

# **A Code of Many Colours: Rationale, Validation and Requirements for a Sound-Based Letter Colour-Code that Might Support Some Children with Dyslexia in Spelling Certain Words**

**by**

**Emily Sarah Cramer**

B.A., University of British Columbia, 2012

Thesis Submitted in Partial Fulfillment of the  
Requirements for the Degree of  
Master of Science

in the  
School of Interactive Arts and Technology  
Faculty of Communication, Art and Technology

**© Emily Sarah Cramer, 2015**

**SIMON FRASER UNIVERSITY**

**Summer 2015**

All rights reserved.

However, in accordance with the *Copyright Act of Canada*, this work may be reproduced, without authorization, under the conditions for "Fair Dealing." Therefore, limited reproduction of this work for the purposes of private study, research, criticism, review and news reporting is likely to be in accordance with the law, particularly if cited appropriately.

## Approval

**Name:** Emily Sarah Cramer

**Degree:** Master of Science

**Title:** *A Code of Many Colours: Rationale, Validation and Requirements for a Sound-Based Letter Colour-Code that Might Support Some Children with Dyslexia in Spelling Certain Words*

**Examining Committee:** **Chair:** Lyn Bartram  
Associate Professor

**Alissa Antle**  
Senior Supervisor  
Associate Professor

---

**Bernhard Riecke**  
Supervisor  
Associate Professor

---

**Alyssa Wise**  
External Examiner  
Associate Professor  
Department of Education  
Simon Fraser University

---

**Date Defended/Approved:** August 6, 2015

## Ethics Statement



The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

- a. human research ethics approval from the Simon Fraser University Office of Research Ethics,

or

- b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University;

or has conducted the research

- c. as a co-investigator, collaborator or research assistant in a research project approved in advance,

or

- d. as a member of a course approved in advance for minimal risk human research, by the Office of Research Ethics.

A copy of the approval letter has been filed at the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library  
Burnaby, British Columbia, Canada

update Spring 2010

## **Abstract**

Dyslexia is a severe impairment in reading and spelling. Despite receiving best-practice remediation, many children with dyslexia fail to surpass the 30<sup>th</sup> percentile in reading and spelling. A major impediment to children's remediation is poor attention, which motivates the development of stronger attentional supports. One intriguing candidate is dynamic colour-coding. We have developed a tangible software system (PhonoBlocks), which could leverage dynamic colour-coding. The present study was undertaken to better understand how to use dynamic colours to support children with dyslexia in learning through PhonoBlocks. I develop a theoretical framework for designing dynamic colour-codes and implement and assess it in a mixed-methods study with PhonoBlocks. My framework addresses a general knowledge gap in how to apply dynamic colour to literacy acquisition in software. I use my findings to identify individual and interface factors that affected children's use of the colours, and recommend general design counter-strategies with specific applications to PhonoBlocks.

**Keywords:** Dyslexia; Software interventions; Colour; Design; Evaluation; Mixed-methods

## Dedication

*To:*

- My parents: Brad and Annette Cramer
- My stepmother, Ginnie Schuster-Cramer
- My brother, Lucas Cramer
- My friends: Eric Pledger, Toni Epp, Spencer Staiger and Sara Tan
- My friend and roommate, Kevin Preston
- My (fictional) role model, Captain Jean-Luc Picard

## Acknowledgements

The study would not have been possible without the gracious cooperation and enthusiasm of the tutors and students at Kenneth Gordon Maplewood School. I am deeply grateful for all the help, insight and kindness that I received.

I am also indebted to my supervisor, Alissa Antle. Alissa has been my main source of insight and encouragement. She has shown me new perspectives on research and design, and the purpose that comes from guiding oneself with the question- *what can we uniquely contribute?*

Thank you.

# Table of Contents

Approval.....	ii
Ethics Statement.....	iii
Abstract.....	iv
Dedication.....	v
Acknowledgements.....	vi
Table of Contents.....	vii
List of Tables.....	xi
List of Figures.....	xii
List of Acronyms.....	xiii
Glossary.....	xiv

<b>Chapter 1. Introduction .....</b>	<b>1</b>
--------------------------------------	----------

<b>Chapter 2. Theoretical Background.....</b>	<b>6</b>
-----------------------------------------------	----------

2.1. A Definition of Dyslexia and Its Social and Emotional Costs.....	7
2.1.1. The Motivation for My Research .....	8
2.2. The Skills that are Needed to Read and Spell English and the General Implications for Literacy Interventions .....	9
2.2.1. Alphabetic Orthographies Represent Phonemes.....	9
2.2.2. Alphabetic Orthographies Require Phonological Awareness.....	10
2.2.3. Orthographic Transparency .....	11
2.2.4. How Anglophone Children Develop Fluency and the Requirements on Visual and Executive Attention .....	13
Determining which Multi-Letter Patterns are Predictive of Sounds .....	15
Executive Functions: Inhibition and Switching.....	16
2.2.5. Implications for Interventions: the Need to focus Visual and Executive Attention .....	17
Summary.....	18
2.3. The Unique Visual and Executive Attentional Challenges of Children with Dyslexia and Implications for Interventions.....	19
2.3.1. Impairments in Visual Processes Involving the Dorsal Stream.....	20
The Dorsal Stream.....	20
Visual Attention Span.....	21
Orienting and Focusing.....	21
Visual Search .....	22
Relevance to Visual Segmentation of Letters.....	23
Relevance to Auditory Segmentation of Phonemes .....	23
Relevance to Orthographic Decoding and Spelling .....	25
Direct (Visual) .....	26
Indirect (Auditory) .....	26
2.3.2. Impairments in Executive Functions That Are Relevant to Reading and Spelling .....	27
2.3.3. Implications for Interventions and Design Requirements.....	29
2.4. Interventions for Struggling Readers.....	30

2.4.1.	Mainstream Interventions: the need for Explicit Instruction in Phonics and Multi-Letter Units and Contexts .....	31
	Lindamood Phonics and Variants (Earobics, FastForWord) .....	31
	Integrated Picture Mnemonics .....	32
	RAVE-O and Other Multi-Letter Unit Approaches .....	33
	Summary .....	36
2.4.2.	Orton-Gillingham Multisensory Curricula .....	36
	Guided Discovery .....	37
	Multisensory Engagement .....	40
2.4.3.	Colour-Coding and Literacy .....	41
	A Rationale for Colour-Coding .....	42
	Attention .....	42
	Memory .....	43
	Information Integration .....	43
	The Potential Application to Literacy Interventions .....	44
	The Typical Colour Perception of Children With Dyslexia .....	46
	Attempts to Use Colour In Literacy Acquisition .....	48
	Reading with Words in Colour .....	48
	Dybuster .....	50
	The Colour Vowel Chart .....	52
	Colour to Discriminate Letters .....	54
	Colour to Highlight Multi-Letter Units .....	55
	Colour to Highlight Abstract Categories .....	57
2.4.4.	The Need for a Principled Framework for Using Colour in Literacy Interventions .....	58
	My Framework for Designing Colour-Coding Schemes .....	60
	General Principles: .....	60
	Design Elements: .....	61
<b>Chapter 3.</b>	<b>Methodology .....</b>	<b>70</b>
3.1.	Research Design, Questions and Hypotheses .....	70
3.2.	Research Instrument Design .....	73
3.2.1.	PhonoBlocks .....	73
	Hardware and Physical Interface .....	73
	Software and Digital Interface .....	74
3.2.2.	Literacy Rules and Rationale .....	75
	Spelling Or Reading? .....	77
3.2.3.	Colour-Coding Schemes and Rationale .....	78
	Scheme 1: particular .....	79
	Scheme 2: categorical .....	79
	Short and Long Vowels .....	80
	Consonant-Le Units .....	81
3.2.4.	Tutor Design Sessions .....	83
	Methodology .....	83
	Results and Revisions .....	84
	Consonant-LE Gemination .....	84
	Short Vowel Discrimination .....	84
3.3.	User Study at Kenneth Gordon Multisensory School .....	85
3.3.1.	Study Design .....	85
3.3.2.	Participants .....	86
	Child Profiles .....	87

