging nature, the *Multiple-choice cloze task: non-words* (Task 2) task needed to be administered on an individual basis to ensure children had as clear an understanding as possible of the task.
- The Sentence repetition (Task 4a), sentence correction (Task 4b) and plural oral production (Plurals Task 3) would be administered together and on an individual basis, in as quiet an area as possible to ensure no distractions or interference when recording.
- The paper based Plurals tasks would be administered in groups of no more than 15 due to children's ability to complete them independently once instructions had been given.

- A practice sheet for the *Multiple-choice cloze task: non-words* task (Task 2) was also created for the final study to ensure participants could try practice items before completing the Task on the computer.

- The addition of the instruction to say 'pass' into the microphone if unsure of the answer during the plural production task was also included to ensure time was efficiently used.

- The order of the items on Plurals Task 1 would also be adjusted to ensure participants were not discouraged if they could not answer the first question. A simpler item was placed at the beginning of the task.

The next Chapter (5) presents the findings of Study 2, involving children's performance on Welsh mutation tasks and Chapter 6 presents the findings of Study 3, involving children's performance on the Welsh plural tasks. Chapter 7 presents an overall discussion of the findings of Studies 1-3 in relation to the wider literature, and presents recommendations for practice and future research avenues.

CHAPTER 5

STUDY 2

Exploring the relationship between Welsh-English bilingual children's performance on Welsh mutation tasks and tests of reading ability.

This chapter presents a description of each of the novel tasks that were designed specifically for this study. This is followed by a results section, which presents the findings of Study 2 in relation to the research questions under investigation. The chapter ends with a discussion of the findings.

Introduction

In the previous chapter the background measures for Studies 2 and 3 were described as well as the rationale for each study. The focus of this chapter will be on Study 2, which explores the relationship between children's performance on novel mutation tasks and their reading scores on both the All Wales Reading Test (AWRT) in English and Welsh and the National Reading Tests (NRT), also in English and Welsh. The relationship between the background measures used (phonological awareness, IQ, age and short term memory) and performance on the mutation tasks is also explored. Given there are no other research studies that explore mutation in this way, the study takes an explorative approach to investigate what the relationships are between mutation, reading and the background measures as listed above.

The chapter will cover each of the tests carried out as part of Study 2 which includes a multiple choice real word cloze task (Task 1), a multiple choice non-words cloze task (Task 2), a grammaticality judgement task (Task 3a), a sentence correction task (Task 3b and 4b) and a sentence repetition task (Task 4a). Each of the novel tasks relate to the mutation system in Welsh. Due to the absence of appropriate assessment tools in Welsh it was necessary to produce original tests to explore the research questions under investigation.

It was decided that computer-based tasks would be most appropriate given that the design would be more appealing to children and would be presented more as a game/activity rather than a test. This would enable children to feel more relaxed whilst completing the task and therefore able to work at a pace they were comfortable with. Computerised tasks also enable quick access to quantitative data to avoid the researcher being seen to 'mark' the children's responses in front of them. The use of computer based tasks in language learning is currently increasing and research has shown benefits of using computer based tasks in children's learning (see e.g. Reinhardt, 2018 and Butler, 2015). A computer-based task also provides a useful tool for teachers within a classroom to enable children to develop their learning independently and provide teachers with direct access to their ability scores.

The aim of this study was to answer the following research questions:

- Is there a relationship between children's performance on morpho-phonological tasks and their reading abilities? If so, what is the nature of this relationship and does it vary between language groups?
- Can tests of morpho-phonological awareness predict children's performance on reading tests?

Regarding Research Question 1, it was predicted that those with higher reading abilities and/or phonological awareness abilities would perform better on tasks requiring knowledge of mutation than those with lower reading and/or phonological awareness skills. It was also predicted that the nature of this relationship would vary between language groups due to the varying levels of exposure to Welsh potentially affecting their performance on tasks.

In relation to Research Question 2, assuming that there is a link between reading scores and/or phonological awareness abilities and mutation due to the high phonological awareness load incurred by the mutation system, it was predicted that performance on mutation tasks would predict children's performance on reading.

Measures

Study 2 Novel Experimental Tasks: Mutation

A total of four (non-standardised) mutation tasks and one plural morphology task (see Chapter 6 for results of plural tasks) were created using an original, computer based format which had been created specifically for Studies 2 and 3. The software used for the design of these tasks was Adobe Director v12 with Audacity used for the recording of the audio used for the characters. I designed each page using PowerPoint and the idea was inspired by a study of sentence repetition where Powerpoint slides were used within the test design (see Marinis and Armon-Lotem 2015). The PowerPoint design was then converted into an interactive computer-based task with the support of an IT specialist using Adobe Director v12. The voices of the two central characters were pre-recorded using two native Welsh speakers. These five tests were all linked to a central pirate story – *Y Goriad* 'The Key' – that was based around two central characters (brother and sister Ben and Ceri) who enjoy venturing to the beach to explore. One task was created as a second part to Task 4 (Task 4b) and a further set of three tasks (Plurals Task 1, Plurals Task 2a and 2b) were created to elicit children's knowledge of plural morphology in Welsh. These were not part of the computer-based suite. To measure children's knowledge and application of mutation in Welsh a series of tasks were designed that explored mutation in different ways (these are described in more detail below). Task 1 and 2 were multiple choice cloze tasks. Cloze tasks were chosen due to being tests that are regularly used within language and literacy assessment and therefore would be familiar to the children completing the tasks. As outlined in Chapter 2, cloze tasks provide useful information about a child's reading abilities and can be used in bilingual assessment (see Trace, 2020). For Task 1 I decided to use the format of cloze/gap fill tasks and apply it to mutation to explore whether children could identify the correctly mutated word to complete the sentence. This concept was also used within Task 2 in relation to being able to apply mutation correctly to non-words.

Task 3a was a grammaticality judgement task which tapped into children's ability to identify whether a sentence was grammatically correct in relation to mutation. Task 3b supported Task 3a in assessing whether children could identify the incorrect word in those sentences which were incorrect. These types of assessment are regularly used within education to assess children's grammar abilities. Task 4a was a sentence repetition task which involves children using their knowledge of the grammatical structure of a sentence in order to repeat it (see Chapter 2 for further detail regarding sentence repetition tasks). This task was used to ensure an oral production element was included within assessment to avoid over-reliance on reading to assess children's abilities. Task 4b supported Task 4a by assessing whether children could identify the incorrectly mutated word in a sentence they had repeated aloud, and also produce the correctly mutated version of that word. The tasks used in this study are based on typical language testing methodologies and aim to tap into children's morpho-phonological awareness. There have been no previous research where these tasks have been used to assess Welsh mutation in this way. It is important therefore to note this is an exploratory study and the findings should be treated with that in mind.

The findings of the plural morphology tasks will be presented and discussed in the next chapter. The design of the computer-based story is presented below.

Y Goriad

Visual design.

The computer-based design began with a written introduction on screen (in Welsh) explaining to the children that there are a range of puzzles to complete. As they read through the story they would be asked to fill the gaps on the screen using real words or non-words and that all the puzzles were related to mutations or plurals in Welsh. The introduction then went on to remind the children that if they decided that they did not want to take part they were welcome to tell us at any time. It also stated that if they did not

understand a puzzle they could ask any questions and that if they wanted to take a break at any time to just ask the researcher. They were then asked if they had any questions before starting. This introduction was read out by the researcher before the tasks began to check for understanding from the children.

The remainder of the story was laid out in a cartoon type arrangement, with the main characters on the screen and written text next to the characters if they were speaking. A sound icon was present on the screen to enable children to listen to the narrative of the story if they wished, if the characters were speaking then children needed to click on the character to hear the speech. This was explained to all children before they began the task. Their attention was also drawn to the *nesaf* ('next') button to enable them to move to the next page. They were reminded of the importance of not clicking the *nesaf* button until they had attempted all the items on the screen as there was no back button to enable them to go back a page. Although clear instructions had been given on this there were occasions where children did click on the *nesaf* button before they had completed the task. This happened approximately 8 times in total although there is a possibility it may have happened without the researcher being informed. Also, there were occasions (approximately 21 in total across all tasks) where children wanted to change the answer they had given. In such cases, the error, and what they wanted to change the answer to, were noted and manually changed at a later date. This was a limitation of the test as it was dependent on the child informing the researcher of the mistake.

Narrative leading to the tasks.

The story begins when they come across a key in the sand, which Ben is convinced belongs to a pirate ship. Ceri brushes off this thought and convinces Ben to leave the beach and return home. Ben, however, is unsettled whilst at home and decides to sneak out to the beach that night. Ceri catches him on his way out and decides to go with him. Whilst on the beach the moon appears

from behind a cloud and a door appears in the sand. The children realise that this is what the key must be for, to open the magical door. The children open it and soon find themselves on a pirate ship and face to face with Captain Caradog. They discover that these are no ordinary pirates, they are a type of hybrid human/monster pirate from a faraway land. The novel names given to the pirates *pirallau*/pɪrɑːɬɑːi/ from the *Goffugion*/gɔfəgɪɔn/ tribe introduce non-words into the story, in order that the participants are able to engage meaningfully with non-words in some tasks. The story continues to a point where Captain Caradog will not allow them to leave the ship until they have completed all the tasks. At this point participants reading the story realise that they are completing the tasks in order to help Ben and Ceri get back home. The test then leads the participant on to the different tasks, most of which are linked to the story in some way. Once all tasks are completed, the participants are shown the end of the story where Ben and Ceri are able to return home thanks to their help in completing the tasks set by Captain Caradog.

The computer-based tests were administered on a one-to-one or two-to-one basis depending on the nature of the task. All computer-based tasks were in Welsh and so children were spoken to, and instructions were given in Welsh at all times in order that the children, being bilingual, remained in Welsh 'mode' for the duration of the testing (Grosjean, 1998). The four mutation tests that formed the basis of Study 2 are described below.

Example pages from 'Y Goriad'

Ceri   Ben

Roedd Ben a'i chwaer Ceri wrth eu boddau yn mynd lawr i'r traeth. Byddan nhw'n treulio oriau yn chwarae yn y tywod ac yn chwilio am drysor.

NESAF

Un bore, aethon nhw i lawr i'r traeth fel arfer i chwarae ond y tro yma cawson nhw hyd i oriad yn cuddio ymysg y creigiau.

NESAF
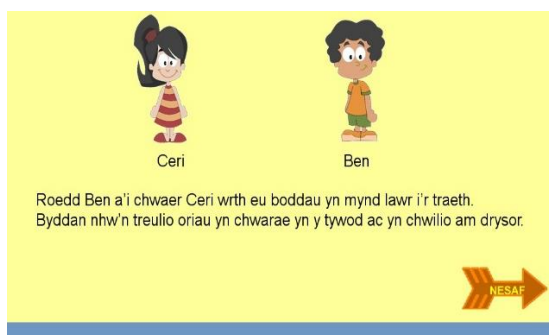
Linguistic stimuli involved in each task

Task 1 – Multiple choice real words cloze task

Task 1 was a multiple choice cloze task based on the story 'Y Goriad'. The focus of Task 1 was Soft Mutation, since this is the most prevalent mutation in the language, and the one that is used most commonly and most regularly by speakers (Ball & Müller, 1992). Children were asked to select the word (by clicking on the word and dragging it to the gap within the sentence) from a selection presented on the screen that they felt best completed a given sentence. This was a multiple choice task where children chose from a non-mutated word, a correctly mutated word or an incorrectly mutated word (see Examples 1 and 2 below). There were 24 items to complete in the task. At least one example of each sound/letter that undergoes Soft Mutation (sounds: /p, t, k, b, d, g, ɬ, r̥, m/; letters: p, t, c, b, d, g, ll, rh, m – see Chapter 3) was included within the task. 19 target words were preceded by a trigger to Soft Mutation and 5 were preceded by words that do not trigger mutation. The child's task was to identify what form the target word should take after the preceding item – mutated or non-mutated. There are 31 Soft Mutation triggers alone (2 Nasal Mutation triggers and 11 Aspirate Mutation triggers – see Lewis, 1993). Therefore, in order to keep the test short, bearing in mind that this was one task out of a large battery of tasks given to the participants, it was important to include the most common mutation triggers that feature in the language children use and hear in Key Stage 2, some of which are referenced within the LNF. For example, within the oracy and writing sections of the framework children are expected to mutate correctly after prepositions and pronouns by the end of Key Stage 2 (Welsh Government, 2013a) and

which would fit in neatly with the *Y Goriad* story so as not to appear too much as an aside. For that reason, the following Soft Mutation triggers were included:

- after the prepositions (*am, ar, at, dan, dros, drwy, i, o, wrth, gan, heb, hyd, tan*)

- after the numeral *dau* 'two'

- after personal possessive pronoun *ei* 'his/its'

- in adjectives and nouns (not verbs) after *yn* 'is'

- after the word '*na* ('yna'/there - from the form 'bod'/to be).

- feminine nouns after the definite article *y/yr/'r* 'the'

Although the task focused on Soft Mutation, the words from which children were asked to select an answer included a version of the word mutated in its nasal and aspirate form, which introduced another dimension to the task. This involved children not only identifying whether a mutation was required but also an ability to identify *which* mutation was required.

Task 1 Example 1

Task 1 Example 2

In order to minimise discomfort for those children who were poor readers, the test included an audio element to enable those children who were not confident readers to listen to the text by clicking on the sentence/word accordingly. This ensured that they were able to access the test and still be able to use their knowledge of mutation to complete the task and were not necessarily hindered by their reading ability. The audio information was recorded with the researcher and a male colleague's voice using support from the Bangor University AV department. This task was administered primarily on a two-to-one basis; however, some were completed on a one-to-one basis depending on need and number of children (i.e. any odd numbers in groups would result in a participant completing the task on a one-to-one basis). When administering on a two-to-one basis children were seated at a table opposite each other (or further part on larger tables in larger classrooms) with the volume on the laptop at a level which would allow them to hear the audio without causing disruption to the other participant in the room. Although some children initially used the sound buttons to hear the voices this appeared to only be for a short time out of curiosity and they then continued without it.

Participants scored one point for each word they selected correctly and the total raw score was calculated. Although children were not given an example to practice this task the researcher supervised them individually for the first sentence to demonstrate how to complete the sentence.

Task 2 – Multiple choice non-words cloze task

Task 2 was another gap fill exercise. However, on this occasion, participants had to try to apply their knowledge of mutation to the non-words in the text. The sentence included a gap, next to which would be the non-word in its original form. Participants would then have to consider whether the word in its original form would need to be mutated in some way or not at all depending on the preceding trigger (See Example 3 and 4 below). The triggers were chosen as triggers for Soft and Nasal Mutation that participants should be familiar with, although actual familiarity with these triggers was not tested prior to selection due to the volume of other tasks which were already being administered as part of the study. Also, general mutation ability was already being assessed as part of the tasks within 'y Goriad'. Due to the more complex nature of this task it was administered on a one-to-one basis with participants. Participants were shown a laminate sheet with three example sentences, they were then given instructions in Welsh on how to complete the task. Once they understood what they were required to do they were then given the opportunity to carry on with the task on the computer.

There were 10 sentences containing non-words, with each non-word having a mutatable onset. Each word was no more than three syllables in length. Each of the option words contained a choice of a non-mutated, correctly mutated and incorrectly mutated word. The non-words included particular language specific characteristics (or phonotactics) found in Welsh which children would have been familiar with such as the double-letter sound *ch* and the *y* sound variability such as in the word *ynys* /ənɨs/ 'island'. A score of 1 was recorded for a correct response and a 0 for an incorrect response. A total raw score was recorded.

Task 2 Example 3

Task 2 Example 4

Task 3a Grammaticality judgement task; Task 3b: Sentence correction Task

Task 3 contained 40 sentences, the distribution of which is listed below. 30 contained a mutation error and 10 did not. The four sets differed in the following way:

Set 1 contained 9 sentences with an unnecessary mutation (ungrammatical)

Set 2 contained 12 sentences without a mutation in an obligatory context for mutation (ungrammatical)

Set 3 contained 10 sentences with correct mutations (grammatical)

Set 4 contained 9 nonsense mutations i.e. sound alterations - 'mutations' - that would never occur in Welsh (ungrammatical).

Each set contained at least 9 sentences that included words beginning with each of the sounds/letter that undergo Soft Mutation in Welsh. There were 27 examples where there was a trigger for Soft Mutation, 2 for Nasal and 2 for Aspirate Mutation. All sentences contained real words, and all target words (errors) were real words.

In Task 3a children were asked to read the sentence that appeared on the screen and to click on the tick or cross underneath to indicate whether they believed the sentence to be correct or incorrect. If they believed it was incorrect they were then prompted by Captain Caradog to click on the word that they believed sounded incorrect (Task 3b). Only one mistake was present in the incorrect sentences. The mistakes within the sentences included using mutations when they were not required (see Example 5), using the incorrect type of mutation (see Example 6) and absurd/silly mutations (see Example 7). Children completed this task on a one-to-one or two-to-one basis depending on need and number within the group.

Task 3 Example 5
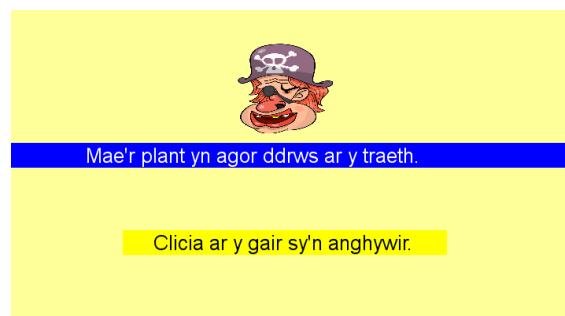
(Task 3a)   (Task 3b)

In the above example the word *drws* 'door' has been softly mutated to *ddrws* in a non-mutatable context. Participants would need to click on the cross (if they identified it as an erroneous sentence) and then the image next to it

would appear on the screen where the participant would be asked to click on the word they believed was incorrect – *clicia ar y gair sy'n anghywir* – 'Click on the word that is incorrect'.

Task 3 Example 6
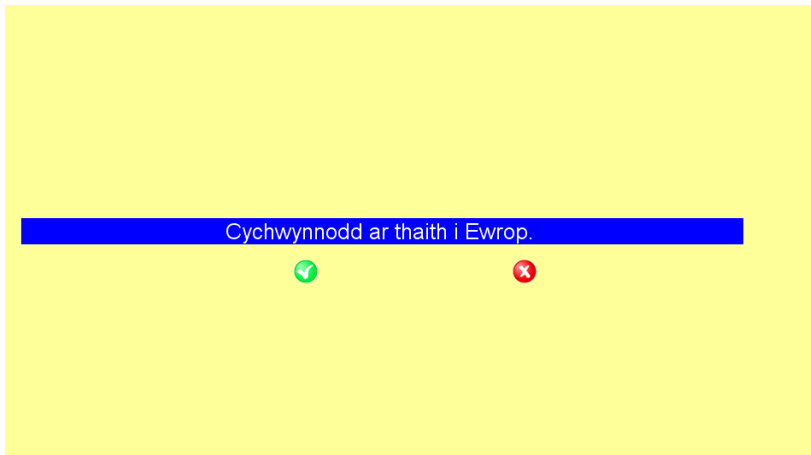
'*Taith*' – 'journey', has undergone an inappropriate Aspirate Mutation into 'thaith' in the above example. The target form is the Soft Mutation form - *daith*.

Task 3 Example 7

The word *befrith* is a made-up word and is nonsensical in this sentence. The real target word here is *llefrith* 'milk', which has undergone mutation to *befrith* (a non-existent mutation of 'll').

In both parts of the task children scored 1 for a correct answer and 0 for an incorrect answer. A total raw score was calculated and recorded separately for each part.

Task 4a: Sentence repetition task; Task 4b: Sentence correction Task

Task 4 was an oral task administered in two parts. In Task 4a, children were instructed to click on a treasure chest from a selection of chests, listen to the sentence attached to the chest, and asked to repeat the whole sentence aloud into a microphone (see Example 8 for illustration). There were 18 sentences in total within the task, 9 sentences were correct and contained a Soft Mutation, 9 were incorrect and contained words that had undergone Soft Mutation in a non-mutatable context. Sentences were simple and ranged from two to six words to avoid an effect of memory on their ability to repeat the sentence (e.g. see Marinis & Armon-Lotum, 2015 regarding length of sentences). The researcher took responsibility for ensuring that children's voices were recorded using the microphone icon on the screen and were allowed to listen to the sentence only once (Marinis & Armon-Lotum, 2015). Children scored 1 if the sentence was repeated exactly as they heard it and 0 if there were any errors such as saying an incorrect word or automatically mutating a word which had been incorrectly mutated within the sentence. A total raw score was recorded.

Task 4b asked the children to listen to the sentences again, one at a time, and then asked to state whether they thought the sentence sounded correct or incorrect. The researcher recorded their response on a record sheet (see appendix 8). The child was then asked to suggest which word they believed to be incorrect in the sentence and state what they believed the correct version of the word should be. Responses were recorded by the researcher. Participants were not permitted to read the sentences; the task was to assess their ability to hear the mutation in the sentence. On a very small number of occasions (approximately less than 5) children misinterpreted the question 'do you think the sentence is correct or incorrect?' (wyt ti'n meddwl bod y

frawddeg yn gywir neu anghywir?) as 'do you think the sentence is true or not?' This is an issue which did not arise in the Pilot study and so was not anticipated. If it was suspected that their interpretation of the question was incorrect, the question was repeated or the researcher explained further what the question was asking. Task 4 was administered on a one-to-one basis and a score of 1 was recorded for a correct response and a 0 for an incorrect response. A total raw score was calculated and recorded.

Task 4 Example 8
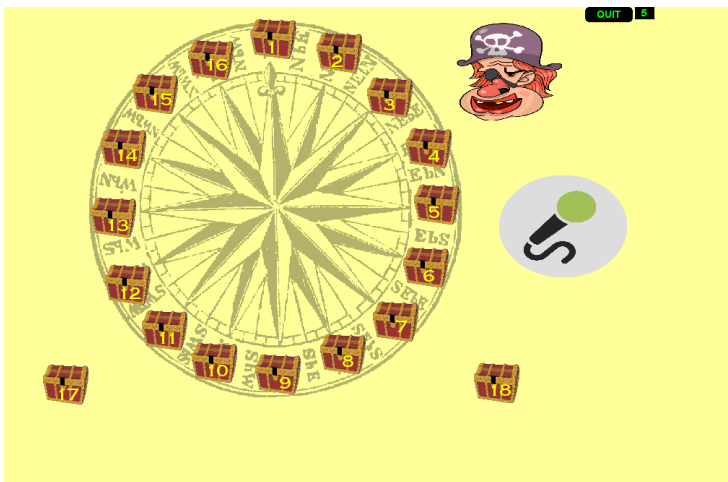
Examples of the types of sentences children were asked to repeat included *Roedd y llong ar dân* (*the boat was on fire*), **Mae'r eliffant yn chwythu ei trwyn* ('The elephant is blowing his nose' – incorrect mutation*)* and *Enillodd y falwoden y ras* ('The snail won the race'*).*

Results

IBM SPSS Statistics 25 was used to run an initial correlation analysis of the whole sample, exploring the scores on the reading tests used within the study to discover whether they showed a strong relationship with each of the variables. Further correlation analyses were then run to explore the relationships between all background measures and children's performance on the mutations tasks within the sample as a whole. Given that not all

variables within the study were normally distributed, a non-parametric Spearman's rho (two-tailed) analysis was carried out.

Table 9: Summary of whole sample correlation analysis according to reading scores and background measures.

Variable	AWRTW[11] (n64)		AWRTE[12] (n64)		NRT Welsh[13] (n 42)		NRT English[14] (n 37)	
	r	p	r	p	r	p	r	p
AWRTW Std					.824	.000	.669	.000
AWRTE Std					.695	.000	.834	.000
Age	-.112	.377	.174	.158	.020	.901	.060	.723
IQ	.271	.031	.402	.001	.303	.051	.450	.005
Letter Recognition English	-.048	.708	-.190	.129	.012	.939	-.146	.389
Letter Recognition Welsh	-.289	.023	-.346	.005	-.278	.079	-.368	.025
Memory for digits	.238	.060	.468	.000	.207	.195	.390	.017
Elision English	.416	.001	.507	.000	.238	.133	.467	.004
Blending Words English	.192	.134	.178	.156	.379	.016	.479	.003
Phoneme Isolation English	.365	.004	.371	.003	.439	.005	.614	.000
Initial Sound Welsh	.223	.080	.100	.424	.141	.380	.148	.380
Final Sound Welsh	.313	.012	.456	.000	.319	.042	.335	.043

[11] All Wales Reading Test Welsh
[12] All Wales Reading Test English
[13] National Reading Test Welsh
[14] National Reading Test English

Phoneme Blending Welsh	.262	.038	.098	.433	.222	.162	.161	.342
Phoneme Deletion Initial W	.210	.098	.357	.003	.113	.481	.110	.518
Phoneme Deletion Final Welsh	.441	.000	.468	.000	.427	.005	.359	.029

The first finding of note is the strong positive correlation between the All Wales Reading Tests and the NRT (National Reading Tests) for both Welsh and English. A significant strong positive correlation was found between the AWRTW test and the NRT Welsh (NRTW) ($r=.824$, $p=<.001$) and between the AWRTW Test and NRT English (NRTE) ($r=.669$, $p=<.001$). A significant strong positive correlation was also found between the AWRTE test and the NRTE ($r=.834$, $p=<.001$) and also between the AWRTE test and the NRTW ($r=.695$, $p=<.001$). This demonstrates that children's scores on these tests are very closely linked and suggest that both tasks, despite tapping into different literacy abilities, provide similar indicators of children's reading ability in both languages. Any differences in correlations with various tasks are therefore unlikely to be due to a problem/issue with what these tasks measure.

Background measures and reading scores

Age, Letter recognition English (LRE) and the Welsh Initial sound task did not correlate with reading scores. These variables therefore were not explored further in relation to children's reading scores within this study.

IQ

Analysis revealed a significant correlation between children's IQ and their performance on all reading tests across the whole sample. A significant weak positive correlation was found between IQ and the AWRTW test ($r=.271$,

p=.031) and a stronger correlation between IQ and the AWRTE test (r=.402, p=.001). There was also a significant moderate positive correlation between IQ and the NRTW (r=.303, p=.051) and between IQ and the NRTE test (r=.450, p=.005).These findings indicate that as children's IQ increased so too did their reading test scores. This would suggest that children with higher IQ perform better on reading tests than children with a lower IQ.

Letter recognition (LRW) Welsh

A significant correlation was revealed between all reading scores except the NRTW and children's ability to recognise letter sounds in Welsh. A significant weak negative correlation was found between LRW and the AWRTW (r= -.289, p=.023) and moderate negative correlation found between LRW and AWRTE tests (r=-.346, p=.005). A significant moderate negative correlation was also found between LRW and NRTE (r= -.368, p=.025). This would suggest that higher ability readers make fewer errors when identifying letter sounds in Welsh than lower ability readers.

Memory for Digits

A significant moderate positive correlation emerged between children's scores on the memory for digits task and AWRTE test (r=.468, p=<.001) and the NRTE (r=.390, p=.017). This implies that higher ability English readers perform better on this task than lower ability readers. Interestingly, no significant correlations emerged between children's scores on the Memory for Digits task and their Welsh reading scores. Ellis and Hooper (2001) found that digit length and time to read digits is longer in Welsh than in English (see Chapter 2), which could explain this finding.

Phonological Awareness

In order to gain a general measure of the validity of the ELDEL a correlation analysis was run on the whole sample between the tasks on the CTOPP-2 and tasks on the ELDEL. The findings are presented in the table below.

Table 10: Correlations output between scores on the CTOPP2 and ELDEL measures.

	Elision		Blending words English		Phoneme isolation English	
	r	p	r	p	r	p
Initial sound Welsh	.142	.258	.069	.583	.223	.076
Final sound Welsh	.382	.002	.089	.480	.053	.678
Phoneme Blending Welsh	.263	.034	.167	.184	.284	.023
Phoneme deletion initial Welsh	.440	.000	.088	.488	.216	.086
Phoneme deletion final Welsh	.293	.018	.260	.036	.197	.118

A positive significant moderate correlation was found between the Elision task and the final sound Welsh (r= .382, p= .002), Phoneme blending Welsh (although weak) (r= .263, p= .034), Phoneme deletion initial Welsh (r= .440, p= .000) and Phoneme deletion final Welsh (r=.293, p= .018) (also relatively weak). Only one significant correlation was found between the Blending words task on the CTOPP-2 and the ELDEL task namely Phoneme deletion final Welsh (r=.260, p=.036), however this was relatively weak. No significant correlations were found between the blending words English and Phoneme blending Welsh task which brings into question the validity of the ELDEL in relation to this task, given that they are fundamentally measuring the same thing. The CTOPP-2 Phoneme isolation task only significantly correlated with the Phoneme blending task on the ELDEL (r=.284, p= .023), however this was also weak. This could be due to the nature of the tasks varying slightly, which

could also account for other correlations not emerging. When considering any findings within this study in relation to children's Welsh phonological awareness, any conclusions drawn must be taken with caution due to the ELDEL not being a standardised measure and not correlating significantly with relevant items of the CTOPP-2.

Phonological awareness and reading scores

English

Elision

Significant correlations emerged between children's performance on the Elision task and their reading scores in both English and Welsh (except NRTW). A significant moderate positive correlation was found between Elision and the AWRTW reading test ($r=.416$, $p=.001$) and between Elision and the AWRTE reading test ($r=.507$, $p=<.001$). A significant moderate positive correlation was also found between Elision and the NRTE ($r=.467$, $p=.004$). This suggests that higher ability readers perform better on this task than lower ability readers.

Blending Words

Correlation analysis revealed a significant moderate positive correlation between children's scores on the English blending words task and their NRT reading scores in both English ($r=.479$, $p=.003$) and Welsh ($r=.379$, $p=.016$). This would suggest that higher ability readers perform better on this task than lower ability readers.

Phoneme isolation

A significant correlation was revealed between children's reading scores on all reading scores. A significant moderate positive correlation was found between phoneme isolation and the AWRTW reading test ($r=.365$, $p=.004$), the AWRTE reading test ($r=.371$, $p=.003$), the NRTW ($r=.439$, $p=.005$) and the NRTE

(r=.614, p=<.001), the latter being the strongest. This suggests that higher ability readers perform better on this task than lower ability readers.

Phonological processing Welsh Final sound

Significant correlations were revealed between all reading tests and children's performance on this task. A significant moderate positive correlation was found between final sound and the AWRTW test (r=.313, p=.012), the AWRTE test (r=.456, p=<.001) the NRTW (r=.319, p=.042) and the NRTE (r=.335, p=.043). Although the NRT scores revealed a weak significant effect, this difference could be linked to the lower sample size within the NRT. This would suggest that lower ability readers have more difficulty completing this task than higher ability readers.

Phoneme blending Welsh

A significant positive correlation was found between children's AWRTW reading scores and their scores on the Welsh phoneme blending words task (r=.262, p=.038), suggesting that higher ability readers perform better on this task than lower ability readers. However the strength of this correlation was weak. No significant correlations were found between phoneme blending Welsh and any other reading test scores. Although the reason for this is unclear, it could again be explained by the smaller sample within the NRT and the fact that the Welsh phonological awareness tests are not standardised measures.

Phoneme deletion initial

Analysis revealed a positive moderate significant correlation between phoneme deletion initial and the AWRTE reading test (r=.357, p=.003). No other significant correlations were found between children's reading scores and their performance on this task.

Phoneme deletion final

A significant correlation was revealed between children's reading scores across all reading tests and their performance on this task. A significant moderate positive correlation was found between phoneme deletion final and the AWRTW test ($r=.441$, $p=<.001$), the AWRTE test ($r=.468$, $p=<.001$), the NRTW ($r=.427$, $p=.005$) and the NRTE ($r=.359$, $p=.029$). This suggests that lower ability readers would not perform as well as higher ability readers on this task.

The tables which follow present the correlation analysis output of the whole sample between background measures and the mutation tasks. Significant relationships are highlighted in yellow and results are then discussed.

Mutation tasks results

Table 11: Results of correlation analysis between mutation tasks and scores on the AWRT and NRT.

Variable	AWRTW (n64)		AWRTE (n64)		NRT Welsh (n 42)		NRT English (n 37)	
	r	p	r	p	r	p	r	p
Task 1	.647	.000	.296	.016	.516	.001	.399	.013
	(n63)		(n66)		(n41)		(n38)	
Task 2	.351	.005	.297	.018	.323	.045	.218	.202
	(n61)		(n63)		(n39)		(n36)	
Task 3a	.533	.000	.343	.007	.440	.005	.302	.074
	(n57)		(n60)		(n40)		(n36)	
Task 3b	.527	.000	.374	.003	.485	.002	.360	.031
	(n57)		(n60)		(n40)		(n36)	
Task 4a	.336	.009	.288	.023	.414	.010	.457	.006

	(n60)	(n62)	(n38)	(n35)
Task 4b	.423 ==.001== .221 .085	.351 ==.029== .313 .063		
	(n61)	(n62)	(n39)	(n36)

Table 12: Results of correlation analysis between mutation tasks and IQ, Age, Letter recognition (English/Welsh) and Memory for digits.

	IQ		Age		LRE		LRW		Memory for digits	
	r	p	r	p	r	p	r	p	r	p
Task 1	.225	.069	.154	.217	-.182	.150	-.136	.284	.087	.489
Task 2	-.022	.862	-.001	.994	-.061	.640	-.241	.059	.093	.466
Task 3a	.282	==.029==	.298	==.021==	-.084	.530	-.115	.389	.204	.121
Task 3b	.348	==.007==	.360	==.005==	-.186	.162	-.122	.362	.221	.093
Task 4a	.052	.689	.055	.669	-.092	.480	-.058	.658	.237	.063
Task 4b	.079	.543	.221	.084	-.075	.564	-.125	.337	-.047	.718

Table 13: Results of correlation analysis between mutations tasks and tasks measuring phonological awareness in English and Welsh.

	Elision		Blending Words		Phoneme isolation		Initial Sound W		Final Sound W		Phoneme blending W		Phoneme deletion initial W		Phoneme deletion final W	
	r	p	r	p	r	p	r	p	r	p	r	p	r	p	r	p
Task 1	.392	.001	.359	.004	.290	.021	.285	.021	.171	.173	.337	.006	.163	.196	.334	.006
Task 2	.21	.095	.322	.011	.084	.519	.103	.421	.047	.715	-.123	.338	.300	.017	.214	.092
Task 3a	.297	.024	.275	.036	.469	.000	.140	.292	.051	.702	.277	.034	.131	.322	.288	.027
Task 3b	.363	.005	.322	.014	.553	.000	.163	.216	.068	.610	.281	.031	.156	.239	.312	.016
Task 4a	.202	.118	.317	.013	.195	.135	.240	.060	.147	.255	.343	.006	.096	.459	.439	.000
Task 4b	.223	.084	.374	.003	.198	.130	-.026	.844	-.018	.891	.267	.036	.045	.731	.293	.021

Task 1: Multiple choice real word cloze task

Analysis revealed a significant positive correlation between children's performance on Task 1 and their reading scores on both the AWRTW reading test (r=.647, p=<.001) and the AWRTE test (r=.296, p=.016). However this was strongest with the AWRTW. A significant positive correlation was also found between Task 1 and the NRTW (r=.516, p=.001) and the NRTE (r=.399, p=.013). This suggests that as children's scores in reading increased so too did their scores on Task 1, which implies that higher ability readers have a better awareness of mutatable contexts (within the parameters of Task 1) than lower ability readers.

A significant correlation emerged between children's scores on Task 1 and their scores on all phonological awareness tasks with the exception of Welsh final sound and phoneme deletion initial in Welsh (see Table 13). This implies that as children's scores on Task 1 increase so too do their scores on these phonological awareness tasks, suggesting that Task 1 may be an indicator of children's phonological awareness abilities (in relation to the areas of phonological awareness which were examined).

No significant correlations were revealed between children's scores on Task 1 and their IQ, letter recognition (Welsh/English) or memory for digits scores. Furthermore there was no significant relationship between children's age and performance on Task 1 implying that performance on this task is not affected by the child's age, IQ or short term memory

Task 2 – Multiple Choice non-words cloze task

A significant correlation was revealed between children's reading scores and their scores on Task 2 across all reading tests except the NRTE. A significant positive correlation was revealed between Task 2 and the AWRTW test (r=.351, p=.005) the AWRTE test (r=.297, p=.018), and with the NRTW (r=.323, p=.045) (although the weakest correlation was with the AWRTE). This would suggest that as children's scores in their Welsh reading tests increased so too did their scores on Task 2. This was also the case with regards

to the AWRTE which also revealed a significant positive correlation with scores on Task 2. However, English reading skills, as measured by the NRT did not reveal the same tendency, despite there being a strong correlation between the NRT and the AWRT.

In contrast with Task 1, very few significant correlations were revealed between children's scores on Task 2 and their scores on the phonological awareness tests. A positive correlation emerged between children's performance on Task 2 and their scores on the English blending words task ($r=.322$, $p=.011$) and the Welsh phoneme deletion initial task ($r=.300$, $p=.017$). Given no other significant correlations were found, a non-word cloze mutation task may not yield results that relate strongly with children's phonological awareness skills in Welsh or in English.

No significant correlations were found between children's performance on Task 2 and their age and IQ, letter recognition (Welsh/English) or memory for digits scores (although a weak significant negative correlation was present with letter recognition in Welsh).

Task 3a – Grammaticality judgement task

Results revealed a significant correlation between children's scores on this task and their reading scores in both the All Wales Reading Test (English and Welsh) and their scores on the Welsh NRT reading scores only. A significant positive correlation was found between Task 3a and the AWRTW test ($r=.533$, $p=<.001$) the AWRTE ($r=.343$, $p=.007$) and the NRTW ($r=.440$, $p=.005$). No significant correlation emerged with scores in the NRTE. This would suggest that higher ability Welsh readers perform better on this task than lower ability readers, but their performance in English may not yield the same type of relationship.

A significant positive correlation was found between children's scores on tasks 3a and Elision ($r=.297$, $p=.024$), Blending words ($r=.275$, $p=.036$), Phoneme isolation ($r=.469$, $p=<.001$), Phoneme blending Welsh ($r=.277$, $p=.034$) and Phoneme deletion final Welsh ($r=.288$, $p=.027$). This suggests that as

children's scores increase on this task so too do their scores on these specific phonological awareness tasks, although this was strongest with phoneme isolation.

A significant positive correlation was revealed between children's performance on Task 3a and IQ scores (r=.282, p=.029) and Age (r=.298, p=.021). This suggests that children who have higher IQ scores perform better on these tasks than those with lower IQ scores. Also, older children perform better than younger children on these tasks, which could relate to the grammatical nature of the tasks as older children will have been exposed to more grammatical structures which would enable them to identify inaccuracies within a sentence. However this correlation was weak and so findings should be interpreted with caution.

Task 3b – Sentence correction task

A significant correlation was present between all reading test scores and children's scores on this task. A significant positive correlation was found between Task 3b and the AWRTW (r=.527, p=<.001), the AWRTE (r=.374, p=.003) the NRTW (r=.485, p=.002) and the NRTE (r=.360, p=.031) suggesting that higher ability English and Welsh readers perform better on this task than lower ability readers.

A significant positive correlation was found between Task 3b and Elision (r=.363, p=.005), Blending words (r=.322, p=.014), Phoneme isolation (r=.553, p=<.001), Phoneme blending Welsh (r=.281, p=.031) and Phoneme deletion final (r=.312, p=.016). This suggests that as children's scores increase on this task so too do their scores on these specific phonological awareness tasks which could suggest that this task may be a good proxy test for phonological ability

A significant positive correlation was found between Task 3b and IQ scores (r=.348, p=.007) and age (r=.360, p=.005). This suggests that children who have higher IQ scores perform better on these tasks than those with lower IQ scores. Also, older children perform better than younger children on these

tasks. This task therefore appears to differentiate between groups which could be useful as an item within diagnostic testing.

Task 4a – Sentence repetition task

A significant positive correlation was present between Task 4a and the AWRTW (r=.336, p=.009), AWRTE (r=.288, p=.023), NRTW (r=.414, p=.010) and NRTE (r=.457, p=.006), suggesting that higher ability English and Welsh readers perform better on this task than lower ability readers, although the strength of this correlation was lowest with the AWRTE.

A significant correlation was revealed between children's scores on Task 4a and Blending words (r=.317, p=.013), phoneme blending Welsh (r=.343, p=.006) and phoneme deletion final Welsh (r=.439, p=<.001). No other significant correlations were found between the phonological awareness tasks and Task 4a or between Task 4a and IQ, age, letter recognition (Welsh/English) or memory for digits

Task 4b – Sentence correction Task B

Correlation analysis revealed a significant relationships between children's Welsh reading scores (All Wales and NRT) and their performance on this task. A significant positive correlation was found between Task 4b and the AWRTW (r=.423, p=.001) and the NRTW (r=.351, p=.029). This suggests that higher ability Welsh readers perform better on this task than lower ability Welsh readers. No significant correlations emerged between English reading scores and children's performance on this task. A significant correlation was revealed between children's scores on Task 4b and Blending words (r=.374, p=.003), phoneme blending welsh (r=.267, p=.036) and phoneme deletion final Welsh (r=.293, p=.021) (although this was weakest). No other significant correlations were found between the phonological awareness tasks and Task 4b.

151

No significant correlations were found between children's scores on task 4a and 4b and their IQ, age, letter recognition (Welsh/English) or memory for digits scores. Research question 1 asked 'Is there a relationship between children's performance on morpho-phonological tasks and their reading abilities? If so, what is the nature of this relationship and does it vary between language groups?

In relation to part 1 of the question it is clear that there is a significant positive correlation between children's reading scores and their performance on the mutation tasks. In particular, Task 1 correlated with all reading measures as did Task 3b and 4a. Interestingly, each of these particular tasks assess different skills in relation to mutation: Task 1 is a multiple choice cloze task, which relies on children's abilities to correctly identify the word (either in its mutated form or not) which would complete the sentence; Task 3b is a sentence correction task, which assesses children's ability to identify the error in a sentence and correct that error; and 4a is a sentence repetition task, which assesses children's ability to repeat a sentence they have heard exactly as they have heard it (although the sentence may be incorrect in terms of the mutation used or not used within the sentence).

The remainder of the mutation tasks - Tasks 2, Task 3a and Task 4b - correlated significantly with AWRT in Welsh and in English and NRT in Welsh but not with the NRT in English, despite both the NRT and AWRT tests being strongly correlated. It is not clear why no correlation emerged between these tasks and the NRT English, as all reading tests within the study correlated with each other, suggesting that there were no significant differences with what the tests measured. However, the sample size for this particular reading test was smaller when compared with the AWRT. The mutation tasks also tap into very specific Welsh grammar items which would bear no relationship to the items within the English NRT. What is clear from the above analysis is that current standardised measures of Welsh reading abilities used in schools in Wales correlate significantly with children's performance on a variety of mutation tasks, and so too do measures of English reading comprehension as measured by the AWRT.

Another important finding is the significant correlation between children's performance on the mutations tasks and their performance on the CTOPP-2, a standardised measure of phonological awareness. Task 1, Task 3a and Task 3b all revealed a significant positive correlation with the three phonological awareness tasks of the CTOPP-2 (Elision, Blending words and Phoneme isolation), supporting the notion that these three mutations tasks, which measure a morpho-phonological process, could be valid measures.

In relation to part 2 of research question 1 I wanted to explore the role of language and whether L1 Welsh speakers performed better than 2L1 English/Welsh speakers. The sample was therefore split into 2 groups: Mostly Welsh/Group 1 and Mostly English/Group 2 (see Chapter 4 for full description of each group). The raw scores for each of the mutation tasks can be seen in Table 14 below followed by the results of the correlation analysis output according to language group.

Table 14: Mean raw scores for each task according to language group.

	Group 1 (L1 Mostly Welsh)	Group 2 (Mostly English)
Task 1 average score (24)*	18.10 (n51) S.D: 4.011	16 (n15) S.D: 3.071
Task 2 average score (14)*	8.88 (n49) S.D: 3.327	9.79 (n14) S.D: 2.486
Task 3a average score (40)*	25.27 (n47) S.D: 7.453	23.62 (n13) S.D: 4.770
Task 3b average score (30)*	14.53 (n47) S.D: 8.270	14.08 (n13) S.D: 5.678
Task 4a average score (18)*	14.35 (n49) S.D: 2.411	13.31 (n13) S.D: 1.974
Task 4b average score (18)*	10.98 (n48) S.D: 2.756	10.21 (n14) S.D: 1.968

*denotes highest possible score

An independent samples t-test was run to determine whether any differences between the mean scores of the groups carried any statistical significance. Findings revealed no statistically significant differences between mean group scores and performance on the mutations tasks.

As might be predicted, results show that the average raw score across each of the tasks was higher in the Mostly Welsh group than the Mostly English group with the exception of Task 2 where the Mostly English group performed better. There were limitations to Task 2 which could explain this difference, these are discussed later in the chapter. The mean scores appear to show very little variety across groups which could be explained by the difference in group size. Although on the surface the mean scores appear to be very similar, the standard deviations (SD) vary considerably between groups in Task 3a and 3b where the SD in the Mostly Welsh group is almost three points higher than in the Mostly English group, suggesting a high level of variability between group means.

Correlation results per language group

Table 15: Group 1 (Mostly Welsh) Significant Correlations between reading test scores and mutations tasks.

Variable	AWRTW (n50)		AWRTE(n50)		NRT Welsh (n 32)		NRT English (n 28)	
	r	p	r	p	r	p	r	p
AWRTW Std	-	-	.699	.000	.848	.000	.687	.000
AWRTE Std	.699	.000	-	-	.685	.000	.801	.000
Task 1	.612	.000	.401	.004	.523	.003	.415	.028
Task 2	.343	.018	.378	.007	.238	.206	.167	.405
Task 3a	.558	.000	.443	.002	.410	.022	.226	.256
Task 3b	.549	.000	.431	.002	.427	.017	.241	.227

Task 4a	.303	.038	.320	.025	.432	.017	.435	.023
Task 4b	.409	.004	.226	.122	.267	.155	.260	.190

Table 16: Group 2 (Mostly English) Correlations results.