3.3.3.	Activities, Colours, PhonoBlocks .....	88
	Worksheet Implementation .....	88
	Consonant-LE colours: .....	89
	Letter flashes.....	90
3.3.4.	Procedure .....	90
3.3.5.	Data Collection.....	94
	Quantitative Measures .....	94
	Pre and Post Assessments .....	94
	Software Event Logs .....	100
	Qualitative Measures .....	102
	Aggregate Analysis and Case Studies .....	104
<b>Chapter 4.</b>	<b>Results .....</b>	<b>105</b>
4.1.	Quantitative Analysis.....	105
4.1.1.	Pre and Post Assessments.....	105
	Ceiling Effects for Vowel Discrimination .....	105
	Performance and Transfer, by Group Medians .....	108
	Pre and Post Performance .....	108
	Transfer .....	110
	Paper and Pencil .....	118
	Summary of Overall Effects: .....	119
	Performance and Transfer, by Individuals .....	120
	Pre and Post Performance: .....	120
	Transfer .....	121
	Paper and Pencil .....	124
	Summary .....	124
4.1.2.	Software Event Logs.....	125
	Correct Second Submissions.....	125
	Correcting Erroneous First Submissions .....	128
	Unproductive Errors .....	129
	Misled by Colour .....	129
4.2.	Qualitative Analyses.....	129
4.2.1.	Hardware Failures.....	130
4.2.2.	Tutor Interviews .....	131
	T1 .....	131
	T2 .....	132
	T3 .....	132
4.2.3.	Observations .....	133
	Children's Memory of the Colour Codes.....	133
	Children's Comprehension of the Rules .....	134
	Children's Use of the Colour Codes.....	137
4.2.4.	Individual Cases: <i>why might have colour worked or not?</i> .....	138
	A Note about the Profiles .....	139
	Children Who Spontaneously Attempted to Use the Colour Codes: P2, P5,	
	P9 and P7 .....	140
	P2 .....	140
	P5: .....	141
	P7 .....	144
	P9: .....	145
	Children who did not spontaneously use colour .....	147
	P1: .....	147
	P8: .....	149

P6:	150
P4	151
P3:	152
P10:	153
4.2.5. Summary of Qualitative Analysis:	153
<b>Chapter 5. Discussion</b>	<b>157</b>
5.1. Competition with Alternate Representational Systems	158
5.2. Lack of Incentive for Correct first responding (which either learning the colour codes or the rules would enable) or Punishment for Incorrect first Responding	159
5.3. Availability of Alternate Strategies	160
5.4. Hardware Failures and Design Limitations	160
5.5. Individual Factors	161
5.6. Implications for Design	165
5.6.1. Engage Low-Level Attention to the Novel Feedback	166
Increase Engagement with the System	166
Reduce Discomfort with the System	166
Increase Perceived Age-Inappropriateness	168
Increase Low-Level Salience of Onscreen Elements	168
Integrate Correlated Sensory Events	168
Incentivize Effective Exploitation of Novel Feedback	170
5.6.2. Engage High-Level Mental Integration	170
Encourage Reflection on the Connections Between Alternate Representations	171
Incentivize Abstraction	173
Assess the Intuitiveness of the Visual Codes	174
5.7. Implications for Study Design	175
5.7.1. Pre-Screen Children on Their Ability to Spell the Target Words	175
5.7.2. Match Children on Age, Favoured School Subject and Motivation to Engage with the System	176
5.7.3. Remove All Feedback that Enables Trial-and-Error Responding	176
<b>Chapter 6. Conclusion</b>	<b>178</b>
<b>References</b>	<b>180</b>
Appendix A. Initial Session Script	191
Appendix B. Assessment and PhonoBlocks Session Words	199
Appendix C: Tutor Interview Questions	204

## List of Tables

Table 2-1.	A summary of some previous attempts to apply colour to literacy. ....	58
Table 4-1.	Summary of Child Profiles .....	139

## List of Figures

Figure 2-1.	How colour might draw attention to the rime unit (top), or the presence and positional restrictions on front and end blends (bottom). .....	44
Figure 2-2.	How colour might support learners in noticing the rule that consonant-le syllables only appear at the end of the word (top), or support them in “grouping” (noticing the relationship between) all closed and all vowel-consonant-e syllables, respectively.....	45
Figure 2-3.	A software application that enables children to spell words could leverage dynamic colours to draw children's attention to the changes in sound that correlate key changes the child makes to the word's letters.....	46
Figure 2-4.	An example of a solution to the design problem of communicating the hard versus soft “c” generalization, developed by applying my framework for designing colour-codes.....	67
Figure 3-1.	The integrated screen and tangible interfaces. ....	74
Figure 3-2.	The screen interface. The word the children spell appears in the middle of the screen. Tapping the “check” button submits the word. Completed words appear in the word history (upper right). .....	75
Figure 3-3.	How the consonant-le (top) and vowel discrimination (bottom) words appeared in the categorical (right) and particular (left) colour-coding schemes.....	83
Figure 4-1.	Pre and post accuracies for short vowel discrimination, split by scheme (categorical or particular) and assessment word appearance (coloured, uncoloured and paper and pencil). ....	107
Figure 4-2.	Left, the median number of consonant-le and consonant gemination errors at pre and post-test, split by colour coding scheme. Right, median accuracies at pre and post-test, split by scheme. ....	109
Figure 4-3.	(Right) Pre and Post Test Accuracies, split by word familiarity and scheme. (Left) pre and post consonant-le formation and gemination errors, split by word familiarity and scheme. ....	112
Figure 4-4.	(Right) Pre and Post Test Accuracies, split by word familiarity and scheme. (Left) pre and post consonant-le formation and gemination errors, split by word familiarity and scheme. ....	115
Figure 4-5.	(Right) post-test accuracies for familiar and unfamiliar words, split by word appearance and scheme (Left) consonant-le formation and gemination errors for familiar and unfamiliar words, split by word appearance and scheme. ....	118