Variable	AWRTW (n14)		AWRTE (n14)		NRT Welsh (n 9)		NRT English (n 9)	
	r	p	r	p	r	p	r	p
AWRTW Std	-	-			.885	.002	.895	.001
AWRTE Std	.670	.009			.726	.017	.884	.001
Task 1	.628	.016	.578	.024	.710	.021	.691	.027
Task 2	.496	.072	-.065	.824	.293	.444	.281	.464
Task 3a	.416	.178	.256	.399	.407	.277	.617	.077
Task 3b	.525	.080	.377	.204	.613	.079	.711	.032
Task 4a	.335	.263	.437	.135	.568	.142	.812	.014
Task 4b	.479	.083	.394	.164	.735	.024	.612	.080

At first glance it is clear to see that the majority of significant correlations exist within the Welsh only group, with very little emerging within the Mostly English group. Nevertheless, Task 1 appears to be significantly correlated with both language groups which may suggest that children's language background may not be playing a significant role in relation to their abilities to complete this task. The strength of this correlation is also moderate to high across all reading tasks and Task 1, with the highest being with AWRTW in the Mostly Welsh group and NRTW in the Mostly English group. Task 4b also has a strong correlation with NRTW in the Mostly English group and Tasks 3b and 4a have

a strong correlation with NRTE, all of which are much stronger correlations than when the sample was analysed as a whole.

In relation to part 2 of research question 1 it is apparent that there are considerable differences between the groups in terms of the significant relationship between the mutation tasks and reading scores. Although the exposure of each group to Welsh may be very different, it appears their ability to apply mutation rules within written texts is very similar (within the parameters of Task 1). This may support the notion that Task 1 may be a significant marker of literacy ability as it does not appear to be affected by level of Welsh language exposure.

Correlation analysis has shown there is a strong relationship between mutation abilities and reading skills among L1 Welsh speakers, and some indication of a relationship - but not across-the board - for children for whom Welsh may not be an L1. This provides further evidence to support the need to identify relevant linguistic profiles for different types of bilinguals and to ensure appropriate norms. In order to identify which of the task variables serve as potential predictors of reading abilities, a series of regression analyses were performed based on significant correlations.

Regression Analysis

Results of simple regression analysis

A simple linear regression analysis was run to explore whether any of the variables that had a significant correlation with reading scores within Study 2 could be considered a predictor of children's reading scores. The results can be seen below in tables 17 to 20.

Table 17: Simple regression Analysis results according to group. Dependent variable: AWRTW.

	Group 1					Group 2				
Task	R^2	Sig	SE	Unstandardised B	Standardised Beta	R^2	Sig	SE	Unstandardised B	Standardised Beta
Task 1	.509	.000	.402	2.803	.713	.303	.042	.948	2.163	.550
Task 2	.088	.043	.662	1.377	.296	.130	.205	1.351	1.810	.361
Task 3a	.282	.000	.276	1.135	.531	.183	.165	.846	1.266	.428
Task 3b	.305	.000	.246	1.069	.552	.213	.131	.673	1.106	.461
Task 4a	.113	.021	1.038	2.488	.336	.215	.110	1.757	3.051	.464
Task 4b	.138	.010	.780	2.094	.372	.220	.091	1.617	2.970	.469

AWRTW Reading Test

As revealed in the table above, Task 1 accounts for the highest level of variance within Group 1 when compared with the other regression models within this study (r^2=.509), which indicates that Task 1 accounts for 50.9% of the variance in children's scores on the AWRTW Reading test. Task 1 also accounted for 30.3% of the variance in AWRTW reading scores in Group 2 along with Task 3a accounting for 28.2% of variance and Task 3b accounted for 30.5% of variance in Group 1. A multiple regression analysis will explore this further later in this chapter.

Table 18: Simple regression Analysis results according to group. Dependent variable: AWRTE.

Task	Group 1					Group 2				
	R^2	Sig	SE	Unstandardised B	Standardised Beta	R^2	Sig	SE	Unstandardised B	Standardised Beta
Task 1	.176	.002	.428	1.385	.420	.283	.041	1.316	2.977	.532
Task 2	.097	.030	.538	1.206	.311	.005	.817	2.040	.484	.068
Task 3a	.165	.005	.249	.743	.406	.103	.284	1.076	1.211	.321
Task 3b	.165	.005	.224	.670	.407	.145	.199	.882	1.205	.381
Task 4a	.069	.069	.753	1.404	.262	.267	.070	2.305	4.618	.517
Task 4b	.013	.439	.693	.541	.114	.213	.096	2.291	4.133	.462

AWRTE Reading Test

Although all tasks except Task 4b run within the simple regression models in the table above show a significant relationship with the scores on the AWRTE Reading test within Group 1, and Task 1 has a significant relationship with the AWRTE in Group 2, all tasks fall below 30% of variance (known as a low/weak effect) and are therefore disregarded within this analysis as predictors of English AWRT reading test scores given.

Table 19: Simple regression Analysis Results according to group. Dependent Variable: NRTW.

Task	Group 1					Group 2				
	R^2	Sig	SE	Unstandardised B	Standardised Beta	R^2	Sig	SE	Unstandardised B	Standardised Beta
Task 1	.269	.003	.530	1.731	.519	.371	.062	1.343	2.916	.609
Task 3a	.179	.018	.318	.799	.423					
Task 3b	.214	.009	.290	.814	.462					
Task 4a	.184	.018	1.196	3.005	.429					
Task 4b						.372	.081	2.224	4.531	.610

NRTW reading scores

Although a significant relationship was revealed between all tasks included in the simple regression analyses and the NRTW reading scores in Group 1 (as indicated in table 19 above) the strength of the variance was low (less than 30%) and therefore not considered as potential predictors of Welsh NRT scores. No significant relationship was present between the tasks run within a simple regression analysis in Group 2 and the NRTW reading scores.

Table 20: Simple regression analysis results according to group. Dependent variable: NRTE.

| Task | Group 1 | | | | | Group 2 | | | | |
	R^2	Sig	SE	Unstandardised B	Standardised Beta	R^2	Sig	SE	Unstandardised B	Standardised Beta
Task 1	.182	.024	.588	1.413	.426	.376	.060	1.510	3.314	.613
Task 3b						.510	.031	.905	2.443	.714
Task 4a	.245	.009	1.210	3.450	.495	.720	.008	1.860	7.300	.848

NRTE Reading Scores

A low value of r^2 was revealed within Group 1 simple regression analysis (Task 1 18.2% of variance and Task 4a 24.5% of variance) and so these tasks are disregarded as predictors of English reading test scores within the study. However, in Group 2 Task 3b and Task 4a had a significant relationship with the NRTE, accounting for 51% of variance (Task 3b) and 72% of variance (Task 4a) respectively. This was therefore explored further within multiple regression analysis, discussed below. Table 25 provides a summary of the findings of the simple regression analysis of Study 2.

Table 25: Summary of findings from Study 2 simple regression analysis.

Task	Reading Test	Percentage of variance	Group
Task 1: Multiple choice real words cloze task	AWRTW	50.9%	Group 1
		30.3%	Group 2
	AWRTE	17.6%	Group 1
		28.3%	Group 2
	NRTW	26.9%	Group 1 only
	NRTE	18.2%	Group 1 only
Task 2: Multiple choice non-words cloze task	AWRTW	8.8%	Group 1 only
	AWRTE	9.7%	Group 1 only
Task 3a: Grammaticality judgement task	AWRTW	28.2%	Group 1 only
	AWRTE	16.5%	Group 1 only
	NRTW	17.9%	Group 1 only
Task 3b: Sentence correction Task	AWRTW	30.5%	Group 1 only
	AWRTE	16.5%	Group 1 only
	NRTW	21.4%	Group 1 only
	NRTE	51.0%	Group 2 only
Task 4a: Sentence repetition task	AWRTW	11.3%	Group 1 only
	AWRTE	6.9%	Group 1 only
	NRTW	18.4%	Group 1 only
	NRTE	24.5%	Group 1
		72.0%	Group 2
Task 4b: Sentence correction Task	AWRTW	13.8%	Group 1 only

Multiple Regression Analysis

A multiple regression analysis was run in order to assess which of the tasks would be the best predictor of children's reading when compared within a single model. Table 21 below presents the findings. The variables included in each of the models are those which were revealed earlier in the Chapter as having a statistically significant correlation with the Dependent Variable.

Table 21: Multiple regression analysis Group 1: Dependent variable: AWRTW.

Task	R^2	Sig	SE	Unstandardised B	Standardised Beta
	.537				
Task 1		.001	.623	2.146	.549
Task 2		.628	.579	.283	.060
Task 3a		.703	1.087	-.418	-.199
Task 3b		.526	1.025	.657	.346
Task 4a		.370	.864	.785	.121
Task 4b		.987	.829	.014	.002

*ANOVA significant at .000 (Group 1)

AWRTW Reading Test

The table above shows the multiple regression model based on the variables which were revealed as significantly correlated with the AWRTW within Group 1. As Task 1 was the only significantly correlated task within Group 2 no multiple regression analysis was run for Group 2. The model accounts for 53.7% of the variance in scores on the AWRTW and is statistically significant at $p=<.001$. Analysis reveals that Task 1 makes a statistically significant unique contribution to the prediction of scores on the AWRTW, making it a potential marker of reading abilities in Welsh.

Table 22: Multiple regression analysis Group 1: Dependent Variable: AWRTE.

		Group 1			
Task	R^2	Sig	SE	Unstandardised B	Standardised Beta
	.252				
Task 1		.370	.648	.589	.179
Task 2		.239	.601	.719	.182
Task 3a		.972	1.128	.040	.022
Task 3b		.787	1.041	.283	.177
Task 4a		.368	.899	.818	.150

*ANOVA significant at .041 (Group 1)

AWRTE Reading Test

As above within the AWRTW reading test analysis, Group 2 only showed significant correlations between Task 1 and AWRTE reading test scores and so was not further explored within this multiple regression analysis. The model for Group 1 accounted for 25.2% of the variance in AWRTE reading scores. It was concluded therefore that these tasks were not a significant predictor of children's reading scores on the AWRTE.

Table 23: Multiple regression analysis Group 1 and Group 2: Dependent Variable: NRTW.

	Group 1					Group 2				
Task	R^2	Sig	SE	Unstandardised B	Standardised Beta	R^2	Sig	SE	Unstandardised B	Standardised Beta
	.416					.430				
Task 1		.311	.728	.752	.219		.466	2.203	1.712	.350
Task 3a		.164	1.287	-1.845	-.996					
Task 3b		.101	1.187	2.018	1.209					
Task 4a		.041	1.020	2.196	.384					
Task 4b							.461	3.436	2.705	.355

*ANOVA significant at .007 (Group 1) not significant in Group 2

NRTW Reading Scores

The model within Group 1 accounts for 41.6% of variance in the NRTW scores. Task 4a makes the only statistically significant unique contribution within this model. None of the variables within the Group 2 model were revealed as having a statistically significant unique contribution and so these variables were disregarded as predictors of children's reading scores on the NRTW for children within this group.

Table 24: Multiple regression analysis Group 1 and Group 2: Dependent Variable: NRTE.

	Group 1					Group 2				
Task	R^2	Sig	SE	Unstandardised B	Standardised Beta	R^2	Sig	SE	Unstandardised B	Standardised Beta
	.316					.804				
Task 1		.130	.607	.954	.284		.330	2.455	2.719	.493
Task 3b							.624	1.698	-.901	-.302
Task 4a		.040	1.010	2.192	.393		.071	2.967	7.261	.846

*ANOVA Significant at .011 in Group 1, not significant in Group 2

The model above (table 24) accounts for 31.6% of the variance in scores on the NRTE. Task 4a makes the only statistically significant unique contribution within this model and so further exploration of the use of this kind of task as a marker of literacy difficulty could be worthwhile. In Group 2 none of the variables were revealed as making a statistically significant contribution to the model however the model did account for 80.4% of variance in NRTE scores implying this model is a good predictor of children's English reading ability for children who are from a mostly English background.

Discussion

The aim of this study was to answer the following research questions:

1. Is there a relationship between children's performance on morpho-phonological awareness tasks and their reading abilities? If so, what is the nature of this relationship and does it vary between language groups?
2. Can tests of morpho-phonological awareness predict children's performance on reading tests?

In order to answer these questions analysis of the data involved exploring the correlations between children's scores on the novel tasks, which were designed specifically for this study, and children's reading scores on the All Wales reading tests (both English and Welsh) and the National Reading Tests (both English and Welsh). Children were also assessed on their phonological awareness skills, IQ and memory for digits. The sample was first analysed as a whole and then the sample was split into two groups according to the language background - Mostly Welsh (Group 1) and Mostly English (Group 2). The outcomes of the correlation analysis was then used to inform the regression analysis models, which were run in order to answer research question 2.

The initial focus of research question 1 was to explore the relationship between children's ability to complete morpho-phonological awareness tasks and their reading ability levels in both English and Welsh reading tests. The tasks of morpho-phonological awareness assessed children's reading, writing and speaking abilities in relation to applying particular mutation rules in Welsh. The primary focus was Soft Mutation as this is the most common form of mutation used in Welsh. Elements of Nasal and Aspirate mutation were also touched upon to provide a more holisitc view of children's abilities. All mutation tasks within this study were original due to the lack of availability of tests of this nature in the Welsh language, and were designed to be fun and accessible to children from 7-11 years of age.

Phonological awareness, reading scores and morpho-phonological awareness scores.

All morpho-phonological task scores correlated with at least 2 of the phonological awareness tasks, with Task 1 revealing the highest number of significant correlations. Blending words in English was of particular significance as it significantly correlated with all tasks. This suggests that children with higher levels of phonological processing skills, in particular in relation to blending words in English, perform better on tasks of morpho-phonological processing. This would support the notion that Welsh mutation is a higher level phonological processing skill, as those with higher levels of phonological awareness perform better on these tasks than those with lower levels of phonological awareness.

In relation to reading scores, all phonological awareness scores significantly correlated with at least one of the reading test scores suggesting that within the sample, higher ability readers performed significantly better on particular phonological awareness tasks than lower ability readers. Phoneme isolation (English) and phoneme deletion final (Welsh) were of particular significance as they correlated significantly with all reading tests. This would support the notion that higher ability readers have higher levels of phonological awareness than lower ability readers on these particular tasks; however, this relationship was not across-the-board for all phonological awareness tests and for both languages. The language of testing as well as the type of test can affect the outcome of such relationships. For example, not all Welsh phonological awareness tasks correlated significantly with Welsh reading tests and neither did the English phonological awareness tasks with the English reading tests. It is difficult to determine why this was the case from these data alone; however, the Welsh phonological awareness tests were not standardised and the nature of the reading tests may not correspond with the content of the phonological awareness tasks.

Reading Scores and Task 1 Multiple choice real words cloze task

Task 1 was a cloze task which measured children's abilities to identify the correct mutated/non-mutated form of the word to complete the sentence. This task emerged as a task of significance within this study. It was significantly correlated with both the All Wales Reading tests and the NRT in both English and Welsh, this applied also within both language groups. This would suggest that children who perform well on reading tests in both English and Welsh also perform well on Task 1, regardless of whether they were from a Welsh medium home or not. Following this finding a simple regression analysis was run to explore the predictability of Task 1 in relation to each of the reading scores. Task 1 had the highest predictability with the AWRTW test at 50.9% within the Mostly Welsh group (30.3% within the Mostly English group), suggesting that Task 1 accounts for 50.9% of the variance in children's scores on the AWRTW reading test within this group. This a promising result as it suggests there is a clear and significant link between children's reading scores and their performance on this kind of morpho-phonological test, to a degree where Task 1 could predict more than 50% of the variance in the AWRTW reading test. The fact that this task has emerged as significant regardless of children's language background is important, as children attending schools in Wales come from homes with varying degrees of exposure to Welsh and so it is essential that any screening test which is developed can be used with all children within the classroom. Further research in this area is needed to determine what elements could be accounting for the remaining 49.1% of variance.

Simple regression analysis also revealed a significant relationship between children's performance on Task 1 and the AWRTE within both language groups, although this was not as highly predictive as that revealed for AWRTW reading tests (17.6% variance for AWRTE in Mostly Welsh group and 28.3% of variance in the Mostly English group).

The multiple regression model based on the variables which were revealed as significantly correlated with the AWRTW reading test within Group 1 was found to account for 53.7% of the variance in Welsh reading scores. Task 1 was

found to make a statistically significant unique contribution to the prediction of scores on the AWRTW reading test, making it a potential candidate as a marker of literacy abilities in Welsh.

Task 1 was not revealed to be making a statistically significant contribution to any other models analysed in relation to children's reading scores on the AWRTE, NRTW and NRTE, and so it is important to proceed with caution when concluding the role of Task 1 as a predictor of literacy difficulties in children from varying Welsh/English bilingual backgrounds. Further research is now needed in this area.

Limitations

As Task 1 was a multiple choice task with two, three or four answers to select from (chance performance set at 50%, 33.3% or 25%) it may not be a completely accurate reflection of children's morpho-phonological awareness. In future studies, a written or oral task to accompany this multiple choice task would provide a more holistic assessment. Also, with the absence of a 'back' button or ability for children to change their answers there may have been instances where they may have entered an incorrect answer in error and were unable to change it. Steps had been taken to try to avoid this through giving the children a clear instruction at the beginning that they could not change their answers themselves and if they wanted to change their answers they would need to inform the researcher. There may therefore have been instances where they did not want to admit that an error had been made and it would then be left incorrect although the child may have known the correct answer. Issues like these could affect the accurateness of the children's abilities. Nevertheless, this task provides an excellent starting point for exploring the relationship between Welsh mutation and children's reading abilities.

Reading Scores and Task 2 Multiple choice non-words cloze task

Task 2 was the most challenging of all the mutation tasks as it measured children's abilities to identify the correctly mutated/non-mutated novel word to complete the sentence. Across the whole sample Task 2 was significantly correlated with both English and Welsh All Wales reading tests and the NRTW. This result varied considerably when the sample was split into the two language groups. In the Mostly Welsh group significant correlations were only present with the All Wales reading tests (both Welsh and English) and within the Mostly English group there were no significant correlations. This would suggest an effect of language on children's performance on this task in that children from Mostly Welsh backgrounds who were higher ability readers performed significantly better on Task 2 than lower ability readers within the group. What is interesting to note is the raw score output (see Table 14) in relation to the mean scores of children within each of the groups. Contrary to performance on other tasks where the Mostly Welsh group performed better, the Mostly English group performed slightly better on Task 2 than their Mostly Welsh counterparts. This could be related to a group size effect or potentially that children who are from a predominantly English speaking background have better phonological awareness due to the opacity of English and are therefore better able to apply that awareness to what was a very difficult task. This does not however *explain* how children from English/Bilingual backgrounds would perform better than Welsh-speaking children on tasks related to Welsh mutation. Perhaps this could be explained by previous research into Welsh language acquisition (Bellin, 1988) which found that L2 children who learn Welsh have a better understanding of grammatical rules than L1 Welsh speakers due to being taught them rather than just acquiring them through language they've heard in the environment.

A simple regression analysis found Task 2 to account for 8.8% of the variance in scores in the AWRTW test and 9.7% of the variance in the AWRTE tests, both within Group 1 only. As Task 2 was not revealed as significantly correlated with the NRTE and NRTW it was not included in a simple regression

with these reading tests. A multiple regression analysis showed that Task 2 did not make a significant contribution within any of the models, in which case, it may not offer itself as a significant marker of literacy difficulties in Welsh within this study.

Limitations

Some difficulties emerged during the administration of Task 2. Despite using the practice items before carrying out the task some children struggled with the 'non-word' concept and so their completion of the task at times may have been based purely on guessing. Due to the nature of the task, when administering the practice items it was difficult to be clear whether the child did not understand the task or whether they did not have an awareness of the mutation, and so they were only corrected if all practice items were answered incorrectly. If they answered at least 1 correctly then this demonstrated an understanding of the task. However, this again could have been a 'lucky guess' and not a true reflection of the child's abilities. Due to the difficulties encountered by some of the children, it could be recommended that any future research exploring the use of non-words in mutation would be more appropriate for older children and may not yield 'clean' enough data in the assessment of children's literacy abilities at a young age.

Reading Scores and Task 3a: Grammaticality judgement task, and 3b: Sentence correction Task

Task 3 was administered in two parts and so both parts were analysed separately within this study. Task 3a involved children identifying whether the sentence they read on screen was correct or incorrect in relation to grammar (mutation being the focus) and Task 3b asked children to identify which word they believed was incorrect. Correlation analysis revealed that Task 3a was significantly correlated with all reading tests except the NRTE and Task 3b was significantly correlated with all reading tests. Some variation in this relationship emerged once the sample was split according to language group.

Within the Mostly Welsh group significant correlations emerged between all children's reading test scores except NRTE and their scores on Task 3a (as with the whole sample). A slight difference was seen within this group in that no significant correlation emerged between scores on Task 3b and the NRTE, which had been present when the sample was analysed as a whole. A significant change was seen within the Mostly English group. However, no significant correlations emerged between children's scores on the reading tests and their Task 3b scores, with the exception of the NRTE. This exception is interesting considering the NRTE was not correlated with either task within the Mostly Welsh group. A further simple regression analysis showed that within the Mostly Welsh group the scores on Task 3a accounted for 28.2% of the variance in the AWRTW reading test scores and Task 3b accounted for 30.5%, indicating very little difference between their predictability of the AWRTW scores. Task 3a accounted for 16.5% of the variance in the AWRTE scores and Task 3b also accounted for 16.5% of the variance in these reading scores, again showing very low predictability of reading scores. Task 3a accounted for 17.9% of the variance on the NRTW scores and Task 3b accounted for 21.4% of the variance in this reading test within the mostly Welsh group, which shows low predictability once again. Within the Mostly English group however, Task 3b accounted for 51% of the variance in the NRTE scores. This would suggest that Task 3b could be considered a significant predictor of NRTE Scores but not Welsh reading scores for children who are from predominantly English speaking backgrounds.

To explore this further Task 3a was included within multiple regression models related to AWRTW, AWRTE and NRTW. It was not found to make a significant contribution within any of the models. As a result, Task 3a was not considered to be a potential predictor of Welsh literacy difficulties within this study.

Limitations

During the administration of this task children at times would change their minds about their responses and due to the computer-based design of the

task they were unable to change their response once it had been selected. Although it was emphasised to the children that they needed to think carefully about their response before selecting as they could not change it afterwards, inevitably some children selected the wrong option and wanted to change their response. On these occasions I made a note of the change they wanted to make and updated their responses manually at a later date. The limitation here however is that this system depended on the children telling me about their error and there may have been occasions where they did not share this with me and therefore their scores may not have been completely representative of their abilities. Future tests of this nature would need to ensure there is an option for children to change their response themselves if required.

Reading Scores and Task 4a: Sentence repetition task, 4b: Sentence correction Task B

The oral element of the mutations tasks was administered in two parts. First was a sentence repetition task which asked children to listen to a sentence and repeat the sentence exactly as they had heard it, this was Task 4a. The second element of the task asked children to listen to the sentence again and judge whether they believed the sentence was grammatically correct or incorrect. If they believed it was incorrect they were asked to state what they believed was incorrect and provide the correct answer. Any errors that were present within a sentence were all related to mutation. Correlation analysis on the whole sample revealed that Task 4a (sentence repetition) was significantly correlated with all reading tests. When the sample was split into language groups there was a clear difference between the Mostly Welsh and Mostly English groups with the Mostly Welsh group displaying a significant correlation between all reading test scores and children's scores on Task 4a, and the Mostly English groups only revealing significant correlations between the NRTE and this task. This suggests a clear difference between children's performance on Task 4a and the relationship between these scores and the reading scores within each of the language groups. Simple regression analysis showed that

Task 4a accounted for 11.3% of the variation in scores on the AWRTW, 6.9% on the AWRTE, NRTE 24.5% and 18.4% in NRTW within the Mostly Welsh group. This would suggest that Task 4a was not a strong predictor of English or Welsh reading abilities within this group. Within the Mostly English group however this task accounted for 72% of the variance in the NRTE scores. This suggests that Task 4a is a strong predictor of children's scores on the NRTE for children who are L2 Welsh and a significant predictor of NRTE scores for L1 Welsh although not as strong (24.5%).

Task 4a was included within multiple regression models for predictions of AWRTW (Group 1 only), AWRTE (Group 1 only), NRTW (Group 1 and 2) and NRTE (Group 1 and 2). The task was found to make a significant unique contribution to the model for NRTW and NRTE scores within Group 1, however no significant contribution was found within any of the other models in relation to Task 4a in Group 2. There is potential therefore for sentence repetition tasks to play a role within screening tests which measure children's literacy abilities in Welsh.

Within the whole sample Task 4b was significantly correlated with the AWRTW and NRTW reading tests. When the sample was split according to language group it was revealed that Task 4b was significantly correlated with the AWRTW reading test only within the Mostly Welsh Group but with the NRTW within the Mostly English Group. This would suggest a significant link between children's Welsh reading scores and their performance on Task 4b. During simple regression analysis Task 4b was revealed as significant in relation to children's performance on the AWRTW task within the Mostly Welsh group, and accounted for 14% of the variance in these reading scores. It was also revealed as significant in the Mostly Welsh group accounting for 1.3% of the variance in the AWRTE scores. None of the findings of the simple regression analysis within the Mostly English group were significant in relation to this task.

Further analysis, including the types of errors which were made during the sentence repetition task may reveal some interesting findings as children were often noted to have automatically corrected words which had not been

mutated when they should have been. However, when questioned about whether the sentence was correct in Task 4b they were not always able to identify the automatic correction they had made when saying the sentence aloud.

Task 4b was included within the multiple regression models relating to prediction of scores on the AWRTW test (Group 1 only) and the NRTW (Group 2 only). No significant unique contribution was found in terms of Task 4b's predictability of reading test scores. This task was therefore not considered to be a potential predictor of reading abilities within this study.

Limitations

There was a significant negative relationship between the total number of children who repeated the sentence correctly and the length of the sentence (p=.002, r=-.690). This implied that as the sentence length increased the number of children who repeated it correctly decreased. This could suggest an effect of sentence length on the outcomes of this task. To explore this further a correlation analysis (Spearmans rho) was run between scores on the memory for digits task and Task 4a (sentence repetition) in both language groups. Results found that there was a significant positive correlation between Task 4a and scores on the Memory for digits tests within the Mostly English group (p=.568, r=.043). This may suggest a significant relationship between memory and children's scores on this task. This would also support the notion that sentence length may have played a significant role in the outcomes of this task. It would be recommended therefore that a task of this nature may benefit from using shorter sentences to avoid an effect of sentence length on test outcomes. It is worth noting however that this significant relationship only emerged when the sample was split and not when the sample was analysed as a whole.

Summary

The analysis within this study shows a clear correlation between children's performance on these tasks, which require morpho-phonological awareness skills, and their performance on the Welsh reading tasks in particular, as all tasks in this study have been significantly correlated with the AWRTW tests and NRTW (when the sample is analysed as a whole). This outcome is to be expected due to the tasks being based on the morpho-phonological system of mutation, which is a complex system within Welsh grammar. It would be assumed therefore that children who perform better on Welsh reading tests would also perform better on tests related to Welsh grammar as being a better reader would suggest having a better understanding or knowledge of grammar structures. What is interesting to note is that when the sample is split according to language group, Task 1 still continued to be significantly correlated with all reading tests whereas variation appeared between all other tasks and the English and Welsh reading tests. Task 1 was a comparatively high predictor of L1 Welsh children's Welsh reading scores as revealed by the regression analysis, accounting for 50.9% of the variance in scores on the AWRTW reading test. This type of test would therefore require further investigation to determine whether adaptations to the test would result in a different or stronger predictor outcome. Although other significant relationships were revealed, no other variables within the study were revealed to be strong predictors of L1 Welsh children's reading scores.

With regards to L2 Welsh children within the sample (Mostly English group), the sentence repetition task (Task 4a) was revealed to be the strongest predictor of the NRTE test within the Mostly English group whereas its comparative predictability for the Welsh reading tests was much lower. Sentence repetition tasks have been previously used as tests of language impairment (see Chapter 2) in various languages but in this particular study, although significantly correlated with all reading tests (in particular Welsh reading tests) it does not appear to be a strong predictor of children's Welsh reading abilities. This raises the questions as to why is it therefore a strong predictor of English reading abilities within the Mostly English group? There

could be a link here with the fact that some children in the sample automatically corrected the sentences when they were being repeated rather than repeating the sentence exactly as they heard it. It could be assumed that the Welsh L1 children would predominantly do this due to naturally mutating when speaking their mother tongue. English L1 children may however be more inclined to repeat the sentence as they hear it as they would not have the same automaticity when speaking Welsh and therefore may not mutate automatically in the same way as an L1 Welsh speaker. However, it is not possible to determine this within the parameters of this study as this element was not formally recorded when the children carried out the task. It is also important to note the small group size within the study, which would make it unwise to draw any definitive conclusions. Further investigation with a larger group of Mostly English speaking children would be required to determine whether this finding is representative of a larger sample.

It is worth noting that there was a correlation between English reading tests and scores on the Memory for digits task across the whole sample, but there was no correlation between scores on this memory task and Welsh reading tests at all. Could this suggest that English reading tests require more engagement with memory due to its opaque nature and therefore tests which are reliant on good levels of short term memory, such as the sentence repetition task, would be significantly correlated with English reading tasks within a group whose predominant language is English but who have some degree of Welsh knowledge? This suggestion may go some way to explaining why Task 4a was revealed to be a strong predictor of children's scores on the NRTE within the mostly English group.

With regards to the tasks as a whole it would appear that for children who were Mostly Welsh speakers Task 1 was revealed to be the best predictor of Welsh reading scores, in particular in relation to the AWRTW reading test. As such, a further exploration of the use of this kind of test, perhaps with other mutation triggers and other elements of literacy (i.e. writing and oral) would be beneficial.

Conclusion

This study raised some interesting findings in relation to children's performance on reading tests and tasks which measure morpho-phonological ability. Significant correlations with other standardised tests used within the study, such as those measuring phonological awareness, showed that some of the novel tasks did carry some validity as a measure of morpho-phonological awareness in Welsh. Although there were some clear limitations to the tests, the findings do suggest that tests of children's Welsh mutation ability could play a role in the development of future screening tests which measure children's literacy abilities. Further research however is needed in this area with a larger sample and using differing triggers of Welsh mutation to further develop understanding of the role of mutation in diagnosing children's literacy difficulties. Use of qualitative analysis would also be beneficial in relation to identifying the types of errors children made in these tasks and combining this with other methods such as time taken on task, pace, effort and number of stop starts to provide a more holistic view of a child's literacy abilities.

The next chapter presents the findings of Study 3 which explores the findings of the plural morphology tasks and their relationship with children's reading scores and phonological awareness.

CHAPTER 6

STUDY 3

Exploring the relationship between Welsh-English children's performance on plural morphology tasks and their reading abilities

Introduction

The previous chapter presented the findings of Study 2 that explored children's performance on a range of novel tasks measuring children's knowledge of Welsh mutation, and the relationship between performance on these tasks with reading scores and other background measures (age, IQ, short term memory and phonological processing in English and Welsh). The focus of Chapter 6 is on Welsh plural morphology and its relationship with reading scores and other background measures (see chapter 4 for description of measures). There is currently only a small body of research into Welsh plural morphology (for examples see Thomas et al., 2014; Binks & Thomas, 2019; Chondrogianni & John, 2018) but no previous research has explored the relationship between children's reading ability and their knowledge of or ability to apply plural morphology in a written task. For this reason, this study also takes an explorative approach, investigating relationships between performance on a set of novel plural production tasks and reading, and also how performance on these novel tasks are related to their performance on a series of background measures. As no tests currently exist to assess children's abilities in relation to plural morphology it was necessary to create original tests and these measured children's knowledge of plural morphology both orally and in writing. These tasks are detailed below.

This study aims to address the following research questions:

- Is there a relationship between children's performance on Welsh plural morphology tasks and their reading abilities? If so, what is the nature of this relationship and does it vary between language groups?
- Can tests of Welsh plural morphology predict children's performance on reading tests?

In relation to the first research question, and assuming that reading difficulties are linked to phonological awareness difficulties, it was predicted that there would be a significant correlation between children's reading ability and their performance on tasks that involve knowledge of plural morphology. It was also predicted that this relationship would vary across the two language groups given that children from L1 Welsh backgrounds would have more exposure to plural forms than other children.

In relation to research question 2, it was predicted that, if plural morphology tasks required high level phonological awareness skills and correlated strongly with reading abilities, plural morphology tasks would be a predictor of reading abilities.

Measures

Novel Experimental Tasks: Study 3 – Plural Morphology

A series of paper-based written tasks and one computer based task were used to measure children's ability to apply their knowledge of plural morphology in Welsh to real words and non-words. Task 1 was a simple singular to plural conversion task to measure children's basic ability to pluralise. Task 2a developed this further by using more complex real word singular to plural conversion patterns and Task 2b assessed children's ability to identify the plural pattern in Task 2a and apply it in a non-word scenario in Task 2b. The design of these tasks was loosely based on previous research by Thomas *et al.* (2014) who used an oral production task to assess children's knowledge of plurals in Welsh.

Plurals Task 1: singular-plural conversion in written form

Task 1 contained a series of 5 pictures with the image of the item and the name underneath in its singular form. Children were asked to write the plural version of this item underneath the image.

Ysgrifennwch y gair am mwy nag un… ('Write the word for more than one…')

Iâr = _____   Ffon = _____

The aim of this task was to put children in the mind-set of thinking about plurals through introducing them to a simple task, hence the use of only 5 items. This was regarded as more of a practice task to assess children's ability to complete the tasks. If they had great difficulty in completing this task then it was likely that Task 2 would have been beyond their abilities. However, none of the participants had difficulty with this introductory task, and they were all, therefore, introduced to the subsequent tasks.

Plurals Task 2: singular-plural conversion application to novel words

Plural Task 2 was in two parts: Part 1 (a) involving real words, and Part 2 (b) involving novel words. In both cases, children were required to read the singular noun (real or novel) provided and then write next to it what they believed the plural version of that noun to be. A total of 16 real singular nouns were presented and a further 16 novel words based on the structure of the real noun, for example, real word: *cwpan* (cup); novel word: *lwpan*. The items

were selected to include the common suffixes often found in Welsh plural morphology e.g. *iau* (*het-het__iau__* 'hat – hats'), *od* (*cath-cath__od__* 'cat-cats'), progressing to more complex items as the task developed e.g. removal of the circumflex followed by the addition of '*i* as in *pêl-p__e__l__i__* 'ball-balls', single vowel conversion with suffix e.g. *cwch – c__y__ch__od__* 'boat-boats', *cloch-cl__y__ch__au__* 'bell-bells', *draig-dr__eig__iau__* 'dragon-dragons', double vowel conversion with suffix e.g. *cwpwrdd – c__y__p__y__rdd__au__* 'cupboard-cupboards' and a double vowel conversion only e.g. *castell-c__e__st__y__ll* 'castle-castles'. The task became gradually more difficult and it was decided that 16 items would be sufficient to enable an assessment to be made on children's knowledge of plural morphology in Welsh through ensuring at least one example was provided of each plural conversion type. The novel words were created to ensure their structure was the same as the singular real word in terms of letter sounds/syllables and location of the vowels which would need to be converted, and included Welsh sounds such as the /ð/ *dd* and /r̥/ *rh* to ensure they were recognisable as Welsh non-words. The LNF in Wales refers to children being able to spell an increasing number of plural forms (both regular and irregular) correctly within the writing element of the framework and so will be an area which children will have been taught and therefore have some familiarity with.

The first two items (see Examples 1 and 2 below) were practice items to ensure participants understood what they were being asked to do.

Both tasks were administered in small groups under exam conditions.

Practice items provided in Plural Task 2a and 2b:

Rhif (No.)	Gair unigol (singular noun)	Gair lluosog (Plural word)	Gair-ffug unigol (Singular non-word)	Gair-ffug lluosog (Plural non-word)
1	cath	*cathod*	path	*pathod*
2	het	hetiau	let	*letiau*

Plurals Task 3: Oral production

Plurals Task 3 was related to the computer based activity – *Y Goriad* – described in Chapter 5. Children were asked to listen to a word by clicking on the treasure chest coin on the screen. They were then asked to say the plural form of the singular word they had just heard into the microphone. There were 11 items within this task, a number which enabled at least one example of vowel change within the singular-plural conversion to be included. The aim of this task was to assess whether children were able to correctly convert a singular noun into a plural. The items on the task focused on singular-plural conversions that involved a change in the vowel sound e.g. *mynydd* /mənɨ̞ð/ 'mountain' – *mynyddoedd* /mənə̞ðɔɨð/ 'mountains', d*yn* /d*ɨ̞n*/ 'man' - *dynion* /d*ə*nɪɔn/ 'men' and also a complete change in the vowel letters e.g. *car* /k*æ*r/ 'car'– *ceir* /k*eɪ*r/ 'cars', *bwrdd* /b*u*rð/ 'table'-*byrddau* /bərðæɨ/ 'tables'. Most of the words involved adding one of the common suffixes used in Welsh when converting from singular to plural e.g. *-oedd, -au, -io, -iau* and so the task was assessing children's ability to identify the correct suffix and also pronounce the vowel sounds correctly when converting the singular word into a plural. As the focus was primarily on the pronunciation of the vowel sound, the task consisted of a range of examples where this happens in Welsh nouns and added to this was some more complex yet easily identifiable conversions such as *car-ceir* which the majority of Welsh children would be able to identify as referred to within the LNF.

Children were given the option to listen to the word twice if necessary. The researcher took responsibility for the recording of the audio by clicking on the microphone icon. This task was administered directly after the mutations oral task (Task 4) to enable all oral tasks to be completed in one session (see appendix 9 for full list of items).

Results

Due to the data not being normally distributed in all cases, a Spearman's rho correlation analysis was run on the sample as a whole (using IBM SPSS Statistic 25) to explore whether any relationships existed between children's scores on the plurals tasks and their reading scores, as well as the background measures of IQ, memory for digits and phonological awareness. These results are presented below.

Due to the very short nature of the task (5 items), and it being used primarily as an initial gauge of plural ability, Plurals task 1 was not included in any analysis as no clear conclusions could be drawn from any results that emerged.

Plurals tasks and Age, IQ, Memory for digits and letter recognition

The first set of analyses, presented in Table 26 below, explored the correlational relationship between the three plural morphology tasks and the background measures of Age, IQ, Letter recognition (Welsh and English), and memory for digits.

Table 26: Correlation results of Plurals tasks and background measures.

	Age		IQ		LRE		LRW		Memory for digits	
	r	p	r	p	r	p	r	p	r	p
Plurals Task 2a	.256	.036	.260	.033	-.118	.350	-.408	.001	-.025	.844
Plurals Task 2b	.305	.012	.253	.039	-.226	.070	-.387	.001	.140	.263
Plurals Task 3	.221	.084	.079	.543	-.075	.564	-.125	.337	-.047	.718

As can be seen in Table 26, a significant positive correlation was found between Plurals Task 2a and age (r=.256, p=.036), and Plurals Task 2b (r=.305, p=.012) and age, suggesting that as the sample increased in age, so did their performance on both parts of Task 2, although the stronger correlation was with Plurals Task 2b. A significant positive correlation was also present between Plurals Task 2a and IQ (r=.260, p=.033) and Plurals task 2b and IQ (r=.253, p=.039), suggesting that children with higher IQ perform better on both parts of this task than those with lower IQ, although the strength of these correlations are weak. A significant negative correlation was found between LRW and Plurals task 2a (r=-.408, p=.001) and Plurals Task 2b (r=-.387, p=.001), suggesting that children who performed better on these plurals tasks made fewer errors when identifying letter sounds in Welsh. No significant correlations were found between Plurals Tasks 2a or 2b and LRE or Memory for digits.

No significant correlations were revealed between Plurals Task 3 and age, IQ, letter recognition or Memory for digits suggesting that children's performance on this task is not related to these variables. It is unclear why no relationship was found, particularly with age and IQ. However, one possible explanation could relate to language, in that children who have a higher exposure to Welsh are more likely to have been exposed to the plural forms included in Task 3.

Phonological awareness tasks correlations

Performance on the plural morphology tasks was correlated with performance on the phonological awareness tasks. Results are presented below in Table 27.

Table 27: Correlation results of Plurals tasks and phonological awareness tasks.

	Elision		Blending Words		Phoneme isolation		Initial Sound W		Final Sound W		Phoneme blending W		Phoneme deletion initial W		Phoneme deletion final W	
	r	p	r	p	r	p	r	p	r	p	r	p	r	p	r	p
Plurals Task 2a	.306	.013	.356	.004	.400	.001	.241	.051	.063	.613	.319	.009	.326	.008	.301	.014
Plurals Task 2b	.439	.000	.213	.088	.365	.003	.159	.202	.130	.300	.288	.019	.431	.000	.262	.034
Plurals Task 3	.223	.084	.374	.003	.198	.130	-.026	.844	-.018	.891	.267	.036	.045	.731	.293	.021

Plurals Task 2a (real words)

A significant positive correlation was found between Plurals Task 2a and Elision ($r=.306$, $p=.013$), Blending words ($r=.356$, $p=.004$), Phoneme Isolation ($r=.400$, $p=.001$), Phoneme Blending Welsh ($r=.319$, $p=.009$), Phoneme Deletion Initial W ($r=.326$, $p=.008$) and Phoneme deletion final ($r=.301$, $p=.014$), and a weak significant correlation with Initial Sound Welsh ($r=.241$, $p=.051$). This suggests that children who perform better on these tasks of phonological awareness also perform better on Plurals Task 2a. What is interesting to note here is the correlation with the English phonological awareness tests. Given that these are derived from a standardised measure (CTOPP-2) it provides some strength to the validity of Plurals Task 2a.

Plurals Task 2b (non-words)

In relation to Plurals Task 2b, a significant positive correlation was revealed between this task and Elision ($r=.439$, $p=<.001$), Phoneme Isolation ($r=.365$, $p=.003$), Phoneme Blending Welsh ($r=.288$, $p=.019$), Phoneme Deletion Initial W ($r=.431$, $p=<.001$) and Phoneme Deletion Final ($r=.262$, $p=.034$) with the latter two having the weakest correlation. This suggests that children who perform better on Plurals task 2b also perform better on these particular tests of phonological awareness.

Plurals Task 3

A significant moderate positive correlation emerged between Plurals Task 3 and Blending words ($r=.374$, $p=.003$), with weaker correlations emerging with Phoneme blending Welsh ($r=.267$, $p=.036$) and Phoneme deletion final W ($r=.293$, $p=.021$). No other significant relationships emerged in relation to Plurals Task 3 and phonological awareness scores.

Plurals tasks and reading scores

Table 28 displays the output of a correlational analysis between the plurals tasks and children's performance on the AWRT and NRT in both English and Welsh.

Table 28: Whole sample Plurals tasks and reading scores correlations results.

Variable	AWRTW (n64)		AWRTE (n64)		NRTW (n 42)		NRTE (n 37)	
	r	p	r	p	r	p	r	p
Plurals Task 2a	.595	.000	.389	.001	.582	.000	.507	.001
Plurals Task 2b	.498	.000	.430	.000	.360	.019	.441	.006
Plurals Task 3	.352	.006	.108	.407	.190	.254	.144	.410

Significant correlations emerged between children's scores on Plurals Task 2a and 2b and their reading scores in both English and Welsh on both the AWRT and NRT. A positive significant correlation was found between Plurals Task 2a and the AWRTW (r=.595, p=<.001), the AWRTE (r=.389, p=.001), NRTW (r=.582, r=.000) and the NRTE (r=.507, p=.001). This suggests that lower ability readers, regardless of the language being tested, did not perform as well as higher ability readers on these tasks. The same also applied to Plurals Task 2b where significant positive correlations emerged with the AWRTW (r=.498, p=<.001), AWRTE (r=.430, p=<.001), the NRTW (r=.360, p=.019) and the NRTE (r=.441, p=.006).

In relation to Plurals Task 3, a significant positive correlation was found with the AWRTW only (r=.352, p=.006), suggesting that children who scored higher on this particular reading task also performed better on Plurals Task 3. However, as was also the case with the other variables included in the

previous correlations, no significant correlations were revealed between Plurals Task 3 and the AWRTE, NRTW and NRTE, which may suggest that Plural Task 3 is weak at distinguishing between different types of learners in relation to their reading ability.

The first research question under investigation was

1. Is there a relationship between children's performance on Welsh plural morphology tasks and their reading abilities? If so, what is the nature of this relationship and does it vary between language groups?

In relation to part 1, there is clearly a relationship between children's performance on the Welsh plural morphology tasks and their reading abilities. In particular Plural Task 2a (real word) and Plural Task 2b (non-word) revealed significant positive correlations with both the AWRT and NRT in English and in Welsh, suggesting that higher ability readers perform better on these tasks than lower ability readers and therefore distinguishing between groups of learners.

In order to answer part two of research question 1 the sample was split into two groups according to language background. Group 1 was 'Mostly Welsh' and Group 2 was 'Mostly English' (see chapter 4 for a breakdown of each group). A further analysis was run on the results of these groups, this is presented below.

Raw scores

Table 29 below displays the mean average raw scores per child in Group 1 (Mostly Welsh) and Group 2 (Mostly English).

Table 29: Group 1 and 2 raw scores for all plural tasks.

	Group 1 (52)	Group 2 (15)
Plurals Task 2a (16)	8.38 (52) S.D: 4.020	7.40 (n15) S.D: 3.397
Plurals Task 2b (16)	6.25 (52) S.D: 3.597	6.60 (n15) S.D: 2.823
Plurals Task 3 (11)	7.51 (n49) S.D: 2.053	4.92 (n12) S.D: 2.314

Results show very little difference between Group 1 and 2's performance on Plurals Task 2a, but the mean score is higher within Group 1 as would be expected due to this being a mainly Welsh-speaking group. Group 2 appear to have performed marginally better than Group 1 on Plurals Task 2b but on the whole Group 1 appear to be the higher scorers, as would be expected. What's interesting to note however is that in Plurals Task 2b there is a marginal difference in scores but it is Group 2, the mostly English group, which appears to have performed slightly better. This was not predicted due to this being a non-word task dependent on children's ability to use the morphological pattern identified from Plurals Task 2a and applying this pattern to the non-word, and therefore it would be expected that children from Welsh L1 backgrounds would have a better knowledge and understanding of these plural forms, although group size may be accountable for this difference. Our predictions are satisfied with Plurals Task 3 however where Group 1 average scores are much higher than those of Group 2, which is interesting given the very few correlations which emerged between Task 3 and reading scores, background measures and phonological awareness scores. An independent t-test revealed no significant differences between children's language group and the mean scores on Plurals Task 2a and 2b. A significant difference was however revealed between group mean scores on Plurals Task 3 ($p<.001$) with a high Cohen's d point estimate of 1.233. This suggests a statistically significant difference between children's performance on Plurals Task 3 and their language background as in indicated by the mean scores in Table 29.

Correlations per group

Group 1: Mostly Welsh

Table 30: Group 1 Plural tasks correlation results.