## List of Acronyms

OG	Orton-Gillingham
KGMS	Kenneth Gordon Maplewood School

## Glossary

Term	Definition
Orthographic Pattern	A multi-letter unit that has a regular sound correspondence (e.g., “tion” sounds like /shun/), or a contextual rule (e.g. vowels sound short in closed syllables)
Closed Syllable	A syllable consisting of one or zero consonants followed by a vowel and one additional consonant or consonant digraph, cluster or blend. E.g. “Cat”, “odd”, “end”.
Open Syllable	A syllable consisting of one or zero consonants followed by a vowel. E.g., “ba”, “ta”.
Stable Syllable	A multi-letter unit that stands as a completed syllable. Words are syllabicated around a stable syllable. E.g., “cap/tion”, “sta/ble”.
Consonant-Le syllable	A stable syllable consisting of a consonant, followed by the letters “l” and “e”. For example, “gle”, “ple”, “dle”.
Consonant gemination	Doubling a consonant.
Consonant gemination error	Failing to geminate a consonant or geminating a consonant when unnecessary. In the present study gemination errors occur in the context of spelling consonant-le words. For example, spelling “stable” (consonant “b” should not be geminated) as “stabble”, or spelling “stubble” (consonant “b” should be geminated) as “stuble”.
Consonant formation error	Failing to correctly spell the final syllable in a consonant-le word as the stable consonant-le syllable. Some examples include: spelling the final syllable phonetically, e.g. “stabul” versus “stable” or omitting the final “e”, “stabl” versus “stable”.
Orton-Gillingham Multisensory programs	The major current mainstream research-based curricular intervention for children with dyslexia. Orton-Gillingham (OG) programs (including their derivatives, such as Alphabetic Phonics) apply the empirically-validated techniques of multiple sensory cues (seeing, tracing, pronouncing letters simultaneously), repetition and content blocking to the introduction and rehearsal of literacy concepts.

Dorsal Stream	A visual subsystem composed of projections from primary and secondary visual regions to the dorsal (top) parietal cortex, and which is responsible for attentional selection (focusing, inhibition), visuo-spatial navigation and peripersonal awareness. Functionally comprised in many individuals with dyslexia.
Executive Functions	A class of cognitive processes involved in self-regulation and goal-directed behaviour. Includes sensory (inhibiting irrelevant information, maintaining information in memory), motor (inhibiting or switching motoric responses) and associative (planning, organizing information) processes.

## Chapter 1. Introduction

Dyslexia is a severe impairment in reading and spelling that is discrepant with the individual's IQ and general academic ability (Lyon, Shaywitz & Shaywitz, 2003; Dehaene, 2009). Dyslexia affects about 10% of children in countries that use alphabetic orthographies (Reid, 2013). In alphabetic orthographies, letters represent phonemes. Phonemes are elementary speech sounds. English poses especial challenges to children with dyslexia because English does not correspond letters to phonemes consistently: some letters correspond to multiple sounds (consider the sounds of "a" in "cat", "fade" and "star"); some sounds correspond to multiple letters (consider the letters for /s/ in "cite" and "sit") (Ziegler, Bertrand, Tóth, Csépe, Reis, Faísca, & Blomert, 2010). To read and spell English, children must learn additional contextual rules (e.g., vowels sound short in closed syllables and long in open syllables) or multi-letter units that have consistent sounds (Archer, Gleason & Vachon, 2003; Ziegler & Goswami, 2006). The emphasis in English-speaking cultures on reading and writing prevents many children with dyslexia from realizing their intellectual potential and exacts considerable social and emotional costs (Reid 2013).

Although dyslexia is heterogeneous, there is a consensus among researchers that a core cause is poor phonological awareness (Snowling, 1981; Wagner & Torgesen, 1987; Bruck, 1992). Poor phonological awareness can be traced partly to compromised selective attention and executive control (Altemeier, Abbott & Berninger, 2008; Vidyasagar & Pammer, 1999). Compromised visual attention and executive control also contribute to the difficulties Anglophone children face in learning the multi-letter rules and contexts of English (Cestnick & Coltheart, 1999; Valdois, Bosse & Tainturier, 2004). Accordingly, practitioners and researchers have applied the insights of researchers by developing intervention strategies that target the dysfluent reading and spelling at the level of visual (Hines, 2009; Berninger, Abbott, Brooksher, Lemos, Ogier, Zook & Mostafapour, 2000; Bhattacharya & Ehri, 2004) and executive attention (Birsh, 2011; Cox, 1974).

Orton-Gillingham (OG) multisensory curricula are the mainstream example. OG programs apply the empirically-validated techniques of multiple sensory cues (seeing, tracing, pronouncing letters simultaneously), repetition and content blocking to the introduction and rehearsal of literacy concepts (Birsh, 2011; Cox, 1974; Henry, 1998). Over the long term, OG interventions can boost the reading and spelling of children with dyslexia to at least the 30<sup>th</sup> percentile (Oakland, Black, Stanford, Nussbaum & Balise, 1998; Ritchey & Goeke, 2006). Unfortunately, many children fail to achieve even these gains, and few completely close the achievement gap (Alexander & Slinger-Constant, 2004).

Poor attention has been identified as the main impediment to children's benefitting from classroom interventions (Alexander & Slinger-Constant, 2004). A key task of interventions is directing children's attention to linguistic information that is relevant to learning specific reading or spelling rules. Although the strategies and goals of mainstream interventions are effective, the paper and pencil medium may provide insufficient attentional support. What is needed is a medium that provides stronger attentional support.

Software mediums enable dynamic, immediate and variegated feedback. Developers have used software to guide the attention of children with other learning challenges, such as ADHD (Blok, Oostdam, Otter & Overmaat, 2002; Bender & Bender, 1996). Developers have implemented certain learning activities for children with dyslexia (Alexander & Slinger-Constant, 2004). Software supports for dyslexia use various visual and auditory cues, but are generally limited to teaching isolated letter or letter-unit and sound correspondences (Blok, Oostdam, Otter & Overmaat, 2002; Ecalle, Magnan & Calmus, 2009; Magnan, & Ecalle, 2006; Regtvoort & van der Leij, 2007; Hook, Macaruso & Jones, 2001; Diehl, 1999). It is unclear what kinds of sensory supports would be effective for teaching children English's more complex orthographic rules.