Variable	AWRTW (n50)		AWRTE (n50)		NRTW (n 32)		NRTE (n 28)	
	r	p	r	p	r	p	r	p
Plurals Task 2a	.606	.000	.465	.001	.564	.001	.449	.016
Plurals Task 2b	.519	.000	.462	.001	.324	.070	.303	.117
Plurals Task 3 Oral	.360	.013	.196	.178	.120	.527	.178	.373

Group 2: Mostly English

Table 31: Group 2 Plurals tasks correlations results.

Variable	AWRTW (n50)		AWRTE (n50)		NRTW (n 32)		NRTE (n 28)	
	r	p	r	p	r	p	r	p
Plurals Task 2a	.416	.139	.405	.134	.722	.018	.829	.003
Plurals Task 2b	.562	.036	.345	.208	.520	.120	.756	.011
Plurals Task 3 Oral	.140	.665	.113	.726	.733	.039	.518	.188

Plurals Task 2 (groups)

As shown in Table 30 above there was a positive significant correlation between children's performance on the AWRTW reading test and their performance on all plurals tasks within Group 1. This is similar to the findings of the whole sample analysis which also found significant correlations between performance on Plurals tasks and scores on the AWRTW. However, this is not echoed in the analysis of Group 2 (table 31) which showed only Plurals Task 2b to be significantly correlated with the AWRTW ($r=.562$, $p=.036$).

There was a positive significant correlation between the AWRTE and Plurals Task 2a ($r=.465$, $p=.001$) and Plurals Task 2b ($r=.462$, $p=.001$), with almost identical output. This suggests that higher ability readers on the AWRTE perform better on these tasks than lower ability readers. No significant correlations emerged between Plurals task 2a and 2b and the AWRTE scores within Group 2.

Scores on the NRTW and Plurals Task 2a were positively correlated in both Group 1 ($r=.564$, $p=.001$) and Group 2 ($r=.722$, $p=.018$), suggesting that higher ability Welsh readers (according to the NRTW scores) performed better than lower ability Welsh readers on this task, regardless of language background. There were no significant correlations between the NRTW scores and Plurals Task 2b scores in either group, although the sample as a whole showed a positive correlation.

The NRTE scores showed a positive significant correlation with Plurals Task 2a within Group 1 ($r=.449$, $p=.016$) and also within Group 2 ($r=.829$, $p=.003$), indicating that higher ability English readers (according to scores on the NRTE) performed better on Plurals Tasks 2a than lower ability readers regardless of language background. In relation to Plurals Task 2b however no significant correlation was found with NRTE scores in Group 1 but there was a significant positive correlation within Group 2 ($r=.756$, $p=.011$), showing a difference in relationship between the two groups.

Plurals Task 3 was found to have only one significant positive correlation in Group 1 which was with the AWRTW ($r=.360$, $p=.013$), suggesting that higher ability Welsh readers were better able to correctly convert and articulate Welsh singular nouns into plural nouns than lower ability Welsh readers. In Group 2 however the only significant correlation was with NRTW ($r=.733$, $p=.039$). Results indicate that this task therefore was only significantly correlated with Welsh reading scores and not with English reading scores.

In order to further explore the significant correlations between the plurals tasks and children's reading scores a simple regression analysis was run on

each variable that had emerged as significantly correlated with each of the reading tests.

Regression Analysis

A simple regression analysis was run between each of the plurals tasks and their performance on each of the reading tests, in both Group 1 and Group 2, to explore whether any of the tasks could be considered predictors of reading ability. Only variables which had arisen as significantly correlated with each of the reading tests within this study were included within each analysis. The same process was applied to the simple regression used for Group 2; however, in this case, fewer regression analyses were required due to fewer significant correlations emerging within this group between reading scores and performance on the plurals tasks.[15] A total of 7 simple regression analyses were run with Group 1 participants and a total of 5 were run for Group 2. The results of these analyses are presented below.

---

[15] It is also worth noting that, given the unequal samples across both groups, that the results relating to Group 2 should be treated with caution, despite there being on-going debates about the minimum number of samples that are necessary for meaningful analyses.

Table 32: Results of simple regression analysis between Plurals tasks and dependent variable: AWRTW.

Task	Group 1					Group 2				
	R^2	Sig	SE	Unstandardised B	Standardised Beta	R^2	Sig	SE	Unstandardised B	Standardised Beta
Plurals Task 2a	.374	<mark>.000</mark>	.440	2.357	.612	-	-		-	
Plurals Task 2b	.274	<mark>.000</mark>	.528	2.248	.523	.197	.112	1.143	1.963	.444
Plurals Task 3	.090	<mark>.041</mark>	1.086	2.286	.299	-	-		-	

Simple regression analysis shows that all Plurals tasks significantly account for the variance in reading scores within Group 1. In particular, Plurals Task 2a accounts for the highest level of variance at 37.4% followed by Plurals Task 2b at 27.4% and Plurals Task 3 at 9%. No significance emerged between the Plurals tasks and the AWRTW in Group 2, suggesting that this trend applied only to children from a Welsh L1 background.

Table 33: Results of simple regression analysis between Plurals tasks and dependent variable: AWRTE.

Task	R^2	Sig	SE	Unstandardised B	Standardised Beta
				Group 1	
Plurals Task 2a	.193	.001	.417	1.440	.439
Plurals Task 2b	.211	.001	.460	1.685	.460

Plurals Task 2a and 2b significantly accounted for the variance in reading scores on the AWRTE in Group 1. Task 2a accounted for 19.3% of the variance and Task 2b accounted for 21.1% of variance in scores. No correlations had emerged between the Plurals Tasks and the AWRTE in Group 2 and therefore these were not included in any regression analysis.

Table 34: Results of simple regression analysis between Plurals tasks and dependent variable: NRTW.

Task	Group 1					Group 2				
	R^2	Sig	SE	Unstandardised B	Standardised Beta	R^2	Sig	SE	Unstandardised B	Standardised Beta
Plurals Task 2a	.283	.002	.531	1.826	.532	.360	.067	1.249	2.652	.600
Plurals Task 2b	-	-		-		-	-		-	
Plurals Task 3	-	-		-		.262	.195	2.275	3.318	.512

Regression results within Group 1 found that Plurals Task 2a accounted for 28.3% of the variance in NRTW scores. In Group 2 however, there was no significant effect of Plurals task 2a on NRTW scores. Plurals Task 3 also appeared to have no significant effect.

Table 35: Results of simple regression analysis between Plurals tasks and dependent variable: NRTE.

Task	Group 1					Group 2				
	R^2	Sig	SE	Unstandardised B	Standardised Beta	R^2	Sig	SE	Unstandardised B	Standardised Beta
Plurals Task 2a	.241	.008	.572	1.644	.491	.637	.006	1.062	3.981	.798
Plurals Task 2b						.560	.013	1.408	4.490	.748

Plurals Task 2a was found to significantly account for the variance in scores on the NRTE in Group 1 (24.1%) and in Group 2 (63.7%). Plurals Task 2b also accounted for 56.0% of the variance in NRTE scores in Group 2. Suggesting that these tasks could be useful predictors of the English reading scores of children from a Mostly English speaking background. Table 36 below provides a summary of the findings from the regression analysis for Study 3.

Due to the outcomes of the simple regression analysis it was decided that multiple regression analysis would not be necessary based on the small number of tasks being analysed and the low level of variance shown in the simple regression output.

Table 36: Summary of Study 3 regression analysis findings.

Task	Reading Test	Percentage of variance	Group
Plurals Task 2a: Written singular-plural conversion task (real words)	AWRTW	37.4%	Group 1 only
	AWRTE	19.3%	Group 1 only
	NRTW	28.2%)	Group 1 only
	NRTE	24.1% 63.7%	Group 1, Group 2
Plurals Task 2b: Written singular-plural conversion task (novel words)	AWRTW	27.4%	Group 1 only
	AWRTE	21.1%	Group 1 only
	NRTE	56%	Group 2 only
Plurals Task 3: Oral production	AWRTW	9%	Group 1 only

Discussion

The aim of Study 3 was to explore the relationship between children's scores on tasks that measured their knowledge/understanding of Welsh plural morphology and their reading scores in both English and Welsh. Similar to the questions posed in Study 2, Study 3 focused on Welsh plural morphology to explore the following:

1. Is there a relationship between children's performance on Welsh plural morphology tasks and their reading abilities? If so, what is the nature of this relationship and does it vary between language groups?
2. Can tests of Welsh plural morphology predict children's performance on reading tests?

As with Study 2, the analysis of Study 3 involved initial whole sample correlational analysis of children's performance on plurals tasks with reading scores and background measures of phonological awareness, IQ, age and memory for digits. The sample was then split according to language group and analysis run on children's plurals tasks scores and reading scores. A further regression analysis was then run using the findings of the correlation analysis to inform the regression models.

Phonological awareness and Plurals Tasks scores

Promising results were found in relation to the Plurals Tasks and phonological awareness ability. Plurals Task 2a correlated significantly with all phonological awareness tasks with the exception of final sound W. Task 2b was also significantly correlated with many of the phonological awareness tasks. What is of particular note is that both significantly correlated with all three tasks of the CTOPP-2. As this is a standardised measurement of phonological awareness it shows encouraging results for the validity of the plurals tasks as a measurement of higher level phonological awareness.

Reading scores and Plurals Task 2a and Task 2b

Plural Task 2 was split into 2 tasks. Task 2a was a task that measured children's abilities to convert a singular noun into a plural applying knowledge of different conversion patterns and Task 2b involved children using the same pattern identified in Task 2a and applying that to a non-word (e.g. *cath-cathod, path-pathod)*. Correlation analysis of the whole sample revealed a

significant relationship between children's scores on this task (2a and 2b) and their scores on all reading tests within the Study. This suggested that higher ability readers in both English and Welsh performed better on this task than lower ability readers. In order to explore any language effects analysis was also run on the separate language groups (Group 1 – *Mostly Welsh* Group 2 - *Mostly English*). Analysis revealed that a significant relationship was present between children's scores on Task 2a and their reading scores on all reading tests within the Mostly Welsh group. Within the Mostly English group there was a significant relationship between Task 2a and all reading tests except the AWRTE. This would suggest that higher ability Welsh readers perform better on this task than lower ability Welsh readers regardless of their language group.

Analysis of Task 2b revealed a significant relationship between children's scores on this task and their scores on both the AWRTE and AWRTW within the Mostly Welsh group and between the AWRTW and NRTE within the Mostly English group, suggesting that some difference did exist between language group and performance on this task.

To investigate these relationships further a simple regression analysis was run between all variables which had emerged as significantly correlated within each language group. Within the Mostly Welsh group performance on Task 2a accounted for 37.4% of the variance in the AWRTW Reading test, 28.3% of the variance in the NRTW, 19.3% of the variance in the AWRTE and 24.1% of the variance in the NRTE reading test. In the Mostly English group, however, the only significant result was between children's scores on Task 2a and the NRTE, accounting for 63.7% of the variance in reading scores. Due to the small sample size however it is difficult to draw definite conclusions and therefore further research using a larger sample size may help to explore the significance of this variance for Mostly English bilinguals further.

Scores on Task 2b accounted for 27.4% of the variance in the AWRTW and 21.1% of the variance in the AWRTE within the Mostly Welsh group. Within the Mostly English group the scores on this task accounted for 56.0% of the variance in the NRTE scores, which is closely linked to the outcome of scores

on Task 2a within this group. There is therefore a clear relationship between L2 Welsh children's scores on Plurals Task 2b and their English reading abilities. This might suggest that for L2 Welsh-speaking children, tasks of this kind could have a role in predicting children's English reading scores, which could be useful in the design of future Welsh-English bilingual tests.

Limitations

During the marking of these tasks spelling/punctuation were marked as incorrect, however, this would not necessarily mean that the child was unable to pluralise correctly but simply that they were not able to spell/punctuate the words correctly. One example was when children had not included a circumflex e.g. *pêl* was written as *pêli*, which is incorrect as the 'e' is not elongated within the plural form of the word. This is where an oral production task such as Plurals Task 3 helps to provide an indication of children's plural abilities which are not reliant on spelling/punctuation of the words. Tests which measure both written and oral aspects of plural morphology are therefore useful as a measure of children's abilities.

Reading scores and Plural Task 3

Task 3 was an oral production task which required children to listen to a word in its singular form and then say the plural form of the word into the microphone. Correlation analysis of the whole sample revealed a significant relationship between children's scores on this task and their All Wales reading test scores, implying that higher ability Welsh readers performed better on this task than lower ability Welsh readers. This finding was mirrored within the Mostly Welsh group when the sample was split into language groups however within the Mostly English group the only significant relationship which emerged was with children's scores on the NRTW. When this was explored further using a simple regression analysis the results showed that children's scores on this task significantly accounted for only 9% of the variance of scores on the AWRTW reading test within the Mostly Welsh group. No other

significant relationships were revealed which would suggest that this task is not a strong predictor of children's reading scores regardless of language background.

Limitations

Being a test which required use of a microphone some children chose not to take part and so perhaps a task where they would not be required to do this may be preferable, or perhaps the use of technology which would enable their voices to be recorded from a distance (rather than holding a microphone).

There was a very clear difference in children's mean raw scores within this task which would suggest that L1 Welsh speakers perform better than L2 Welsh speakers. This would support the notion that language exposure plays a significant role in children's ability to apply plural morphology in Welsh.

Conclusion

The results of this study suggest that plural morphology tasks may reveal a significant relationship with bilingual children's reading abilities and may therefore play a role in predicting those reading abilities in bilingual children from varying Welsh-English bilingual backgrounds. Further research is now needed to explore this grammatical structure with a more balanced sample of Welsh-English bilinguals.

The next chapter presents a general discussion of the key findings from all three studies within this thesis and concludes with recommendations for future research/practice.

Chapter 7

<u>**GENERAL DISCUSSION**</u>

The studies presented in this thesis were designed to explore potential markers of literacy difficulties in Welsh. Literacy difficulties in English are often characterised by poor phonological awareness skills that are typically measured via various tests requiring real and non-word processing, manipulation and production. However, due to the transparent nature of Welsh orthography, such tasks may not always identify underlying difficulties, particularly where children can rely on the transparent grapheme-phoneme correspondences that the language affords. Consequently, it was necessary to explore other aspects of Welsh that could be hypothesised to be significantly correlated with literacy abilities, but which also pose a level of phonological awareness difficulty for the learner. In the present thesis, these aspects included mutation and plural morphology in Welsh. Study 1 explored the types of errors bilingual Welsh-English children make in their written work, both in English and in Welsh, and Studies 2 and 3 explored the relationship between children's performance on reading tests, traditional tests of phonological awareness, and novel tests of mutation and plural morphology. The selection of these particular structures was influenced in part by the outcomes of Study 1 (which provided an error analysis of the types of errors made by children aged from 6 to 12 years of age), and previous research in relation to specialist dyslexia teachers' views about the particular aspects posing difficulties to Welsh-English bilinguals in Welsh (Davies, 2016).

The findings of the three studies are summarised in Table 37 below:

Table 37: Summary of thesis findings (Studies 1, 2 and 3).

	General pattern of findings
Study 1	• Errors in English tend to focus around Phonological Error and Spelling Error. • Errors in Welsh tend to be more variable, but are often focused around mutation and spelling. • Mutation errors increase with age, possibly influenced by the vernacular.
Study 2	• All Mutations Tasks correlated significantly with other phonological awareness tasks, in particular Tasks 1, 3a and 3b. • All Mutations Tasks were highly correlated with both Welsh and English reading scores (with the exception of Task 4b which did not correlate significantly with any English reading task) across the whole sample and more so in Group 1 (mostly Welsh) than Group 2 (mostly English) when the sample was split. • Task 1 was found to be the highest predictor of reading abilities in Group 1. • Task 4a was the highest predictor of English reading scores (NRTE) in Group 2.
Study 3	• Plurals Task 2a was highly correlated with phonological awareness tasks. • Plurals Task 2a and 2b were highly correlated with all reading scores. • Plurals Task 2a was highly correlated with all reading scores in Group 1. • Plurals Task 2a significantly accounted for variance in scores on both Welsh and English reading scores, more so in Group 1 than Group 2. • Plurals Task 2a and 2b were the highest predictors of English reading ability (NRTE) in Group 2.

Together, the results of the three studies revealed some interesting findings and lead to some useful implications and recommendations for further study. These findings are discussed further below.

Patterns of errors in Welsh-English bilinguals' written texts (Study 1)

One of the major drivers behind this thesis was to identify any patterns of Welsh language writing that may highlight potential literacy problems to teachers, in order that pupils could be referred for assessment earlier than they tend to do so at present. As outlined in Chapter 3, findings revealed clear differences between children from different home-school language backgrounds who scored lower and higher on standardised reading tests. Age was a particularly significant factor within the Welsh at school bilingual group and correlated significantly with many of the error types, including those requiring higher levels of phonological awareness such as Welsh mutation. Within this particular group of bilinguals older children made fewer errors in their writing than younger children. Higher ability English and Welsh readers within the Welsh at school bilingual group on the whole, made fewer errors than lower ability readers. This group also produced more words on average than the Welsh dominant bilinguals and were almost level with the English dominant group, which supports the notion that bilinguals have a higher vocabulary than children who are dominant in one language (e.g. Junker and Stockman, 2002; Pearson, Fernandez and Oller, 1993).

The types of errors produced across the sample varied, with errors in English writing being primarily related to phonological and spelling errors, which would support the notion that the orthographic depth of a language may influence its acquisition (see Chapter 2). This notion is further supported by the finding that the Welsh writing errors (Welsh being a more transparent language) appeared to be more variable yet focused around elements of the language which require higher levels of phonological awareness, namely mutation, and spelling errors. Within both bilingual groups within this study (Welsh dominant bilingual and Welsh at school bilingual), scores on the English reading tests

(NARA) had a significant negative correlation with the number of phonological errors in both English and Welsh, indicating that higher ability readers made fewer phonological errors in their writing than lower ability readers. However, this finding was not echoed in relation to scores on the Welsh reading test (PGM) and also did not apply to the English dominant group. This may again relate to the notion of the orthographic depth of a language playing a part in children's acquisition of a particular language, however further research into the differences in bilinguals' performance on such tasks, particularly children with a diagnosed literacy difficulty, is needed to gain a fuller picture.

Implications: Exploring the written texts of bilinguals can provide useful information regarding children's literacy profiles in each of their languages.

> *Recommendations for future research*: Whether or not a timed written task may be useful as part of a battery of diagnostic measures for Welsh requires further study. However, the distinguishing features found between bilinguals' written profiles in English and in Welsh demonstrate the potential for such a task as a diagnostic measure, particularly when looking specifically at pupils' adherence to language-specific nuances. Future studies would require collecting examples of written texts from children who have already been diagnosed with a literacy difficulty.

Studies 2 and 3

Studies 2 and 3 explored bilinguals' productive and receptive understanding of two complex structures that each require different levels of phonological awareness skills, namely Welsh mutation and plural morphology. Given the known links between literacy difficulties and phonological awareness skills (see Chapter 2), these structures are of interest given their high phonological load. Additional support for the exploration of these structures comes from the experiences of specialist dyslexia teachers (Davies, 2016) who report words that have 'changing features' (i.e. consonant mutation, or the vowel change in plural morphology) as being particularly problematic for Welsh-

English bilinguals and the results of Study 1, which identified vowel substitution as particularly erroneous for Welsh-English bilinguals in English.

Study 2 explored bilingual children's understanding and application of Welsh mutation in both oral and written contexts. This study consisted of 4 computer based tasks, which included a multiple choice cloze task (Task 1), multiple choice non-word cloze task (Task 2), grammaticality judgement task (Task 3a), sentence correction tasks (Task 3b and 4b) and a sentence repetition task (Task 4a) all based on Welsh mutation with particular focus on the most common form of mutation i.e. Soft Mutation.

Study 3 explored bilingual children's understanding and application of Welsh plural morphology. There were three tasks within this study, the first a singular-plural conversion in written form, which was a practice task to focus children's minds on plurals and to assess their understanding of changing a singular to a plural. The second task was in 2 parts: part a, a real word singular to plural conversion task and part b, a non-word singular to plural conversion task using the same pattern observed in part a. Task 3 was an oral production task where children listened to a singular word and were asked to say the plural version of that word into the microphone.

Results of both studies are discussed below in relation to the research questions under exploration in this thesis.

What is the relationship between tasks of morpho-phonological awareness, in the form of Welsh mutation and plural morphology tasks, and IQ, Age and Memory for digits?

IQ, age and memory for digits were included as background measures. Mutations Task 3a and 3b were the only tasks that significantly correlated with IQ and age, suggesting that children with a higher IQ performed better on these tasks as did older children. Similarly, Plural Task 2a and 2b both significantly correlated with IQ and age suggesting that older children perform better on these tasks as did children with higher IQ.

This might suggest that these types of tasks (grammaticality judgement and sentence correction tasks for mutation and real and non-word singular-plural conversion tasks), may help distinguish between those with high IQ and low IQ, and provide useful developmental progression data, whereas multiple-choice real and non-word cloze tasks, and sentence repetition tasks for mutation and oral plural production tasks do not.

Memory for digits was not significantly correlated with any of the mutation or plural tasks. This could suggest that the task design avoids distinguishing between those with good or poor memory skills, however more research would be required to explore these links due to the limited number of participants involved in the present study.

> *Implications*: Older children and children with higher IQ appear to perform better on some morpho-phonological awareness tasks suggesting that some tasks distinguish between these factors more than others.
>
> *Recommendations*: Future studies need to involve larger sample groups across a wide range of ages and tests of morpho-phonological skills should include measures of IQ.

How do tests of morpho-phonological awareness compare with traditional tests of phonological awareness?

A series of traditional phonological awareness tests, as typically used with children in English, but adapted, where possible, to Welsh, was given to each participant alongside the new mutation and plural tasks. These were included first, because they are part of normed, standardised tests of English, or part of test batteries that are currently used for Welsh, and second, because it was important to know whether the new mutation and plurals tasks provided similar results (i.e. has good criterion validity in terms of phonological

awareness) and could therefore be considered as potential items to be included in future diagnostic tests.

The results revealed that all new tasks (mutation and plural) correlated with at least one of the background Phonological awareness measurements For example, Elision was only significantly correlated with Task 1, Task 3a and 3b whereas the Blending Words task was significantly correlated with all mutation tasks. This was a pleasing result as it supports the validity of the mutation tasks, in particular Task 1, Task 3a and Task 3b, which correlated significantly with all tasks of the CTOPP-2, a standardised measure of phonological awareness. This finding was echoed with the plurals tasks (Study 3) where Plurals Task 2a significantly correlated with all phonological awareness tasks (including tasks on the CTOPP-2), with the exception of Final Sound W. Although not as highly correlated, Plurals Task 2b and 3 also revealed some significant relationships with the tasks of phonological awareness.

> *Implications*: The mutation and plural tasks created for these studies carry some validity in relation to the assessment of children's phonological awareness.

> *Recommendations*: Future studies need to see how children with known literacy difficulties perform on these tasks, and the extent to which performance is correlated across tasks.

Is there a relationship between children's performance on morpho-phonological awareness tasks and their reading abilities?

Previous studies have shown a clear relationship between children's reading abilities and their morpho-phonological awareness (see Chapter 2). In the present set of studies, all mutation tasks were shown to have significant relationships with reading scores across the whole sample, in particular with Welsh reading scores (no significant correlation emerged between Task 2, Task 3a and Task 4b and the NRTE and between 4b and the AWRTE). This

provides further evidence to support the need for testing bilinguals in both of their languages.