My thesis investigates whether dynamic colour-coding in the context of a tangible software application could help children with dyslexia learn English reading and spelling rules. Researchers have used colour-coding in other learning contexts to support learners in remembering and integrating information from multiple sources, for example,

pictorial and textual descriptions of biological processes (Ozcelik, Karakus, Kursun & Cagiltay, 2009). Researchers and educators have also used colour in literacy contexts, where colouring important letters or letter units can help children attend to them (Berninger, Abbott, Brooksher, Lemos, Ogier, Zook & Mostafapour, 2000; Hines, 2009). Colour could be particularly beneficial to children with dyslexia. Distinctive colours support rapid attentional focusing and grouping (Green, & Anderson, 1956; Treisman, 1982; Wolfe & Horowitz, 2004). Children with dyslexia have equivalent colour perception to those without dyslexia (Dautrich, 1993): they can rapidly focus to coloured targets and ignore uncoloured distracters (Vidyasagar & Pammer, 1999) and show a preference for visual encoding strategies (Miller & Kupfermann, 2009), which colour could support (Kast, Meyer, Vogeli, Gross & Jancke, 2007; Kast, Baschera, Gross, Jancke & Meyer, 2010).

My thesis research is part of a larger design project to develop a tangible software system for children with dyslexia. My design goal was developing effective visual supports for my lab's tangible software system. On the basis of earlier speculations about possible links between synaesthesia and learning, my team encouraged me to investigate the potential of colour-coding as a visual interface strategy. My research revealed that, while colour holds promise as an attentional guide for literacy acquisition, the field lacks a guiding framework that would help designers determine what linguistic features to colour-highlight or how to map colours to linguistic information. My research goal was articulating general principles that could serve designers by identifying these questions and procedures for answering them.

To achieve my goals, I analyzed the attentional demands of reading, the attentional challenges of children with dyslexia, and the attentional benefits of colour to develop, apply and thereby assess a candidate theoretical and design framework.

Although the decision to implement my colour codes in a system that uses a tangible (as opposed to a purely graphical) user interface limits my claims about colour to the interaction of colour and tangibility, it increased the applicability of my results to our tangible system. On the basis of the positive effects of kinesthetic engagement (such as tracing (Birsh, 2011; Hulme, Monk & Ives, 1987)) in literacy interventions for dyslexia,

we considered tangibility to be an essential feature of our system. I therefore considered the benefit of increased applicability to a tangible user interface greater than the cost of decreased applicability to purely graphical user interfaces.

The core principles of my framework are that colour codes should be designed around a particular literacy rule, and that colour should highlight all and only the linguistic variables that are relevant to that rule. Four auxiliary “design elements” specify the key design decisions that my general principles entail, and outline the main design alternatives, their trade-offs, and what designers should consider in choosing between alternatives. I applied my framework to develop two colour-coding schemes for two different linguistic rules, which I implemented in “PhonoBlocks”. PhonoBlocks is a tangible software system that implements OG style spelling activities (Antle, Fan & Cramer, 2015).

To assess my framework, and investigate whether my colour-codes supported rule acquisition and transfer to new or uncoloured words, I conducted a month-long, quasi-experimental multiple case-study. Children practiced both linguistic rules, but each child experienced only one colour-coding scheme. Therefore, one rule matched the child’s scheme; the other did not. My framework predicts that children should learn the matched rule better than the mismatched rule. I assessed whether colour-benefits transferred by comparing pre-post intervention performance for instructed versus uninstructed and coloured versus uncoloured words. Although I only assessed two rules, I designed my framework to be general and to guide the design of colour codes for any literacy rule. I provide some examples of how my framework would colour-code alternate rules, but further research is needed to confirm whether the empirical consequences of applying my framework are similar across different rules.

One presupposition was that part of how colour would help was by cueing children’s attention dynamically to different parts of a word. Because dynamic colour codes are most easily implemented in software, I frame my study in the context of designing software applications for children with dyslexia. Although colour might support children without dyslexia in acquiring English reading and spelling rules, the present research is limited to children with dyslexia. Because children with dyslexia have unique

visual attentional challenges, finding a positive effect of colour-coding for children with dyslexia does not imply a positive effect of colour-coding for children without dyslexia. Because I conducted my research in the wider context of designing PhonoBlocks, I implemented my codes in PhonoBlocks. Although colour is a low-level visual feature, and it is likely that my results would generalize to other software systems, it is outside the scope of my thesis to make such claims.

I divide the remainder of my thesis into five chapters. Chapter 2 presents the theoretical background that informed my research questions and my framework of design requirements for colour-codes as literacy supports. Chapter 3 describes my study methodology. Chapter 4 describes the results; chapter 5 discusses the results. Chapter 6 describes my research contribution, the limitations of my study and some directions for future research.

## **Chapter 2. Theoretical Background**

I adopted a theory-based approach to develop my colour codes and how I implemented them in the tangible software system. This section describes the research and previous interventions for children with dyslexia from which I derived the rationale and design requirements for my colour-coding schemes and the research questions for my study.

I divide this chapter into six sections:

First, I define dyslexia and describe its social and emotional costs. I state the motivation for my research, which is designing more effective attentional interventions for helping children with dyslexia to read or spell.

Second, I discuss the processes of reading and spelling English, in terms of the demands upon selective visual and auditory attention and on executive attentional control. I conclude by stating some general implications that are relevant to literacy interventions for children with and without dyslexia.

Third, I summarize research on the core attentional challenges of children with dyslexia and the more specific implications for literacy interventions for children with dyslexia. I translate these implications into design requirements and goals for dynamic colour codes as a software interface feature.

Fourth, I review studies on the kinds of specific intervention features that have yielded improvements in the reading and spelling of children with and without dyslexia, and I review the overall strategies of the contemporary multisensory approaches that synthesize the evidence from the experiments that I review. With reference to the attentional challenges of children with dyslexia, I identify possible limitations in how contemporary interventions implement their strategies.

Fifth, I introduce dynamic colour that is implemented in software as a possible way of overcoming these limitations. I first provide a general theoretical rationale for colour in terms of its uses in multimedia learning applications and information visualization. Next, I provide a more specific rationale in terms of the typical colour-processing of children with dyslexia and how colour could compensate for the difficulties children with dyslexia have in visual inhibition, visually segmenting letter-like stimuli and using letter-like stimuli to re-focus attentional set. The analysis that I describe in this section informs my design framework.

In the sixth and final section, I review attempts to use colour to support various aspects of literacy acquisition. I argue that the knowledge we can glean from these studies is lacking because there have been no systematic assessments of the assumptions that have driven the design and implementation of the codes that were used.

## **2.1. A Definition of Dyslexia and Its Social and Emotional Costs**

Dyslexia is a severe impairment in reading and spelling that is discrepant with the child's IQ or other academic abilities (Lyon, Shaywitz & Shaywitz, 2003; Dehaene, 2009). Dyslexia affects about 10% of children in countries that use Alphabetic orthographies (Reid, 2013). In Alphabetic orthographies, visual units (letters) represent speech sounds (phonemes). I detail Alphabetic orthographies and the challenges they pose in section 2.2.

Dyslexia is diagnosed when there is an absence of socio-economic or emotional factors which might otherwise explain the reading and spelling impairment (Lyon, Shaywitz & Shaywitz, 2003). Although the considerable heterogeneity of children with dyslexia eludes a full causal profile, there is some common etiology. Many children with dyslexia suffer comorbidities in ADHD (Reid, 2013; Birsh, 2011). Those who are not diagnosed with ADHD experience other attentional disturbances, which can be observed in tasks that assess serial visual attention (Vidyasagar, & Pammer, 1999) perceptual set-shifting (Poljac, Simon, Ringlever, Kalcik, Groen, Buitelaar & Bekkering, 2010; Stoet,

Markey & Lopez, 2007) and executive functions involving phonology, such as verbal digit span and rapid automatized naming (Berninger, Abbott, Thomson, Wagner, Swanson, Wijsman & Raskind, 2006; Wolf, Bally & Morris, 1986).

Despite these disturbances, many children with dyslexia perform at or above grade level in tests of non-verbal intelligence (Das, J. P., Mishra, R. K., & Kirby, J. R. 1994) and can perform typically in other domains requiring the acquisition and application of symbolic rules, i.e., mathematics and science (Reid, 2013; but see Joffe, 1981). Despite their symbolic and conceptual skills, children with dyslexia struggle academically (Frederickson & Jacobs, 2001). Contemporary curricula, as well as most careers, require high levels of decoding and writing fluency (Cunningham, 1998). For example, the many long and difficult to decode words in science textbooks (“asphyxiation”, “hydrolysis”) may curtail a child’s interest in science (ibid). Similarly, the poor writing of children with dyslexia prevents them from expressing their often considerable conceptual understanding. Children consequently receive poorer scores than their effort and comprehension warrants, and may lose their motivation to study.

### **2.1.1. The Motivation for My Research**

Many governments have recognized the loss in intellectual potential that dyslexia causes and have overhauled early reading intervention, most notably by emphasizing analytic phonics (Ziegler & Goswami, 2006). Well-intentioned and evidenced as such overhauls may be, they have largely failed to close the fluency gaps between children with and without dyslexia (ibid). The gap is especially pronounced in English-speaking countries, largely because reading and spelling English requires additional skills to analytical phonics and imposes additional attentional demands. The existence of such gaps motivates more specific interventionist strategies, particularly for Anglophone children with dyslexia.

The scope of my thesis is developing strategies for software interventions for Anglophone children with dyslexia. The strategy that I investigate is dynamic colour coding. My research contributions are A) my theory-grounded framework for designing dynamic colour-codes, which explains how to use colour to direct attention to relevant

multi-letter units and contexts, and B) evaluating my framework through a quasi-experimental multiple-cases study.

The remaining sub-sections describe my analysis of theory on the attentional demands of reading English, the attentional challenges of dyslexia, the principles of interventions that have worked and the attentional benefits of colour codes. My analysis grounds the rationale and requirements of the framework I investigate, by which I designed my colour-codes, and the research questions that shaped my study methodology.

## **2.2. The Skills that are needed to Read and Spell English and the General Implications for Literacy Interventions**

### **2.2.1. Alphabetic Orthographies Represent Phonemes**

An orthography is a system of relating visual (squiggles and lines) to auditory (sudden changes in the amplitude and frequency of continuous speech) stimuli (Apel, 2011). The size of the represented auditory units differentiates orthographies. Some (e.g., Korean) represent syllables; some (e.g. traditional Chinese) represent words. English, along with many European languages, uses an Alphabetic orthography. Alphabetic orthographies represent phonemes.

Phonemes are the smallest auditory unit that speakers of a language distinguish and manipulate. Phonemes are not identified to auditory stimuli. Phonemes are abstract mental categories, the boundaries of which are imposed by the speakers of a language. The boundaries cover different regions of the auditory frequency spectrum. Variations of sound falling within these boundaries (“phones”) are not perceived as different sounds; those falling outside the boundaries are (Serniclaes, Ventura, Morais, & Kolinsky, 2005). For example, the sound that corresponds to “t” in the words “teeth” and “stop” are different phones: the frequency spectra differ. Because they both fall within the frequency interval that English classifies as the /t/ phoneme, Anglophones perceive them as the same (ibid).

Alphabetic orthographies correspond letters to phonemes (e.g., “t” represents /t/). Reading an Alphabetic Language involves a) identifying each letter b) “recoding” the letter into a phoneme (e.g., “star” becomes /s/, /t/, /a/, /r/) c) holding and blending the phonemes in auditory working memory into a phonological representation of a word /star/ d) using the phonological representation of the word as a “retrieval cue” for the sensory and cognitive associations that constitute the word’s meaning (Wagner & Torgesen, 1987). Writing is a complementary process of retrieving the phonological representation of a word, holding it in memory and segmenting it into phonemes, and then mapping the phonemes to letters and recording them on the page (ibid).

### **2.2.2. Alphabetic Orthographies Require Phonological Awareness**

Alphabetic Orthographies may be more challenging to learn than those involving larger-grain units, such as syllables. Humans do not naturally segment speech into phonemes (Liberman, Shankweiler, Fischer & Carter, 1974; Read, Yun-Fei, Hong-Yin & Bao-Qing, 1986; Mann, 1986) though even infants show some sensitivity to syllable speech boundaries (Fox & Routh, 1975). Phoneme boundaries are arbitrary, in the sense that acoustic differences of equivalent magnitude might not constitute a boundary, and that different languages can use widely different boundaries (ibid; Serniclaes, Ventura, Morais, & Kolinsky, 2005). Although children spontaneously acquire the capacity to speak their native language, such abilities do not typically coincide with the ability to consciously segment language into phonemes (Fox & Routh, 1975; Liberman, Shankweiler, Fischer & Carter, 1974).

Awareness of the existence of sound units smaller than the syllable is “phonological awareness” (Ball, 1997). “Phoneme awareness”, a more specific subclass, is awareness of the phonemic structure of words. Phonological and phoneme awareness comprise several measureable skills, such as recognizing rhymes, segmenting words into phonemes, deleting phonemes from words and reading nonsense words (Wagner & Torgesen, 1987; Bruck, 1992). Phonological awareness is the strongest predictor of literacy in Alphabetic Languages (Ball, 1997; Snowling, 1981; Schatschneider, & Torgesen, 2004; Ehri & Robbins, 1992; Wagner & Torgesen, 1987; Ziegler, Bertrand, Tóth, Csépe, Reis, Faísca & Blomert, 2010; Casalis, Colé, Sopo, 2004; Blachman, 1991;

Bruck, 1992). Phonological and phoneme awareness are necessary for successful reading and writing given that both of these tasks involve the representation of single phonemes and their letter correspondents [ibid].

### **2.2.3. Orthographic Transparency**

The consistency of the mapping between single letters and phonemes differentiates Alphabetic Orthographies. “Orthographic transparency” refers to the strength of the relationship between the single letters of the orthography’s alphabet and the single sounds of the language’s phonology: languages with strong correlations between letters and sounds are transparent; those with weak correlations are opaque (Ziegler, & Goswami, 2006; Goswami, Ziegler, Dalton, Schneider, 2003; Ehri, 2014). For example, Spanish and Finnish are highly transparent because there is a near perfect correlation between letters and sounds: a given letter always represents the same phoneme; a given phoneme is always represented by the same letter. Accordingly, learning the mappings between letters and sounds equips children to decode any word. Another way of phrasing this is that learners do not need to consider any contextual factors- apart from the specific letter or sound in question - to read or spell a letter or sound. Barring the hurdle of developing phonological awareness, children quickly achieve fluency in reading and writing transparent orthographies (Ziegler, & Goswami, 2006; Seymour, Aro, & Erskine, 2003; Frith, Wimmer & Landerl, 1998).