Task 1, Task 3b and Task 4a all significantly correlated with children's reading performance across all reading tests. This is a promising result as it shows a clear link between children's performance on mutation tasks and their reading abilities. All other mutation tasks were also significantly correlated with the AWRTW and NRTW which shows a strong relationship between these tasks and children's performance on Welsh reading tests. Again this supports the potential validity of the mutations tasks given that all reading tests used within the study are standardised.

In relation to the plural morphology tasks, there was a clear relationship between children's performance on Welsh plural morphology tasks and their reading abilities, particularly in relation to L1 Welsh children. The written plural task (Task 2a and 2b) was the most significant in that it positively correlated significantly with all reading tasks, suggesting that higher ability readers perform better on this task than lower ability readers. This supports the predictions made that higher ability readers would perform better on this task than lower ability readers. Together, these patterns suggests that the types of mutation and plural tests designed for this study may allow testers to distinguish between those who are good and poor readers, and suggest that children's use of these structures in class may provide useful cues to teachers as to potential literacy difficulties, particularly in contexts where the transparency of the orthography may mask such difficulties.

> *Implications*: Children's performance on reading tasks largely correlate with their performance on tasks exploring morpho-phonological awareness in the form of mutations and plurals, as measured here. These patterns suggest that looking at language specific aspects of the target language, particularly those that require skills that are often considered as critical for literacy (i.e. good phonological awareness skills), might be beneficial in identifying at risk children sooner than is currently possible.

Recommendations: Future studies should consider including language specific phonological awareness tasks when developing new or novel testing tools.

Does the relationship between children's performance on mutation and plural morphology and reading tasks vary depending on the type of bilingual they are?

Following the initial correlational analysis, the sample was split into two groups, Group 1 (Mostly Welsh) and Group 2 (Mostly English), to explore whether any differences in relationships emerged. On the whole raw scores revealed that children in the Mostly Welsh group performed better on the mutations tasks than children in the Mostly English group with the exception of Task 2, where the Mostly English group slightly outperformed the Mostly Welsh group. This was as predicted given that the foundation of the novel tasks was Welsh mutation and therefore it would be expected that L1 Welsh speakers would be at an advantage in relation to exposure to Welsh grammatical patterns (cf. Thomas & Gathercole, 2005; Gathercole, Thomas & Laporte, 2001; Gathercole & Thomas, 2009). Task 2 revealed a surprising result in that children in the Mostly English group slightly performed better than those in the Mostly Welsh group. It is unclear why this was the case; however, there were clear limitations to Task 2 (see below and Chapter 5) and also the difference in group size make drawing solid conclusions very difficult.

The relationship between children's performance on reading tests and their performance on the mutations tasks varied according to language group. The Mostly Welsh group revealed the most significant correlations with all tasks correlating significantly with the AWRTW as well as some also correlating with the AWRTE, NRTW and NRTE. Very few correlations emerged between the Mostly English group and reading scores. Task 1 however consistently correlated with all reading tests across both language groups. This is a promising result as it would suggest Task 1 may be a significant indicator of children's literacy abilities, regardless of their language background.

Results show that the nature of the relationship tends to vary considerably depending on the type of test being administered and also the language background of the children completing the tasks. Predictions for this question were on the whole upheld with regards to higher ability Welsh readers performing better than lower ability Welsh readers on novel mutation tasks. Scores on all mutations tasks were significantly correlated with either the AWRTW or the NRTW and some tasks were significantly correlated with both, although this was only within the Mostly Welsh group. Variation occurred when observing correlations between the novel tasks and the English reading tests within this group with significant correlations only emerging between one or both English reading tests and all novel tasks with the exception of Task 4b. The nature of the relationship between scores on the novel tasks and reading test scores within the Mostly English group was very different. Significant correlations only emerged between Welsh reading tests and scores on Task 1 and on Task 4b. No significant correlations emerged with scores on Task 2 and Task 3a. Significant correlations were revealed between scores on Task 3b, Task 4a, and Task 1 and the NRTE (Task 1 correlated with all reading tests within this group) but no other tests. The nature of the relationship between children's performance on novel mutation tasks and reading scores in both English and Welsh appeared to vary considerably between language groups. This could be a result of language exposure (as referred to in Chapter 2) and so this would need to be a consideration when designing any screening test which measures children's literacy abilities in Welsh.

With regards to the plurals tasks, when the sample was divided according to language background, there was a clear relationship between children's performance on reading tests and their performance on singular to plural conversion tasks, although this varied somewhat in relation to language group. Performance on the tasks remained significantly correlated with the AWRTW within the Mostly Welsh group but only Task 2b was significantly correlated with the AWRTW within the Mostly English group. Plurals Task 2a and 2b were, however, both significantly correlated with the NRTE in the Mostly English group and Task 2a was significantly correlated with both the NRTE and NRTW in both language groups. It was interesting to note, however,

that within Task 2a and 2b particularly there was very little difference between the mean scores of each group, with the Mostly English group performing fractionally better than the Mostly Welsh group on the non-word task (Task 2b).

There was a clear difference when the sample was split according to language background, with fewer significant correlations emerging within the Mostly English group when compared with the Mostly Welsh group. This result could suggest that these tasks appear to distinguish well between L1 learners but not between L2 learners who tend to find these tasks more challenging.

Implications: In general, performance on tasks of morpho-phonological awareness varied according to children's Welsh-English language background, although within the current study, mutations Task 1 (a multiple choice cloze task) had a significant relationship with reading scores regardless of children's language background.

Recommendations: The use of multiple choice cloze tasks, such as the one used within Study 2, could be considered in the design of future literacy screening tests. Detailed information regarding the type of bilingual being assessed is also important given the variance in findings in relation to the two language groups (cf. Gathercole, Thomas & Hughes, 2008).

Can tests of morpho-phonological awareness predict children's performance on reading tests?

A simple regression analysis was run between each of the mutations tasks that had arisen as significantly correlated with reading scores on each of the reading tests. The results showed that none of the individual novel tasks alone could be used as a predictor of children's performance on reading tests, although all tasks significantly accounted for the variance in reading scores (AWRTW) within the Mostly Welsh Group. Task 1 emerged as being the highest

predictor of Welsh reading scores (AWRTW) for L1 Welsh speakers in comparison with the other tasks within the study whereas Task 4a was the strongest predictor of English reading scores (NRTE) amongst L2 Welsh speakers. These are promising findings for the use of these tasks as predictors of reading abilities amongst bilinguals. All mutation tasks, with the exception of Task 4b, also significantly accounted for the variance in the AWRTE scores within the Mostly Welsh group implying that these tests can potentially predict children's English and Welsh reading scores (as measured by the AWRT) in bilingual children whose dominant language is Welsh. This is a significant finding which is worthy of further exploration.

A multiple regression analysis further explored the mutation tasks which had arisen as having a statistically significant relationship with the reading tests. These tasks were included within a single model to explore whether any of the tasks could be a predictor of reading ability. Within the Mostly Welsh group the first model included all tasks. Task 1 was found to make a significant unique contribution to the prediction of scores on the AWRTW, suggesting that Task 1 could be a potential predictor of Welsh reading abilities. Cloze tasks have previously been found to be closely linked to reading ability (e.g. Trace, 2020) and so this shows promising results regarding the reliability of Task 1 as a predictor of reading abilities. Task 4a made a significant unique contribution to the prediction of reading scores on both the NRTW and NRTE within the Mostly Welsh Group. None of the mutation tasks made a significant contribution to the prediction of reading test scores within any of the Mostly English models.

In relation to the plurals tasks, according to the simple regression analysis results it would appear that there is some scope for plurals tasks of this nature to be used as predictors of L1 Welsh children's reading ability, with Task 2a significantly accounting for a level of variance (albeit low) in all reading scores within Group 1. Task 2b also significantly accounted for a small level of the variance in reading scores, namely the AWRTW and AWRTE in Group 1. The picture varies considerably in Group 2, however, with Plurals Task 2a significantly accounting for a higher level of variance, although this only

applied to the NRTE task. Task 2b also significantly accounted for the level of variance in NRTE scores. This suggests therefore that Plurals Tasks 2a and 2b show promising results with regards to their prediction of English reading scores for L2 Welsh-speaking children and tasks of this nature may therefore be useful in the design of future Welsh-English bilingual literacy screening tools. Further exploration of using plural tasks of this nature with larger sample sizes is needed to further solidify the role of Welsh plural morphology in identifying children's literacy difficulties.

Implications: Tests of morpho-phonological awareness can provide some indication of children's reading abilities, although this is dependent on the nature of the test.

Recommendations: Further exploration into the role of morpho-phonological tasks in predicting reading ability is needed with a larger sample; however, multiple choice cloze tasks and sentence repetition tasks may be useful starting points in exploring this notion further.

Limitations of the studies

Whilst the implications and recommendations presented above arise directly from the results as obtained in this study, it is important to highlight some issues that may have influenced participants' behaviour. These issues should be taken into account in any further studies in this field.

Within Study 1 the categorisation of errors was a challenge, in particular determining whether a child had made a phonological error or a spelling error. This was unavoidable, however, without being able to question participants directly with regards to their writing.

Within the computer based task 'Y Goriad', due to there being no back button or option for children to alter their responses themselves once they had moved to the next page, any answers that the children wanted to change had to be communicated verbally to the researcher and so depended on children

communicating this. Future tests which use an interactive format, such as the tests used here, should consider incorporating an option that enables participants to change their answer if they wish, without needing to communicate this to the researcher.

Another limitation was related to the multiple choice aspect of the tasks which potentially prevents the researcher gaining an accurate view of the child's abilities due to children potentially guessing answers.

Task 2 emerged as particularly challenging for the children and so it was difficult to gauge the accuracy of these scores as again this could be due to children guessing their responses. The use of non-words within a measure of morpho-phonological awareness in Welsh therefore may not be suitable although this may be worthy of exploration with older children.

Task 4 revealed some limitations in relation to sentence length in the use of sentence repetition tasks. There appeared to be an effect of sentence length on children's performance on this task and so this would need to be taken into consideration in the design of any future tests of this nature.

In Study 3, limitations arose in relation to Plurals task 2a and 2b. At times it was challenging to determine if an error made was a written spelling error or a plural error. The use of oral tasks alongside a written tasks is therefore recommended to acquire a more accurate view of a child's plural abilities. In relation to the oral production task (Task 3), the use of a microphone could be off-putting for children and so the use of technology which enabled the children's responses to be recorded without the need to hold a microphone may be beneficial.

A limitation that arose from studies 2 and 3 was in relation to the categorisation of language background. The first language of the children was identified through parents' identification of what they considered the child's first language to be; it did not provide a full reflection of the level of exposure each child had to each of the languages, which can have an effect on children's performance on tasks (see e.g. Boerma and Blom 2017). Future studies should therefore consider gaining a more in-depth representation of the

bilingual children's exposure to each of their languages before examining their performance on tasks.

Another limitation within studies 2 and 3 was in relation to the group sizes, the Mostly Welsh group was considerably larger than the Mostly English group and so it was difficult to make solid comparisons and draw secure conclusions on the differences between the groups. Further research using more balanced group sizes would therefore be beneficial.

Future recommendations

From a teacher's perspective these studies present some useful findings, in particular in relation to the relationships between mutation and reading scores and Welsh plural morphology and reading scores. There is scope here to develop tests that include the specific types of tasks that seemed most useful according to the results observed here (i.e cloze tasks and singular-plural conversion tasks). Given that children appeared to find the computer based format of the tasks enjoyable it is recommended that more interactive tasks such as these be developed as part of diagnostic tools. For teachers of Welsh-speaking children, the findings of the current thesis are useful in the sense that, if a child appears to be struggling with mutation in their written work, the availability of the types of mutation tasks that seemed best able to distinguish between good and poor readers could be used by the teachers, thereby enabling them to assess the child's abilities in a less formal way, and would potentially provide insight into the child's literacy difficulties. As found in Study 1, written texts provided by children in both languages can yield some useful information for educators, so the inclusion of a bilingual written element within diagnostic testing could also be very beneficial in identifying bilingual children's literacy difficulties.

As highlighted by Butvilofsky *et al.* (2020) it is important to use a variety of assessment methods in order to develop a holistic view of a child's abilities. The inclusion of reading, writing and oral tasks would be beneficial as well as potentially exploring the speed and length of writing and observing children's

approach to literacy. For example, exploring systematically the pace, effort, number and nature of stop-starts and reading repairs etc. performed whilst reading extended texts in combination with the pace, total number of words, mean length of sentential expressions, and the extent of self-correction performed whilst writing. This may help educators, in particular, to identify early issues among their pupils, particularly among those who are learning a language with a transparent orthography, which may mask underlying issues.

Further research using a larger and more balanced sample to explore the elements of mutation and pluralisation which arose as significant within studies 2 and 3, as well as other mutation triggers would provide useful information and build on the findings of this thesis.

Discussions with Welsh teachers regarding their observations when working with children with literacy difficulties within the Foundation Phase, with particular focus on the types of difficulties Welsh-speaking children present in the classroom setting, could also yield useful information for future research.

Conclusion

The findings of this thesis show promising results in relation to the potential benefits of measuring pupils' knowledge of mutation and plural morphology as a means of highlighting to educators those children who may exhibit literacy difficulties in Welsh. The variance in findings when comparing language groups further supports the notion that different types of bilinguals perform differently on language tasks, and is something that should be considered in any testing/support given to different speakers and in our expectations of their ultimate achievements. It also highlights the importance of testing in both languages. Ultimately, children, regardless of language background, should have equal opportunities to reach their potential within education, and developing tests that provide a holistic view of a child's abilities, in both their languages, is one essential step in enabling children to do that.

REFERENCES

Abudarham, S. (1997). The tenability of comparing the receptive lexical proficiency of dual language children with standardised monoglot norms. *Educational Studies*, 23(1), 127-14.

Alexander-Passe, N. (2006). How dyslexic teenagers cope: an investigation of self-esteem, coping and depression. *Dyslexia,* 12, 256-275.

Armon-Lotem, S. de Jong, J., and Meir, N. (2015) *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment.* Bristol: Multilingual Matters.

Aro, M., and Wimmer, H. (2003). Learning to read: English in comparison to six more regular orthographies. *Applied Psycholinguistics, 24*, 621-635.

Baker, C. (2002). *Foundations of Bilingual Education and Bilingualism*, Clevedon: Multilingual Matters.

Ball, M. J., and Müller, N. (1992). *Mutation in Welsh*. Routledge.

Barca, L., Burani, C., Di Filippo, G., and Zoccolotti, P. (2006). Italian Developmental Dyslexic and Proficient Readers: Where Are the Differences? *Brain and Language*, 98(3), 347-351.

Bedore, L. M., & Peña, E. D. (2008). Assessment of bilingual children for identification of language impairment: Current findings and implications for practice. International Journal of Bilingual Education and Bilingualism, 11, 1–29.

Bell, J.F. and Daniels, S. (1990). Are summer-born children disadvantaged? The birthdate effect in education. *Oxford Review of Education, 16,* 67-80.

Bellin, W. (1988). The development of pronunciation. In M. J. Ball (Ed.), *The use of Welsh: A contribution to sociolinguistics.* Clevedon: Multilingual Matters.

Bharick, H.P., Hall, L.K., Goggin, J.P., Bharick, L.E. and Berger, S.A. (1994). Fifty years of language maintenance and language dominance in bilingual Hispanic immigrants. *Journal of Experimental Psychology, 123*(3), 264-83.

Bialystok, E. (2006). Bilingualism at School: Effect on the acquisition of literacy. In *Childhood Bilingualism: Research on infancy through school age*. P. McCardle, and E. Hoff. Eds. 107-24. Clevedon: Multilingual Matters.

Bialystok, E., and Luk, G. (2012). Receptive vocabulary differences in monolingual and bilingual adults. *Bilingualism: Language and Cognition, 15*(2), 397-401.

Binks, H., and Thomas, E. (2019). Long-term outcomes for bilinguals in minority language contexts: Welsh-English teenagers' performance on measures of Grammatical Gender and Plural Morphology in Welsh. *Applied Psycholinguistics*, 1-31.

Boerma, T. and Blom, E. (2017). Assessment of bilingual children: What if testing both languages is not possible? *Journal of Communication Disorders,* 66, 65-76.

Burden, R. (2005). *Dyslexia and Self-Concept: Seeking a Dyslexic Identity.* London: Whurr Publishers Ltd.

Butler, Y.G. 2015b. The use of computer games as foreign language learning tasks for digital natives. *System* 54: 91–102.

Butvilofsky, S. A., Escamilla, K., Gumina, D. and Diaz, E. S. (2020). Beyond Monolingual Reading Assessments for Emerging Bilingual Learners: Expanding the Understanding of Bilteracy Assessment Through Writing. *Reading Research Quarterly*, 0 (0), 1-18.

Caesar, L.G. and Kohler, P.D., (2007). The state of school-based bilingual assessment: actual practice versus recommended guidelines. *Language, Speech and Hearing Services in Schools, 38,* 190-200.

Caravolas, M., (2011). ELDEL Project. https://www.eldel-mabel.net/

Caravolas, M., Lervåg, A., Defior, S., Seidlová Málková, G., Hulme, C. (2013). Different patterns, but equivalent predictors, of growth in reading in consistent and inconsistent orthographies. *Psychological Science*, 24, 1398–1407.

Caravolas, M., Volin, J., and Hulme, C. (2005). Phoneme awareness is a key component of alphabetic literacy skills in consistent and inconsistent orthographies: Evidence from Czech and English children. *Journal of Experimental Child Psychology, 92,* 107-139.

Carroll, S. E. (2015). Exposure and input in bilingual development. *Bilingualism: Language and Cognition*, 20(1), 3–16.

Casalini, c., Brizzolara, D., Chilosi, A., Cipriani, P., Marcolini, S. Pecini, C., Roncoli, S. and Burani, C. (2007). Nonword repetition in children with Specific Language Impairment; A deficit in phonological working memory or in long-term verbal knowledge? *Cortex,* 43, 769-776.

Castles, A. and Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition,* 47, 149-180.

Chiat, S. (2015). Non-word repetition. In Armon-Lotem, S., Jong, J.D. and Meir, N. *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment.* Bristol: Multilingual Matters.

Chiat, S. and Roy, P. (2007). The preschool repetition test: An evaluation of performance in typically developing and clinically referred children. *Journal of Speech, Language, and Hearing Research,* 50(2), 429–443.

Chiat, S., Armon-Lotem, S., Marinis, T., Polisenska, K., Roy, P. and Seeff-Gabriel, B. (2013). Assessment of Language Abilities in Sequential Bilingual Children: The Potential of Sentence Imitation Tasks. In V.C.M. Gathercole (ed.) *Issues in the Assessment of Bilinguals*. Bristol: Multilingual Matters.

Chondrogianni, V., and John, N. (2018). Tense and plural formation in Welsh-English bilingual children with and without language impairment.

International journal of language & Communication Disorders, 53(3), 495-514.

Coady, J.A. and Evans, J.L. (2008). Uses and Interpretations of non-word repetition tasks in children with and without language impairments (SLI). *International Journal of Language and Communication Disorders, 43*, 1-40.

Cobo-Lewis, A. B., Pearson, B.Z., Eilers, R.E., and Umbel, V. C. (2002a). Effects of bilingualism and bilingual education on oral and written Spanish skills: A multifactor study of standardized test outcomes. In *Language and literacy in bilingual children,* ed. D.K. Oller and R.E. Eilers, 98-117. Clevedon: Multilingual Matters.

Cobo-Lewis, A. B., Pearson, B.Z., Eilers, R.E., and Umbel, V. C. (2002b). Effects of bilingualism and bilingual education on oral and written English skills: A multifactor study of standardized test outcomes. In *Language and literacy in bilingual children,* ed. D.K. Oller and R.E. Eilers, 64-97. Clevedon: Multilingual Matters.

Coltheart, M., Davelaar, E., Jonasson, T. and Besner, D. (1977). Access to the internal lexicon. In S. Dornic (ed.) *Attention and performance VI*. New York: Academic Press. As cited in Perez *et al.* (2013). Assessment of Bilinguals' Performance in Lexical Tasks Using Reaction Times. In V.C.M. Gathercole (ed.) *Issues in the Assessment of Bilinguals*. Bristol: Multilingual Matters.

Coltheart, M., Rastle, K., Perry, C., Langdon, R. and Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204-256.

Conti-Ramsden, G., Botting, N. and Faragher, B. (2001). Psycholinguistic markers for specific language impairment. *Journal of Child Psychology and Psychiatry, 42*(6), 741–748.

Core, C., Hoff, E., Rumiche, R, and Senor, M., (2013). Total and conceptual vocabulary in Spanish-English bilinguals from 22 to 30 months:

implications for assessment. *Journal of Speech, Language and Hearing Research, 56,* 1637-1649.

Cummins, J. (1979). Linguistic Interdependence and the educational development of bilingual children. *Review of Educational Research, 49,* 225-251.

Davies, A.E. (2016). *Delivering Welsh medium provision for pupils with specific literacy difficulties.* Unpublished MA Dissertation, Bangor University.

De Houwer, A. (2014). The absolute frequency of maternal input to bilingual and monolingual children: A first comparison. In T. Gruter & J. Paradis (Eds.), *Input and experience in bilingual development* (pp. 37–58). Amsterdam: John Benjamins.

De Lamo White, C., & Jin, L. (2011). Evaluation of speech and language assessment approaches with bilingual children. International Journal of Language and Communication Disorders. 46, 613–627.

Dunn, L.M., Dunn, L.M., Whetton, C. and Burley, J. (1997). *The British Picture Vocabulary Scale: Second Edition.* NFER-NELSON.

Duursma, E., Romero-Contreras, S., Szuber, A., Proctor, P., Snow, C., August, D. (2007). The role of home literacy and language environment on bilinguals' English and Spanish vocabulary development. *Applied Psycholinguistics, 28,* 171-90.

Dworin, J. (2003). Insights into biliteracy development: Toward a bidirectional theory of bilingual pedagogy. *Journal of Hispanic Higher Education, 2*(2), 171-186.

Ellis, N. C. & Hooper, M. (2001). Why learning to read is easier in Welsh than in English: Orthographic transparency effects evinced with frequency matched tests. *Applied Psycholinguistics, 22*, 571-599.

Engel de Abreu, P. (2011). Working memory in multilingual children: Is there a bilingual effect? *Memory, 19*(5), 529-537.

Ezeizabarrena, M.J., Barnes, J., Garcia, I., Barrena, A. and Almgren, M. (2013). Using Parental Report assessment for Bilingual Preschoolers: The Basque Experience. In V.C.M. Gathercole (ed.) *Solutions for the Assessment of Bilinguals*. Bristol: Multilingual Matters.

Fawcett, A., & Nicolson, R. (2004). Dyslexia Screening Test - Junior. Oxford: Pearson.

Fenson, L., Dale, P. S., Reznick, J.S, Thal, D., Bates, E., Hartung, J.P., Pethick, S. and Reilly, J. S. (1993). *The MacArthur Communicative Development Inventories.* San Diego, CA: Singular.

Fenson, L., Dale, P.S., Reznick, J.S., Bates, E., Thal, D., and Pethick, S.J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, 59(5) Serial No. 242.

Fiebach, C.J., Friederici, A.D., Muller, K. and von Cramon, D.Y. (2002). fMRI evidence for dual routes to the mental lexicon in visual word recognition. *Journal of Cognitive Neuroscience, 14,* 11-23.