By contrast, English is opaque. English has 26 letters and 46 phonemes. A bijective mapping between letters and phonemes is therefore impossible. Consequently, many English letters correspond to multiple sounds (consider the sound of “a” in “star”, “cat” and “fade” (Ziegler & Goswami, 2005); many sounds correspond to multiple letters (for example, /s/ may be spelled with “c” or “s”; consonantal sounds are sometimes spelled with two [“bubble”] and sometimes one [“bugle”] consonant). This means that children cannot read English strictly via attending to each letter in a serial fashion, nor spell English by applying single letter-sound correspondences. Predicting a letter’s sound, or a sound’s letter, requires attention to other contextual information (e.g, adjacent letters or syllable boundaries) (Goswami, Ziegler, Dalton, & Schneider, 2003; Treiman & Kessler, 2006). Learning these contexts poses additional challenges to

Anglophone children (Ziegler, & Goswami, 2006; Seymour, Aro, & Erskine, 2003; Frith, Wimmer & Landerl, 1998; Wolf, Miller & Donnelly, 2000; Archer, Gleason & Vachon, 2003).

Although there is some debate about which kinds of contexts are actually primary to learners, theorists agree about which kinds of contexts are statistically predictive of English consonant and vowel pronunciations and spellings. Any of these contexts are candidate foci for intervention.

Vowels are the most inconsistent letters in English orthography and pose the greatest challenge to all children, those with and without dyslexia (Bhattacharya & Ehri, 2004). Accordingly, many researchers have focused on the contexts that predict vowel sounds and how interventions could teach children to exploit them. My intervention also targets vowel sounds. Here, I review two contexts (syllable types and multi-letter units) that my intervention targets.

One context that is relevant to how vowels sound, and that has bearing on children's spelling decisions, is the type of the syllable that encloses the vowel (Archer, Gleason & Vachon, 2003). There are six syllable types: open (ending in a vowel, e.g., "bu" in "bugle"); closed (ending in a consonant, e.g., "cat"), vowel-consonant-e (a closed syllable with an "e" on the end, e.g., "fade"), r-controlled (ending in "r", e.g., "star"), vowel team (involving a vowel digraph, e.g., "peak") and consonant-le (a stable syllable formed from a consonant and "le". For example, "ble", as in "bubble"; "ple", as in "staple"). Barring some exceptions, vowels sound "short", e.g., "ah", "ih", in closed syllables, "long", e.g., "ay", "eye", in open and vowel-consonant-e syllables and they sound "r-controlled" in r-controlled syllables. Consonant-le syllables are indirectly relevant to vowel sounds because they might reveal the syllable that encloses a vowel. For example, recognizing "ble" in the words "stable", "bubble", reveals that the first vowel occurs in an open syllable ("sta"), while the second occurs in a closed one ("bub"). Syllable types are relevant to reading *and* spelling. Children who recognize syllable types know, for example, to pronounce "stubble" as *stuh-bull* (versus *stew-bull*); they also know to spell *stubble* with two "b"s (versus one "b").

English also contains many frequently-occurring sub-word units that have consistent pronunciations, i.e., vowel sounds are stable within them (Goswami, Ziegler, Dalton & Schneider, 2003). This set includes affixes (“pre”, “ness”), rimes (the final vowel and consonants of words, e.g., “et” and “ake” from “pet” and “take”) and the Greek and Latin roots (“phob”, “sect”) that frequent English words. There is evidence that fluent English readers are sensitive to these units and apply them to read or spell unfamiliar words (Devonshire, Morris & Fluck, 2013; Ehri & McCormick, 1998; Ehri & Wilce, 1979; Ehri, 2014).

#### **2.2.4. How Anglophone Children Develop Fluency and the Requirements on Visual and Executive Attention**

Anglophone children become fluent by learning the multi-letter contexts that predict single letter sounds. Share (Share, 2004; Share, 1997; Share, 1995) maintains that children naturally develop fluency through applying their knowledge of single letter-sound correspondences, the “self-teaching” mechanism, to identify the larger letter-units that are phonologically regular and frequently occur. To support this, Share provides data that children are faster and more accurate at reading nonsense words containing units they read previously in different nonsense words than words that do not (Cunningham, Perry, Stanovich & Share, 2002). Share assumed that Anglophone children would learn and apply predictive letter contexts without explicit instruction. Although Share's laboratory studies support this assumption, natural observations do not. After 6 years of reading instruction, Anglophone children read at grade levels three times below Swedish children who have experienced only 1 year of reading instruction (Seymour, Aro, & Erskine, 2003). The discrepancy exists even when comparing students from schools that are subject to the same reading curricula, such as in England and Wales (Welsch being a transparent language) (Ellis, & Hooper, 2001)

The fact that Anglophone children lag behind children learning transparent languages in developing literacy suggests there are some general cognitive impediments to learning predictive *multi* letter-contexts and sound correspondences that are less disruptive to learning *single* letter-sound correspondences (Ziegler, & Goswami, 2006). These impediments affect Anglophone children with and without dyslexia.

Understanding them can guide the design requirements and rationales for systems that seek to improve children's acquisition of English letter-sound correspondences. Other theorists have shed some light on what these impediments might be.

Berninger (Berninger, Vaughn, Abbott, Brooks, Begayis, Curtin & Grahm, 2000; Berninger, Yates & Lester, 1991; Berninger, Abbott, Brooksher, Lemos, Ogier, Zook & Mostafapour, 2000) and Ehri's (Ehri & Wilce, 1979; Ehri & McCormick, 1998; Ehri, 2014) frameworks for how children read and spell English words are centered on two key principles: a) strengthening the connections between specific orthographic patterns and pronunciations and b) doing so in such a way that enables readers to access representations at "lower" levels and thus flexibly shift between different decoding strategies when reading unfamiliar, multi-syllable words. Unlike Share and Stanovich, Berninger and Ehri recognize the corresponding challenges in *noticing* relevant multi-letter units, associating them to their sounds and flexibly shifting attention between different kinds of multi-letter information. Ehri and Berninger recommend that interventions for children, with and without dyslexia, target these challenges.

Berninger's connectionist framework conceptualizes literacy development as the adjustment of weights between input (letter detectors) and output (pronunciations) nodes in a multi-layer neural network. The first layer detects correspondences between high-frequency single letter and digraph or blend units and sounds. The second layer detects correspondences between combinations of the first-layer units (e.g., stable syllables, affixes, rimes) and sounds. The third layer consolidates these into representations of "whole words" that either occur frequently, or are completely irregular (and need to be read "by sight", for example, "yacht") (Berninger, Vaughn, Abbott, Brooks, Begayis, Curtin & Grahm, 2000; Berninger, Yates & Lester, 1991; Berninger, Abbott, Brooksher, Lemos, Ogier, Zook & Mostafapour, 2000). Ehri proposed a "phase" model of sight-word reading that is similar to Berninger's: children progress from representing words as strings of single letters to strings of multi-letter units and finally to whole words. Ehri's model culminates in the "automatization" phase, wherein the printed representation ("orthographic pattern") of frequently encountered word units or whole words enables immediate access to the word's pronunciation (Ehri & Wilce, 1979; Ehri & McCormick, 1998; Ehri, 2014). Ziegler and Goswami (Ziegler & Goswami, 2005) conducted many

experiments to support their claim that readers are sensitive to, and unconsciously learn and apply, their orthography's most statistically regular letter-sound units. Each framework presumes that words activate "multiple levels" of orthographic representation, which learners must coordinate.

### ***Determining which Multi-Letter Patterns are Predictive of Sounds***

Each framework is based on the presupposition that reading English demands visual attention to multiple grains of orthographic unit. These units include single letters, but generalize to frequently-occurring multi-letter units. Children must identify these units via noticing correlations between the units and specific stable pronunciations. Thereafter, children can rapidly read or spell unfamiliar words that follow these memorized patterns. For example, readers can quickly decode or spell words like "grubble" by noticing (and thus dividing the word into) the two sub-units: "ble" and "grub".

There is considerable research to ground such ideas. Fluent English readers are sensitive to multi-letter units when reading unfamiliar words in ways that readers of transparent languages are not (Share, 2004). English readers decode nonsense words with familiar rimes ("pake") faster than words with unfamiliar rimes ("drim") (Goswami, Ziegler, Dalton & Schneider, 2003). English spellers apply consonantal context patterns when selecting the vowels to spell nonsense words (e.g., they will spell the /aw/ sound with "a" after "w", but "o" after "p") (Treiman & Kessler, 2006).

Although children without dyslexia might learn these patterns implicitly, both Berninger and Ehri, as well as other theorists (Birsh, 2011; Archer, Gleason & Vachon, 2003), agree that children would likely achieve literacy quicker if interventions guided attention to the relevant units and contexts. Children's errors in decoding multi-syllable words suggest that they cannot represent syllables effectively, i.e., they might omit a syllable and decode "convention" as "vention" (Archer, Gleason & Vachon, 2003; Bhattacharya 2006). Children may learn a strategy but fail to apply it when faced with a multi-syllable word (Archer, Gleason & Vachon, 2003). Instruction that explicated the proper units for children would take much of the guesswork out of developing an effective multi-letter vocabulary, and could increase the rate at which Anglophone children achieve comparable fluency to children learning transparent languages.

## ***Executive Functions: Inhibition and Switching***

Noticing relevant multi-letter units and contexts is the first cognitive impediment to English children's literacy. The second is flexibly shifting attention between them in reading and spelling multi-syllable words. For example, a word like "stubble" involves a blend ("st"), the stable syllable ("ble") and the need to apply the pattern of doubling the consonant ("b") when vowels sound short. The latter rule requires learners to represent the "closed syllable" ("stub").

The cognitive processes that are responsible for coordinating multiple codes are "executive functions" (Berninger, Yates & Lester, 1991). Executive Functions are cognitive processes involved in self-regulation and goal-directed behaviour. Two executive processes that are involved in reading are *inhibition* and *task switching* (Altemeier, Abbott, & Berninger, 2008; Berninger, Abbott, Thomson, Wagner, Swanson, Wijsman & Raskind, 2006).

Inhibition is the ability to suppress irrelevant information. Inhibition applies to coordinating multiple orthographic codes. When reading or spelling a word like "stubble", which requires the learner to remember the relationship between a doubled consonant and short vowel sound, a beginning reader may need to temporarily inhibit the blend ("st") and the vowel's specific identity ("u" versus "a") so that they can focus on the vowel sound category (short or long) and the consonant of the stable unit. At a lower level, reading or spelling any particular letter-sound unit requires children to ignore letters or sounds that are irrelevant to the task, i.e., to narrow and focus their attentional windows (Brosnan, Demetre, Hamill, Robson, Shepherd & Cody, 2002). In English, where the number of letters (i.e., size of orthographic context needed) to read or spell a single letter varies, readers may also need to dynamically re-adjust the size of their attentional windows (Bosse, Tainturier & Valdois, 2007)

Task switching- also called "set shifting" – involves two abilities. The first is flexibly shifting attention, as well as relevant cognitive strategies or motor responses, between stimuli, as per a task's changing demands. Failures to "set shift" incur "shifting costs" to reaction times for responses that follow the shift. Slowdowns (costs) for changing tasks therefore suggest that these tasks involve different attentional sets, i.e.,

set shifting. Goswami uncovered a slowdown in English readers for switching between reading nonsense words involving familiar (“pake”) to unfamiliar rimes (“drim”). The former could be read by applying memorized rimes. The latter could not, and had to be read letter-by-letter (phonologically). Goswami interpreted her results to mean that a) reading English involves set shifting b) switching attention and strategies between different orthographic units is challenging, even to fluent adults (Goswami, Ziegler, Dalton & Schneider, 2003). Such costs are analogous to those observed in traditional “Stroop” interference assessments of task switching, where participants are slower at naming colours or words in mixed than in blocked trials. Returning to our earlier example, set switching might play a role in helping readers shift from determining whether or not to double the consonant to identifying or spelling the blend (“st), or at a lower level, in helping readers perceptually shift between whole-word, large-units or single-letter decoding (Brosnan, Demetre, Hamill, Robson, Shepherd & Cody, 2002).

Achieving literacy in English therefore imposes considerable demands upon executive functioning. The fact that children do not develop mature executive functioning until about grade seven (Zelazo, Carlson & Kesek, 2008) provides another reason for why Anglophone children take longer to become fluent than children learning transparent languages. Similarly, the involvement of executive functions in reading could explain why children with attentional challenges take longer to become fluent than those without attentional challenges.

#### **2.2.5. Implications for Interventions: the Need to focus Visual and Executive Attention**

The nearly world-wide response to children’s difficulty developing phonological awareness was emphasizing analytic phonics in beginning reading instruction (Ziegler & Goswami, 2005). Early literacy programs for Anglophone children also emphasize phonics, but because English is opaque, developing literacy in English requires mastery of multi-letter sound correspondences and other contextual patterns. The fact that Anglophone children take longer to become literate than those learning transparent languages suggests that implicit learning is insufficient for acquiring English letter-sound correspondences at acceptable rates, even for children without dyslexia.

Interventions that seek to assist children with dyslexia in developing English literacy can benefit from the insights of theorists who applied knowledge of the difficulties of Anglophone children without dyslexia to develop interventions for them. In contrast to Share, neither Berninger, Ehri nor Goswami supposed that children naturally developed fluency with English's multi-letter patterns.