Forbes, S. (1999). *All Wales Reading Test.* https://www.awrt.online/

Forster, K.I. (1976). Accessing the mental lexicon. In E.C.T. Walker and R.J. Wales (eds) *New Approaches to Language Mechanisms.* Amsterdam: North-Holland. As cited in Perez *et al.* (2013). Assessment of Bilinguals' Performance in Lexical Tasks Using Reaction Times. In V.C.M. Gathercole (ed.) *Issues in the Assessment of Bilinguals*. Bristol: Multilingual Matters.

Furnes, B. and Samuelsson, S. (2009). Preschool cognitive and language skills predicting kindergarten and grade 1 reading and spelling: A cross-linguistic comparison. *Journal of Research in Reading, 32*(3). 275-292.

Furnes, B., and Samuelsson, S. (2010). Predicting reading and spelling difficulties in transparent and opaque orthographies: A comparison between Scandinavian and US/Australian children. *Dyslexia, 16*(2), 119-142.

Gathercole, S.E. (2006). Nonword repetition and word learning. The nature of the relationship. *Applied Psycholinguistics, 27,* 513-543.

Gathercole, S. E., and Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: a longitudinal study. *Journal of Memory and Language*, *28*, 200-213.

Gathercole, V.C.M. (2002a). Command of the mass/count distinction in bilingual and monolingual children: An English morphosyntactic distinction. In *Language and literacy in bilingual children,* Oller, D.K. and Eilers, R.E. eds. 175-206. Clevedon: Multilingual Matters.

Gathercole, V.C.M. (2002b). Grammatical gender in bilingual and monolingual children: A Spanish morphosyntactic distinction. In *Language and literacy in bilingual children,* Oller, D.K. and Eilers, R.E. eds. 207-19. Clevedon: Multilingual Matters.

Gathercole, V.C.M. (2002c). Monolingual and bilingual acquisition: Learning different treatments of that-trace phenomena in English and Spanish. In *Language and literacy in bilingual children.* Oller, D.K. and Eilers, R.E. eds. 220-54. Clevedon: Multilingual Matters.

Gathercole, V.C.M. (2013). Assessment of Bilinguals: Innovative Solutions. In V.C.M. Gathercole (ed.) *Solutions for the Assessment of Bilinguals*. Bristol: Multilingual Matters.

Gathercole, V.C.M. and Thomas, E.M. (2007). *Prawf Geirfa Cymraeg, Fersiwn 7-11* (Welsh Vocabulary Test, Version 7-11), 2007. www.pgc-bangor.ac.uk

Gathercole, V.C.M., and Thomas, E.M. (2009). Bilingual first language development: Dominant language takeover, threatened minority take-up. *Bilingualism: Language and Cognition,* 12, no. 1: 213-37.

Gathercole, V. C. M., Kennedy, I. A., and Thomas, E.M. (2016). Socioeconomic level and bilinguals' performance on language and cognitive measures. Bilingualism: Language and Cognition, 1-22.

Gathercole, V. C. M., Laporte, N. I., and Thomas, E. M. (2005). Differentiation, Carry-over, and the Distributed Characteristic in Bilinguals: Structural 'Mixing' of the Two Languages? In J Cohen K T McAlister K

Rolstad & J MacSwan (eds) *Proceedings of 4th International Symposium on Bilingualism, Cascadilla press.*

Gathercole, V.C.M., Thomas, E.M. and Hughes, E.K. (2008). Designing a normed receptive vocabulary test for bilingual populations: A model from Welsh. *International Journal of Bilingual Education and Bilingualism, 11*(6), 678-720.

Gathercole, V. C. M., Thomas, E. M, and Laporte, N. I. (2001). The acquisition of grammatical gender in Welsh. *Journal of Celtic Language Learning, 6,* 53-87.

Gathercole, V.C.M., Thomas, E.M., Thomas, Roberts E.J., Hughes. C.O. and Hughes, E.K. (2013). Why Assessment Needs to Take Exposure into Account: Vocabulary and Grammatical Abilities in Bilingual Children. In V.C.M. Gathercole (ed.) *Issues in the Assessment of Bilinguals*. Bristol: Multilingual Matters.

Gathercole, V.M. Eds. (2007). *Language Transmission in Bilingual Families in Wales.* Cardiff: Welsh Language Board.

Geisinger, K.F. (2003). Testing and assessment in cross-cultural psychology. In J.R. Graham and J.A. Naglieri (eds) *Handbook of Psychology. Vol 10: Assessment Psychology.* Hoboken, NJ: John Wiley & Sons.

Genesee, F. (2001). Bilingual first language acquisition: Exploring the limits of the language faculty. *Annual Review of Applied Linguistics, 21*, 153-68.

Girbau, D. and Schwartz, R.G. (2007). Non-word repetition in Spanish-speaking children with Specific Language Impairment (SLI). *International Journal of Language and Communication Disorders, 42*, 59-75.

Graf Estes, K., Evans, J.L and Else-Quest, N.M. (2007). Differences in the nonword repetition performance of children with and without Specific Language Impairment. A meta-analysis. *Journal of Speech, Language and Hearing Research, 50*, 177-195.

Grosjean, F. (1998). Studying Bilinguals: Methodological and Conceptual Issues. *Bilingualism: Language and Cognition, 1,* 131-149.

Grosjean, F. (2008). *Studying Bilinguals.* Oxford: Oxford University Press.

Gross, M., Buac, M., and Kaushanskaya, M. (2014). Conceptual scoring of receptive and expressive vocabulary measures in simultaneous and sequential bilingual children. *American Journal of Speech-Language Pathology*, *23,* 574–586.

Hakuta, K., and Diaz, R.M. (1985). The relationship between degree of bilingualism and cognitive ability: A critical discussion and some new longitudinal data. *Children's Language,* 319-344.

Hammer, C.S., Hoff, E., Uchikoshi, Y., Gillanders, C., Castro, D., and Sandilos, L.E. (2014). The language and literacy development of young dual language learners: A Critical review. *Early Childhood Research Quarterly, 29*(4), 715-733.

Hammer, C.S., Miccio, A.W., and Rodriguez, B. (2004). *Bilingual Language Acquisition and the Child Socialization Process.* Goldstein, B.A. ed. Baltimore, MD: Brookes Publishing Company.

Hanley, J. R., Masterson, J., Spencer, Ll. H., and Evans, D. (2004). How long do the advantages of learning to read a transparent orthography last? An investigation of the reading skills and reading impairment of Welsh children at 10 years of age. *Quarterly Journal of Experimental Psychology: Part A: Human Experimental Psychology, 57*(8), 1393-410.

Hart, B., and Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Brookes.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development, 74*(5), 1368-1378.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review, 26*, 55-88.

Howard, E. R., Green, J. D. and Arteagoitia, I. (2012). Can yu rid guat ay rot? A Developmental Investigation of Cross-Linguistic Spelling Errors Among Spanish-English Bilingual Students. *Bilingual Research Journal*, *35*(2), 164-178.

Izura, C., Perez, M.A., Agallou, E., Wright, V.C., Marin, J., Stadthagen-Gonzalez, H. and Ellis, A.W. (2011). Age/order of acquisition effects and the cumulative learning of foreign words: A word training study. *Journal of Memory and Language, 64*, 32-58.

Jackson-Maldonado, D.Thal, D., Fenson, L., Marchmann, V., Newton, T. and Conboy, B. (2003). *MacArthur Inventarios del Desarrollo de Habilidades Comunicativas*. User's Guide and Technical Manual. Baltimore: Brookes Publishing.

James, 1975 as cited in Perez, M.A., Izura, C., Stadthagen-Gonzalez, H. and Marin, J. (2013) Assessment of Bilinguals' Performance in Lexical Tasks Using Reaction Times. In V.C.M. Gathercole (ed.) *Issues in the Assessment of Bilinguals*. Bristol: Multilingual Matters.

Johnson and Barry, (2006) as cited in Perez, M.A., Izura, C., Stadthagen-Gonzalez, H. and Marin, J. (2013) Assessment of Bilinguals' Performance in Lexical Tasks Using Reaction Times. In V.C.M. Gathercole (ed.) *Issues in the Assessment of Bilinguals*. Bristol: Multilingual Matters.

Jones, B.M. (1993). *Ar lafar ac ar bapur: Cyflwyniad i'r berthynas rhwng yr iaith lafar a'r iaith ysgrifennedig*. Y Ganolfan Astudiaethau Addysg: Aberystwyth.

Jones, B., and Lewis, W. G. (2014). Language Arrangements within Bilingual Education*. In E.M. Thomas and Ineke Mennen (eds) *Advances in the Study of Bilingualism.* Bristol: Multilingual Matters.

Juhasz, (2005) as cited in Perez, M.A., Izura, C., Stadthagen-Gonzalez, H. and Marin, J. (2013) Assessment of Bilinguals' Performance in Lexical Tasks Using Reaction Times. In V.C.M. Gathercole (ed.) *Issues in the Assessment of Bilinguals*. Bristol: Multilingual Matters.

Junker, D.A. and Stockman, I.J., 2002, Expressive vocabulary of German-English bilingual toddlers. *American journal of speech-language pathology, 11*, 381-394.

Katz, L., and Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. In R. Frost and L. Katz (Eds.), *Orthography, phonology, morphology, and meaning: Advances in psychology,* 67-84. Oxford, UK: North-Holland.

Klassert, A., Gagarina, N., & Kauschke, C. (2014). Object and action naming in Russian- and German-speaking monolingual and bilingual children. *Bilingualism: Language and Cognition*, 17 (01), 73–88.

Kovelman, I., Baker, S.A., and Petitto, L.A. (2008). Age of first bilingual exposure as a new window into bilingual reading development. *Bilingualism: Language and Cognition, 11*(2), 203-23.

Leonard, L. (1998). *Children with Specific Language Impairment*. Cambridge, MA: MIT Press.

Lewis, D.G. (1993). *Y Treigladur*. Gomer.

Lombardi, L. and Potter, M.C. (1992). The regeneration of syntax in short term memory*. Journal of Memory and Language, 31* (6), 713–733.

Lopez-Ornat, S., Gallego, C., Gallo, P. Karousou, A., Mariscal, S. and Martinez, M. (2005). *MacArthur. Inventario de Desarrollo Comunicativo MacArthur.* Madrid: Tesa.

Maratsos, M. (2000). More overregularizations after all. *Journal of Child Language, 27*(1), 183-212.

Marchman, V. A., and Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language, 21*(2), 339-66.

Marinis, T. and Armon-Lotem, S. (2015). Sentence Repetition. In Armon-Lotem, S.,Jong, J.D. and Meir, N *Assessing Multilingual Children:*

Disentangling Bilingualism from Language Impairment. Bristol: Multilingual Matters.

Martin, R.P., Foels, P., Clanton, G. and Moon, K. (2004). Season of birth is related to child retention rates, achievement, and rate of diagnosis of specific LD. *Journal of Learning Disabilities*, *37,* 307–317.

Mattes, L.J. (1995). *Bilingual Vocabulary Assessment Measure*. Oceanside, CA: Academic Communication Associates.

McCarthy, D. (1972). *McCarthy Scales of Children's Abilities.* San Antonio: Harcourt Assessment.

McGregor, K.K. (2009). Semantics in child language disorders. In R. Schwartz (ed.), *Handbook of Child Language Disorders,* New York, NY: Psychology press.

Messer, M.H., Leseman, P.P.M., Boom, J. and Mayo, A.Y. (2010). Phonotactic probability effect in nonword recall and its relationship with vocabulary in monolingual and bilingual preschoolers. *Journal of Experimental Child Psychology, 105*(4), 306–323.

Meyer, D.E., Osman, A.M., Irwin, D.E. and Yantis, S. (1988) Modern mental chronometry. *Biological Psychology, 26*, 3-67.

Montanari, E. G., Abel, R., Graßer, B., Tschudinovski, L. (2018). Do bilinguals create two different sets of vocabulary for two domains? *Linguistic Approaches to Bilingualism, 9*, 1–21.

Muñoz-Sandoval, A.F., Cummins, J., Alvarado, C.G.,   and Ruef, M.L. (2005). *Bilingual verbal ability tests. Normative update*. Boston, MA: Houghton Mifflin Harcourt.

Neale, M. D. (1997). *Neale Analysis of Reading Ability – Revised.* London: GL Assessment.

O'Toole, C. (2013). Using parent report to assess bilingual vocabulary acquisition: A model from Irish. In V.C.M. Gathercole (ed.) *Solutions for the Assessment of Bilinguals*. Bristol: Multilingual Matters.

Oller, D.K. (2005). The Distributed Characteristic in Bilingual Learning. In *Proceedings of the 4th International Symposium,* Cohen, J., McAlister, K.T., Rolstad, K., and Mac Swan, J. eds. 1744-50. Somerville, MA: Cascadilla Press.

Oller, D.K., and Eilers, R.E. (2002). *Language and literacy in bilingual children.* Clevedon: Multilingual Matters.

Oller, D.K., Pearson, B.Z., and Cobo-Lewis, A.B. (2007). Profile effects in early bilingual language and literacy. *Applied Psycholinguistics, 28*, 191-230.

Oshima, T. C., and Domaleski, C. S. (2006). Academic Performance Gap between Summer-birthday and Fall-birthday Children in Grades K - 8. *Journal of Educational Research, 99*(4), 212–217.

Paradis, J., and Genesee, F. (1996). Syntactic Acquisition in Bilingual Children: Autonomous or Interdependent? *Studies in Second Language Acquisition, 18*(1), 1-25.

Paradis, M. and Libben, G. (1987). *The Assessment of Bilingual Aphasia.* Hillside, NJ: Erlbaum.

Paulesu, E., Démonet, J.-F., Fazio, F., McCrory, E., Chanoine, V., Brunswick, N., Cappa, S.F., Cossu, G., Habib, M., Frith, C.D., and Frith, U. (2001). Dyslexia: Cultural diversity and biological unity. *Science, 291*, 2165–2167.

Payne, G. (1998). *Profion Glannau Menai: Asesiadau Safonedig Mewn Darllen a Sillafu Cymraeg*. Gwynedd: Cyd-bwyllgor Anghenion Addysgol Arbennig, Cyngor Sir Ynys Môn a Chyngor Gwynedd.

Pearson, B.Z. (1998). Assessing lexical development in bilingual babies and toddlers. *The International Journal of Bilingualism, 2*(3), 347-72.

Pearson, B.Z., Fernandez, S.C., and Oller, D.K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning, 43,* 93-120.

Pearson, B.Z., Fernandez, S.C., and Oller, D.K. (1995). Cross-language synonyms in the lexicons of bilingual infants: one language or two? *Journal of Child Language, 22*, 345-68.

Peña, E.D., Bedore, L.M. and Kester, E.S. (2016). Assessment of language impairment in bilingual children using semantic tasks: two languages classify better than one. *International Journal of Language and Communication Disorders, 51*(2), 192-202.

Peña, E.D., Gutiérrez-Clellen, V. F., Iglesias, A., Goldstein, B. and Bedore, L. M., (2014*). Bilingual English Spanish Assessment* (San Rafael, CA: A-R Clinical Publ.).

Perez, M.A., Izura, C., Stadthagen-Gonzalez, H. and Marin, J. (2013). Assessment of Bilinguals' Performance in Lexical Tasks Using Reaction Times. In V.C.M. Gathercole (ed.) *Issues in the Assessment of Bilinguals*. Bristol: Multilingual Matters.

Perez-Tattam, R., Gathercole, V.C.M., Yavas, F. and Stadthagen-Gonzalez, H. (2013). Measuring Grammatical Knowledge and Abilities in Bilinguals: Implications for Assessment and Testing. In V.C.M. Gathercole (ed.) *Issues in the Assessment of Bilinguals*. Bristol: Multilingual Matters.

Place, S. and Hoff, E. (2011). Properties of dual language exposure that influence two-year olds' bilingual proficiency. *Child Development, 82*, 1834-1849.

Polišenská, K. (2011). The influence of linguistic structure on memory span: Repetition tasks as a measure of language ability. Unpublished doctoral thesis, City University London.

Posner, M.I. (1978). *Chronometric Explorations of Mind.* Hillside, NJ: Lawrence Erlbaum.

Potter, M.C. and Lombardi, L. (1990). Regeneration in the short-term recall of sentences*. Journal of Memory and Language, 29*, 633–654.

Potter, M.C., and Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *Journal of Memory and Language, 38*(3), 265-282.

Ramus, F. (2004). Neurobiology of dyslexia: A reinterpretation of the data. *Trends in Neurosciences, 27,* 720-726.

Raven, J. (1998). *Standard progressive Matrices-Plus Version*. London: Pearson Education Inc.

Raven, J., Raven, J.C., and Court, J.H. (2004). *Manual for Raven's progressive matrices and vocabulary scales.* San Antonio, TX: Pearson Assessment.

RCSLT (2006). *Communicating Quality 3*. London: Royal College of Speech and Language Therapists.

Redmond, S.M., Thompson, H.L. and Goldstein, S. (2011). Psycholinguistic profiling differentiates specific language impairment from typical development and from attention-deficit/hyperactivity disorder. *Journal of Speech, Language, and Hearing Research*, *54*, 99– 117.

Redmond, S.M. (2005). Differentiating SLI from ADHD using children's sentence recall and production of past tense morphology. *Clinical Linguistics & Phonetics, 19*, 109-127.

Reinhardt, J. (2018). Gameful Second and Foreign Language Teaching and Learning: Theory, Research, and Practice. Springer, Cham.

Rhys, M., and Thomas, E. M. (2013). Bilingual Welsh-English children's acquisition of vocabulary and reading: implications for bilingual education. *International Journal of Bilingual Education and Bilingualism, 16*(6), 633-656.

Richardson, (1976) as cited in Perez, M.A., Izura, C., Stadthagen-Gonzalez, H. and Marin, J. (2013) Assessment of Bilinguals' Performance in Lexical Tasks Using Reaction Times. In V.C.M. Gathercole (ed.) *Issues in the Assessment of Bilinguals*. Bristol: Multilingual Matters.

Riches, N. G., Loucas, T., Baird, G., Charman, T., and Simonoff, E. (2010). Sentence repetition in adolescents with specific language impairments and autism: an investigation of complex syntax. *International Journal of Language & Communication Disorders, 45*(1), 47–60.

Riddick, B., Sterling, C., Farmer, M. & Morgan, S. (1999). Self-esteem and anxiety in education histories of adult dyslexic students. *Dyslexia, 5*, 227-248.

Robert, Elen. 2009. Accommodating "new" speakers? An attitudinal investigation of L2 speakers of Welsh in south-east Wales. *International Journal of the Sociology of Language* 195: 93–115

Rubenstein *et al*., (1971) as cited in Perez, M.A., Izura, C., Stadthagen-Gonzalez, H. and Marin, J. (2013) Assessment of Bilinguals' Performance in Lexical Tasks Using Reaction Times. In V.C.M. Gathercole (ed.) *Issues in the Assessment of Bilinguals*. Bristol: Multilingual Matters.

Sahlén B., ReuterskiöldWagner C., Nettelbladt U., Radeborg K. (1999). Non-word repetition in children with language impairment — pitfalls and possibilities. *International Journal of Language and Communication Disorders, 34*, 337–352

Seeff-Gabriel, B., Chiat, S., and Dodd, B. (2010). Sentence imitation as a tool in identifying expressive morphosyntactic difficulties in children with severe speech difficulties. *International Journal of Language and Communication Disorders, 45,* 691–702.

Seeff-Gabriel, B., Chiat, S., and Roy, P. (2008). *The Early Repetition Battery.* London: Pearson Assessment.

Selleck, Charlotte. 2018. We're Not Fully Welsh: Hierarchies of Belonging and 'New' Speakers of Welsh. In *New Speakers of Minority Languages: Linguistic Ideologies and Practices*. Edited by Cassie Smith-Christmas, Noel P. Ó. Murchadha, Michael Hornsby and Máiréad Moriaty. London: Palgrave Macmillan, pp. 45–65.

Semel, E., Wiig, E. and Secord, W. (1994). *Clinical Evaluation of Language Fundamentals – Revised.* San Antonio, TX: The Psychological Corporation.

Sharp, K. M., and Gathercole, V. C. M. (2013). Can a novel word repetition task be a language-neutral assessment tool? Evidence from Welsh-English bilingual children. *Child Language Teaching and Therapy, 29*(1), 77-89.

Silverman, R. D., Proctor, C. P., Harring, J. R., Hartranft, A. M. Doyle, B. and Zelinke, S. B. (2015). Language skills and reading comprehension in English monolingual and Spanish-English bilingual children in grades 2-5. *Read Writ, 28*, 1381-1405.

Simonsen, H. G., & Haman, E. (2017). LITMUS-CLT – A new way to assess bilingual lexicons. *Clinical Linguistics and Phonetics*, *31*(11–12), 811–817.

Snowling, M. (2000). Dyslexia. Oxford: Blackwell.

Spencer, Ll.H., and Hanley, J.R. (2003). The effects of orthographic consistency on reading development and phonological awareness: Evidence from children learning to read in Wales. *British Journal of Psychology, 94*, 1–28.

Spencer, L.H., and Hanley, J.R. (2004). Learning a transparent orthography at five years old: reading development of children during their first year of formal reading instruction in Wales. *Journal of Research in Reading, 27*, 1-14.

Sprenger-Charolles, L., & Serniclaes, W. (2003). Reliability of phonological and surface subtypes in developmental dyslexia: A review of five multiple cases studies. *Current Psychology Letters*, *10*, 2–8.

Stokes S.F., Wong A.M., Fletcher P., Leonard L.B. (2006). Non-word repetition and sentence repetition as clinical markers of specific language impairment. The case of Cantonese. *Journal of Speech, Language, and Hearing Research, 49,* 219–236.

Summers, C, Bohman, T, Gillam, R., Peña, E. D., Bedore, L. M. (2010). Bilingual performance on nonword repetition in Spanish and English. *International Journal of Language and Communication Disorders, 45,* 480–493.

Taylor, W. L. (1953). Cloze procedure: a new tool for measuring readability. *Journalism Quarterly 30,* 414-438. As cited in Trace (2020).

Taylor, W. L. (1957). "Cloze" readability scores as indices of individual differences in comprehension and aptitude. *Journal of Applied Psychology, 4*(1), 19. As cited in Trace (2020).

Thomas, E.M. and Gathercole, V.C.M. (2007). Children's productive command of grammatical gender and mutation in Welsh: An alternative to rule-based learning. *First Language, 27*(3), 82, 251-278.

Thomas, E. M., and Lloyd, S. W. (2004). Developing language-appropriate task items for identifying literacy difficulties in Welsh-speaking children. *Dyslexia Review, 20*(1), 4-9.

Thomas, E.M., and Mayr, R. (2010). Children's acquisition of Welsh in a bilingual setting: a psycholinguistic perspective. In *Welsh in the twenty-first century.* In D. Morris. Ed. 99-117. Cardiff: University of Wales Press.

Thomas, E. M., Williams, N., Jones, Ll. A., Davies, S., & Binks, H. (2014). Acquiring complex structures under minority language conditions: bilingual acquisition of plural morphology in Welsh. *Bilingualism: Language and Cognition, 17*(3), 478-494.

Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. International Journal of Bilingualism, *15*(4), 426–445.

Thordardotir, E. (2017). Amount trumps timing in bilingual vocabulary acquisition: Effects of input in simultaneous and sequential school-age bilinguals. *International Journal of Bilingualism, 14*(5), 1-20.

Thordardottir, E. (2008). Language-specific effects of task demands on the manifestation of specific language impairment: A comparison of English and Icelandic. *Journal of Speech, Language, and Hearing Research, 51,* 922–937.

Thordardottir, E., and Brandeker, M. (2013). The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *Journal of Communication Disorders 46,* 1-16.

Thordardottir, E., Rothenberg, A., Rivard, M.E. and Naves, R. (2006). Bilingual assessment: Can overall proficiency be estimated from separate measurement of two languages? *Journal of Multilingual Communication Disorders*, *4*(1): 1–21.

Trace, J. (2020). Clozing the gap: How far do cloze items measure? *Language Testing, 37*(2) 235-253.

Umbel, V. M., Pearson, B. Z., Fernandez, M. C., and Oller, D. K. (1992). Measuring bilingual children's receptive vocabularies. *Child Development, 63*(4), 1012-20.

Unsworth, S. (2013). Assessing the role of current and cumulative exposure in simultaneous bilingual acquisition: The case of Dutch gender. *Bilingualism: Language and Cognition, 16*(1), 86–110.

Unsworth, S. (2016). Early child L2 acquisition: Age or input effects? Neither, or both? *Journal of Child Language, 43*(3), 608-634.

Vaessen, A., Bertrand, D., Denes, T., Blomert, L. (2010). Cognitive development of fluent word reading does not qualitatively differ between transparent and opaque orthographies. *Journal of Educational Psychology, 102,* 827–842.

Van den Broeck, W., and Geudens, A. (2012). Old and new ways to study characteristics of reading disability: The case of the nonword-reading deficit. *Cognitive Psychology*, *65*(3), 414–456.

Van Ijzendoorn, M.H., Bus, A.G. (1994). Meta-analytic confirmation of the nonword reading deficit in developmental dyslexia. *Reading Research Quarterly*, *29*, 267–275.

Verachtert, P., DeFraine, B., Onghena, D. and Ghesquiere, P. (2010). Season of birth and school success in the early years of primary education. *Oxford Review of Education*, *36*, 285–306.

Verhoeven, L. and Vermeer, A. (1993). *Taaltoets Allochtone Kinderen Bovenbouw* [Language Tests for Minority Children]. Tilburg, The Netherlands: Zwijsen.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A. and Pearson, N. A. (2013). Comprehensive Test of Phonological Processing Second Edition. Austin, TX: Pro-ed Inc.

Welsh Government, (2007). Defining schools according to Welsh medium provision. https://gov.wales/defining-schools-according-welsh-medium-provision

Welsh Government, (2013a) National Literacy and Numeracy Framework (LNF). Retrieved from https://hwb.gov.wales/curriculum-for-wales-2008/national-literacy-and-numeracy-framework-curriculum-for-wales-2008-version

Welsh Government, (2013b) National Reading Tests. https://hwb.gov.wales/curriculum-for-wales/reading-and-numeracy-assessments/national-reading-and-numeracy-tests

Welsh Government, (2017a). Cymraeg 2050: A million Welsh Speakers. Retrieved from: https://senedd.wales/Laid%20Documents/GEN-LD11108/GEN-LD11108-e.pdf

Welsh Government, (2017b). Professional Standards for Teaching and Leadership. Retrieved from https://hwb.gov.wales/professional-development/professional-standards

Welsh Government, (2017c). Welsh in Education: Action Plan 2017-2021. Retrieved from https://gov.wales/welsh-education-action-plan-2017-2021

Welsh Government, (2020). School census results. Retrieved from https://gov.wales/schools-census-results-january-2020

Welsh Government, (2019). Annual Population Survey results June 2019. https://statswales.gov.wales/Catalogue/Welsh-Language

Wiig, E., Secord, W., and Semel, E. (1992). Clinical Evaluation of Language Fundamentals–Preschool. San Antonio, TX: The Psychological Corporation.

Willis, C. S., and Gathercole, S. E. (2001). Phonological short-term memory contributions to sentence processing in young children. *Memory, 9,* 349-63.

Wimmer, H. (1993). Characteristics of Developmental Dyslexia in a Regular Writing System. *Applied Psycholinguistics, 14*(1), 1-33.

Wimmer, H., & Goswami, U. (1994). The influence of orthographic consistency on reading development: Word recognition in English and German children. *Cognition, 51*, 91–103.

Windsor, J., Kohnert, K., Lobitz, K. F., and Pham, G. T. (2010). Cross-language nonword repetition by bilingual and monolingual children. *American Journal of Speech-Language Pathology, 19*(4), 298-310.

Winter, K. 2001. Numbers of bilingual children in speech and language therapy: Theory and practise of measuring their representation. *International Journal of Bilingualism*, 5: 465–495.

Young, N. E., Rhys, M. & Kennedy, I. & Thomas, E. M. (2017). The linguistic, cognitive, and emotional consequences of minority Language bilingualism. In F. Lauchlan & MC Parafita Couto (Eds.), *Bilingualism and Minority Languages: Current trends and developments*, pp. 120-140. Cambridge Scholar Publications.

Young, Nia E. (2014). *The literacy and self-esteem of children attending Welsh-Medium and English-Medium schools in Wales.* PhD thesis, Prifysgol Bangor University.

Ziegler, J.C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., Saine, N., Lyytinen, H., Vaessen, A., & Blomert, L. (2010). Orthographic depth and its impact on universal predictors of reading: A cross-language investigation. *Psychological Science*, *21*, 551–559.

APPENDIX 1 – Letter to parents and carers (English version only)

<u>LITERACY RESEARCH</u>

Research Title: *Literacy and Welsh: Identifying and exploring the markers of Literacy difficulty in the Welsh language.*

Dear Parent/Guardian,

Carla Owen, a PhD student at the School of Education at Bangor University is currently conducting a piece of research into Welsh literacy difficulties and wondered if you would be interested in your child taking part.

The aim of the study is to identify some of the markers of literacy difficulties in the Welsh language so that first language Welsh-speaking children, who have literacy difficulties such as Dyslexia, can be identified and supported sooner.

Your child would be asked to complete a variety of speaking-reading-writing based activities and each activity would last no longer than 15-30 minutes.

If you would like to be involved in this research **please read the enclosed information sheet**, complete the **consent forms and questionnaire** attached and return it to your child's class teacher by******.**

For further information please contact Carla Owen at edp714@bangor.ac.uk

Thank you very much

APPENDIX 2 – Information sheet for parents and carers (English version only)

Participant information sheet for Parents/Carers

Research Title: Literacy and Welsh: Identifying and exploring the markers of Literacy difficulty in the Welsh language.

Introduction:

Your child is being asked to take part in a study which will be exploring the possible markers of Literacy difficulties in Welsh. Before you make a decision about taking part it is important that you are fully aware of what is involved and what your child will be asked to do. Please take time to read the following information and discuss it with your child and family. Once you have fully understood the information please take time to make a decision about whether you would like your child to take part.

What is the purpose of the study?

The main aims of the study are:

1. To explore potential markers/signs of Literacy difficulties in Welsh through studying and comparing children's reading and writing in English and Welsh.

2. To create a screening tool to be used by professionals working with Welsh-speaking children to help identify Literacy difficulties at an earlier age.

The main purpose of the study is to try to find out what the possible signs are in Welsh that a child has a literacy difficulty such as Dyslexia. The majority of the tests available at the moment are in English and so the study aims to create a screening test in Welsh so that Welsh children with any Literacy difficulties can be identified sooner and get the support they need.

Why has your child been chosen?

Your child has been asked to participate because they are aged between 7 and 11 years of age and speak Welsh as a first language.

Does my child have to take part?

Your child **does not** have to take part in the study. Taking part is completely voluntary and if you or your child changes their mind about taking part at any point in the study they/you have the right to withdraw without giving a reason. If you are happy for your child to take part then please complete the consent forms attached to this information sheet.

What will your child be asked to do?

Carla Owen will be working closely with the school to arrange a suitable time for testing to begin. Testing times will be organised to ensure minimum disruption to the school day. The procedure will be fully explained to your child before they begin and they will be told that they can stop or withdraw from the testing at any time.

Your child will be asked to complete a series of paper and computer based tasks in English and Welsh to measure the following:

Reading (including comprehension), Spelling, working memory, phonological awareness (sounds in words) and IQ. They will also complete a series of tasks based on Welsh grammar, some of which will involve your child saying words into a microphone and their responses recorded as audio files.

Testing will be done in 15-30 minute time slots and your child will be offered frequent rest breaks. It is not anticipated that your child will be tested for longer than 3-3.5 hours in total. The computer based tasks are designed to be fun for children and may involve the use of a touchscreen computer.

You will also be asked to complete a questionnaire about your child's language background.

An adult known to the child will be present or within very close proximity to the testing area and the researcher is fully criminal record checked.

The study will involve approximately 100 children from across North Wales.

Will your child's participation in this study be kept confidential?

Confidentiality is of the utmost importance to us and so will be maintained at all times. Your child's data will be kept on password protected and encrypted files and any hard copies of information about your child will be stored in locked cupboards/lockers within a locked room. Your child will be assigned a unique identification number and so their name will never be associated with the data. Although the data may be published in educational journals, your child's name will never appear in any public document or be presented in a way which would allow your child to be identified.

However, if you or your child shares information that is suggestive of risk to yourself/themselves or others, this will be passed on to a designated individual within your Local Authority.

Are there any risks involved?

There are no expected risks from participation in this study. Your child may experience some fatigue whilst completing the activities; however they will be given regular breaks and encouraged to ask for breaks if they are feeling tired.

What will happen if you (or your child) don't want to carry on with the study?

You have the right to withdraw your child from the study at any time, just inform the researcher. If data has already been collected you may request that these data be removed and destroyed.

What will happen to the results of the study?

The results from the study will be presented in a PhD Thesis and may be published in an educational journal. The results may also be presented at public conferences/presentations. A section has been allocated within the consent form for you to indicate if you wish to receive information about the final results. It is intended that the results from the study will be used to create a Welsh screening test for use by professionals working with children in Wales to help identify children with literacy difficulties such as Dyslexia.

Who is organising and funding the research?

The research is being carried out by Carla Owen, a PhD student at Bangor University under the Supervision of Professor Enlli Thomas and Dr Nia Young. The research is being funded by Bangor University and the ESRC (Economic and Social Research Council).

What happens if I have any concerns about this project?

If you are concerned about any aspect of this project and would like to speak to someone please contact the Head of School of Education - Professor Enlli Thomas tel: 01248 383053 or email enlli.thomas@bangor.ac.uk

Contact for further information:

If you would like more information, or have any questions at all regarding the study then please contact Carla Owen by email at edp714@bangor.ac.uk

Next steps:

If you decide that you would like to take part, please complete and return the enclosed consent forms and questionnaire to your child's class teacher by ***** in the envelope provided.

Thank you so much for taking the time to read this information.

APPENDIX 3 – Bilingual parental consent form

<u>**FFURFLEN CANIATÂD I RHIENI/GOFALWYR**</u>

<u>**PARENTAL CONSENT FORM**</u>

Teitl yr ymchwil: Llythrennedd a'r Gymraeg: Nodi ac archwilio marcwyr o anhawster llythrennedd yn yr iaith Gymraeg

Research Title: Literacy and Welsh: Identifying and exploring the markers of literacy difficulty in the Welsh language.

Rhowch lythrennau blaen eich enw wrth y datganiadau sy'n berthnasol i chi.

Please initial the following statements with which you agree:

	Llythrennau blaen *Please initial*
1. Rwy'n cadarnhau fy mod wedi darllen y wybodaeth a ddarperir, wedi cael amser i fyfyrio ar y wybodaeth ac unrhyw gwestiynau a ofynnir wedi cael eu hateb yn foddhaol. *1. I confirm that I have read the information provided, have had time to reflect on the information and any questions asked have been answered satisfactorily.*	_____
2. Rwyf yn cytuno i ganiatáu i fy mhlentyn gymryd rhan yn yr astudiaeth hon *2. I agree to allow my child to participate in this study*	_____
3. Rwy'n deall bod cyfranogiad fy mhlentyn yn wirfoddol ac y gallant dynnu'n ôl o'r astudiaeth ar unrhyw adeg heb reswm neu gosb. *3. I understand that my child's participation is voluntary and that they may withdraw from the study at any time without reason or penalty*	_____
4. Rwyf yn deall bod gen i hawl i dderbyn crynodeb o ganlyniadau'r astudiaeth ar ddiwedd y prosiect (rhowch gyfeiriad e-bost: _____ **)**	_____

4. I understand that I have a right to receive a summary of the results of the study at the end of the project (please provide an email address: _____)

5. Rwyf yn deall y bydd y manylion cyswllt fy mhlentyn yn cael ei storio ar gronfa ddata gyfrinachol.
5. I understand that my child's contact details will be stored on a confidential database.   _____

6. Byddwn yn fodlon i ti gysylltu a ni ar gyfer astudiaethau yn y dyfodol y gall fy mhlentyn ei gymryd rhan ynddynt.
6. I would be happy to be contacted for future studies that my child may be eligible to take part in.

_____

Dyddiad/ Date: _____

Enw'r plentyn/ Name of Child Participant: _____

Dyddiad geni eich plentyn/Child's date of birth:_____

Enw'r Ysgol a'r flwyddyn dosbarth/ Name of School and Year group:

_____

Enw'r Rhiant neu Gynrychiolydd Cyfreithiol/ Name of Parent or Legal Representative:

_____

Llofnod Rhiant neu Gynrychiolydd Cyfreithiol/ Signature of Parent or Legal Representative:

_____

The researcher and the school will share information about your child's reading scores.

Bydd yr ymchwilydd a'r ysgol yn rhannu gwybodaeth am sgoriau darllen eich plentyn.

APPENDIX 4 – Bilingual parental audio consent form

CANIATÂD CYFRANOGWR I RECORDIO SAIN

PARTICIPANT CONSENT TO AUDIO RECORDING

I gael ei lenwi cyn dechrau'r tasgau.

To be completed prior to completion of tasks.

Rhowch lythrennau blaen eich enw wrth y datganiadau sy'n berthnasol i chi.

Please initial the statements that apply to you.

	Please initial
1. Cytunaf i llais fy mhlentyn cael ei recordio, ac i'r recordiad gael ei ddefnyddio ar gyfer rhesymau sydd wedi eu hegluro i mi *I agree for my child's voice to be recorded and for the recording to be used for the purposes that have been explained to me.*	_____
2. Deallaf y bydd yr holl atebion yn cael eu trin yn **hollol gyfrinachol**. *I understand that all responses will be treated as **strictly confidential.***	_____

Enw'r cyfranogwr:

Participant name/ _____

Llofnod rhiant/

Parent signature: _____

Dyddiad/ Date _____

APPENDIX 5 - Child Information and assent form (English version only)

<u>INFORMATION SHEET FOR CHILDREN</u>

We are asking you to be part of a project about reading and writing in English and Welsh.

WHY ARE WE DOING THIS PROJECT?

We are doing this project to try to look for any signs we can see in your Welsh reading and writing work that shows if you are finding reading and writing difficult. We need to do this so that we can try to find a way of helping children who have reading and writing difficulties at a younger age. This will mean that if you find reading and writing in Welsh difficult you can get some help sooner, without having to wait until you are trying to write in English.

WHAT WILL I BE ASKED TO DO?

You will be asked to do some reading, writing and speaking activities and puzzles in English and Welsh. Some will be on paper and some will be on the computer. The speaking activities will only be in Welsh and we will ask you to say some words and short sentences into a microphone. You will do some of the activities sitting in groups and some on your own. The activities will all be done in school and you will have someone that you know nearby at all times so you will never be completely on your own.

If at any point you don't understand what to do then just ask and we will be happy to answer any questions.

WILL ANYONE KNOW WHAT MY ANSWERS WERE?

No one apart from us the researchers will see the answers to your questions. Your answers will be put together with the answers all the other children give and no-one who reads about the project will know that you took part or what your answers were.

HOW LONG WILL IT TAKE?

Each activity takes about 15-30 minutes. We don't think you will be longer than 3-3.5 hours in total doing all the activities and you will be offered lots of breaks. If you feel like you want to have a break at any time then just let us know.

DO I HAVE TO TAKE PART?

You **do not** have to take part, it is entirely up to you if you want to do the activities or not. If you do decide to take part and then change your mind that is absolutely fine too, just tell someone straight away. No one will be upset or angry with you if you decide not to do the activities.

Thank you so much for taking the time to find out about our project, we hope to meet you soon ☺

CHILD ASSENT FORM

Please read the sentences below and draw a smiley face if you agree with what the sentence says ☺

I have read this information sheet or someone has explained to me what the project is about	
I understand that if I have any questions I can ask them at any time.	
I would like to take part in the project	
I understand that no one will know, after reading about the project, what my answers were or that I was part of the project	
I know that I can stop taking part in the project activities at any time and no one will be angry or upset with me.	
I understand that if I am tired or need to take a break I can tell the adult with me or the person showing me the activities.	

Name: _____

Date: _____

School name: _____

APPENDIX 6 – Parental Questionnaire (English version only)

PARENT/CARER QUESTIONNAIRE

1. Parent/Carer name: _____

2. Child's name: _____

3. In which town/village do you live? _____

4. Which school does your child attend? _____

5. Occupation of parents:

Mother_____

Father_____

6. Annual household income (optional): _____

7. Has your child been diagnosed with Dyslexia? _____

8. Has your child been diagnosed with any other language impairment?

(Please circle)   Yes   No

If you answered yes, please tell us the name of the impairment:

_____

9. Has your child been diagnosed with any other impairment or disorder?

(Please circle)   Yes   No

If you answered yes, please tell us the name of the impairment/disorder:

_____

10. Does your child receive Free School Meals? (Please circle)   Yes   No

11. Does your child speak Welsh? (Please circle)   Yes   No

12. What is your child's first language?

_____

13. Parent first language:

Mother: _____

Father: _____

14. Main language spoken at home:

_____

Thank you very much for taking the time to complete this questionnaire. If you have any questions please do not hesitate to contact the researcher.

APPENDIX 7 - Y Goriad Task 1 and Task 2 Target words

Y GORIAD TASK 1 AND TASK 2 TARGET WORDS

TASK 1

Target answer	Multiple choice options
gannoedd	cannoedd gannoedd channoedd
flynyddoedd	blynyddoedd flynyddoedd mlynyddoedd
ladron	lladron ladron adron
dyfroedd	dyfroedd ddyfroedd nyfroedd
boced	poced boced phoced
mam	mam fam mham
ddau	dau ddau thau
ginio	cinio ginio chinio
wneud	gwneud wneud ngwneud
lawr	llawr lawr ddawr
weld	gweld weld ngweld

fôr-ladron	Môr-ladron Fôr-ladron Mhôr-ladron
wely	gwely wely ngwely
flaenau	blaenau flaenau mlaenau
rhywun	rhywun rywun
traeth	traeth draeth nhraeth thraeth
funud	munud funud ffunud
cotiau	cotiau gotiau chotiau
ddistaw	distaw ddistaw nistaw
gwelodd	gwelodd welodd ngwelodd
ddrws	drws ddrws nrws
blentyn	plentyn blentyn mhlentyn
ryw	rhyw ryw
dlysau	tlysau dlysau thlysau

TASK 2

Original novel word	Target answer	Multiple choice options
Pirall	birall	pirall birall phirall mhirall
Tillog	dillog	tillog dillog thillog nhillog
golldrewop	olldrewop	golldrewop olldrewop ngolldrewop
dorynllyl	ddorynllyl	dorynllyl ddorynllyl norynllyl
crochytob	grochytob	crochytob grochytob chrochytob nghrochytob
dywecos	nywecos	dywecos ddywecos nywecos
llogryfet	logryfet	llogryfet logryfet
rhachol	rachol	rhachol rachol
mysiet	fysiet	mysiet fysiet
bigollych	figollych	bigollych figollych migollych
rhemyl	remyl	rhemyl remyl
tepillt	depillt	tepillt depillt nhepillt thepillt

moledrig	foledrig	moledrig foledrig
tropellig	dropellig	tropellig dropellig thropellig nhropellig

APPENDIX 8 – Sentence Repetition Task items (Mutations Task 4a and 4b)

Sentence Repetition Task/Tasg ail-adrodd brawddegau

Prawf llafar/Oral tasks 4a and 4b

After completing Part a of Task 4 (repeating sentences into a microphone) participants will hear the sentences again one at a time. After each one they will be asked if they think the sentence is correct or incorrect. If they think it is incorrect they will then be asked what sounded wrong in the sentence.

Researcher to put a tick if participant stated was correct or cross if stated was incorrect. A word should be written in the 'incorrect word' column to indicate which word the participant said was incorrect in the sentence.

Sentence	Response (tick/cross)	Incorrect word	Score 0/1
1 Roedd y llong ar dân			
2 Mae'r eliffant yn chwythu ei trwyn			
3 Dwi'n teithio o Caerdydd bob dydd			
4 Torrodd y teigr ei goes			
5 Eisteddwch ar pen y bwrdd			
6 Dyna ddigon!			
7 Dau blodyn hyfryd			
8 Mae'r tân yn boeth			
9 Roedd dwy fuwch yn y beudy			
10 Dy dillad newydd			
11 Enillodd y falwoden y ras			
12 Roedd dwy gwenynen ar y blodyn			

13 Roedd 100 o eiriau yn y llyfr			
14 Ei lygaid glas			
15 Rhedodd y merch i'r ysgol			
16 Pa lliw ydy'r car?			
17 Dwi am redeg yn gyflym!			
18 Dwi'n chwilio am rhaglen ddiddorol.			

APPENDIX 9 – Plurals Tasks items

Tasg Lluosogi/Plurals Task

Plural Task 3: Prawf llafar/Oral Task: (on Computer)

Mynydd – mynyddoedd

Dyn – dynion

Ceffyl – ceffylau

Llyn – llynnoedd

Olwyn – olwynion

Tocyn – tocynnau

Bryn – bryniau

Crys - crysau

Llawr – lloriau

Car – ceir

Bwrdd - byrddau

Tasg 1

Ysgrifennwch y gair am mwy nag un…..

ffon =   ffyn

iâr = ieir

carreg =   cerrig

troed = traed

cadair = cadeiriau

Tasg 2a a 2b

Gwrandwch ar y gair unigol ac ysgrifennwch y gair lluosog.

Gwrandwch ar y gair ffug ac ysgrifennwch y gair lluosog yn defnyddio'r pattrwn rydych chi wedi'i glywed

e.e. ffordd = ffyrdd

    bordd = byrdd

Y pattrwn ydy bod yr 'o' yn newid i 'y'

Rhif	Gair unigol	Gair lluosog	Gair ffug unigol	Gair- ffug lluosog
1	cath	cathod	path	pathod
2	het	hetiau	let	letiau
3	pêl	peli	gêl	geli

4	cwpan	cwpanau	lwpan	lwpanau
5	trwyn	trwynau	ddrwyn	ddrwynau
6	cwch	cychod	twch	tychod
7	cloch	clychau	dloch	dlychau
8	draig	dreigiau	rhaig	rheigiau
9	cwpwrdd	cypyrddau	swpwrdd	sypyrddau
10	llun	lluniau	run	runiau

11	castell	cestyll	mansell	mensyll
12	alarch	elyrch	agarch	egyrch
13	bachgen	bechgyn	wagren	wegryn
14	pabell	pebyll	ffapell	ffepyll
15	corff	cyrff	polff	pylff
16	asgwrn	esgyrn	adwrn	edyrn