Berninger accordingly recommended and developed interventions that strengthen the links between orthography to phonology via a) repetition and immersion in many examples, such as by showing children flashcards with words printed upon them and b) visual explication of the connections between letter units and sounds, for example, by colour-highlighting the units in words to which children should attend, separately enunciating each sound, and visually connecting each sound to its unit (Berninger, Abbott, Brooksher, Lemos, Ogier, Zook & Mostafapour, 2000). Similarly, Ehri developed and explored instructional techniques that focused on explicating the connections between the letter-units and sounds that were particular to the child's current "phase". When children mastered a specific phase, the focus of instruction switched: children's attention was then drawn to units of the next-largest grain (Ehri & McCormick, 1998). Goswami interpreted the finding that English speakers are sensitive to rime analogies as warranting rime-analogy training in early literacy intervention (Ziegler & Goswami, 2006).

## **Summary**

Reading Alphabetic Languages requires segmenting the speech stream into phonemes, segmenting text into letters, and linking letters to phonemes

Reading opaque languages, such as English, requires the acquisition of additional multi-letter sound correspondences and contextual patterns.

Acquiring these additional correspondences poses two main cognitive challenges. These are: detecting the multi-letter contexts that are statistically correlated to sounds and coordinating between various levels of multi-letter unit in reading and spelling single words or sentences involving different units.

Even children without dyslexia struggle to acquire English's additional contextual rules. Many theorists argue that interventions should better explicate relevant letter-sound units and patterns, and should emphasize executive skills.

Although this section focused on challenges that are shared by children without dyslexia, children with dyslexia have them as well. Interventions for Anglophone children with dyslexia must target phonological awareness as well as learning and applying the multi-letter units of English.

Although the tasks of children with and without dyslexia learning English are similar, children with dyslexia have some additional sensory and executive challenges that interventions must address. My design requirements and rationale are based primarily on the interaction between the attentional challenges of children with dyslexia and the attentional demands that English imposes. The next section details the attentional challenges of children with dyslexia.

### **2.3. The Unique Visual and Executive Attentional Challenges of Children with Dyslexia and Implications for Interventions**

Children with dyslexia present impairments in phonological decoding (reading regular nonsense words, such as “sert”, via matching letters to phonemes) and orthographic or “sight” word decoding (memorizing and retrieving the pronunciations of irregular words, such as “yacht”) (Dehaene, 2009). Although some children present only one impairment, most present both (Manis, Seidenberg, Doi, McBride-Chang & Peterson, 1995). Similarly, drawing an analogy between sight-word reading and recognizing multi-letter units (Ehri, 2014), most words demand the flexible coordination of phonological and orthographic strategies (Casalis, Colé & Sopo, 2004; Ehri & Robbins, 1992; Dehaene, 2009). I have accordingly focused this section upon the visual and executive attentional challenges that might disrupt both phonological and orthographic decoding and spelling, where “orthographic” decoding and spelling includes the exploitation of frequently occurring stable multi-letter units and contexts that is necessary for English literacy.

My intervention targets two potential causes of the phonological and orthographic impairments of children with dyslexia: disturbances in visual processes involving the dorsal stream, a visual subsystem involved with focused attention, and disturbances in

executive functions involving phonology and perceptual inhibition and switching. The next two sub-sections describe these causes, their relevance to phonological and orthographic decoding, and the design requirements that they imply.

The impairments of dyslexia likely have multiple causes, and I do not claim that my intervention targets every potential cause. I focused on visual attentive and executive causes because they are consistently found in children with dyslexia, seem amenable to multimedia remediation, and have been extensively studied by perceptual researchers, whose insights I could leverage.

### **2.3.1. Impairments in Visual Processes Involving the Dorsal Stream**

Children with dyslexia present impairments in selective visual attentive processes involving the dorsal stream (Valdois, Bosse & Tainturier, 2004; Hansen, Stein, Order, Winter & Talcott, 2001; Vidyasagar & Pammer, 2009). These include: visual attention span (Bosse, Tainturier & Valdois, 2007), exogenous spatial orienting (Lallier, Tainturier, Dering, Donnadieu, Valdois & Thierry, 2010; Facoetti, Trussardi, Ruffino, Lorusso, Cattaneo, Galli, Molteni & Zorzi, 2009; Hari, Valta & Uutela, 1999) and serial visual search (Jones, Branigan & Kelly, 2007; Vidyasagar & Pammer, 1999). Visual attention deficits can occur independently of poor performance on phonological or phoneme awareness tests, but they predict the reading and spelling abilities of children with dyslexia, for words involving both letter-sound (phonological decoding) and orthographic (sight) decoding (Valdois, Bosse & Tainturier, 2004). This section explains the relevance of dorsal visual-attentional impairments to visual and auditory aspects of phonological and orthographic decoding and motivates interventions that compensate for it.

#### ***The Dorsal Stream***

The dorsal stream is a visual system located in the posterior parietal cortex that receives inputs from large-diameter occipital magnocellular neurons and communicates with the prefrontal cortex. The dorsal stream, also called the “where” pathway, processes visual-spatial location. The dorsal stream allocates visuo-spatial attention to

different spatial locations and determines the size of the visual-attentional window (Hansen, Stein, Order, Winter & Talcott, 2001).

### ***Visual Attention Span***

Visual-attention span is the number of visual elements that an individual can simultaneously process (Bosse, Tainturier & Valdois, 2007). Researchers measure visual attention span by exposing individuals to briefly presented arrays of visual elements. Individuals are subsequently asked to report the entire array (whole-report) or a specific element that is cued (partial report). Typically, the partial report conditions indicate that participants encoded a larger region of the display than what their whole-report performance suggests<sup>1</sup>. An advantage for partial to whole report performance is typical of individuals with normal visual attention spans. Individuals with dyslexia present equally poor whole and partial report performance, indicating a compromised visual attention span. Attention span deficits correlate pseudo (phonological decoding) and sight-word reading (Bosse, Tainturier & Valdois, 2007).

### ***Orienting and Focusing***

Attentional orienting is the ability to disengage attention from a given location and re-allocate it elsewhere. Exogenous orienting is the ability to do so on the basis of a peripheral perceptual cue, such as a flash or sound. Children with dyslexia present “sluggish” attentional orienting (Lallier, Tainturier, Dering, Donnadieu, Valdois & Thierry, 2010; Facoetti, Trussardi, Ruffino, Lorusso, Cattaneo, Galli, Molteni & Zorzi, 2009). They require a longer period between the onset of a peripheral cue and a target to present a “cueing” effect, (Facoetti, Paganoni, Turatto, Marzola & Mascetti, 2000), or to avoid attentional blink (a species of backwards masking, or inter-stimulus interference) (Hari, Valta & Uutela 1999), both of which indicate a need for more time to disengage and re-allocate attention. Individuals with dyslexia also have difficulty narrowing their attentional window. When a pre-cue indicates the size of an upcoming target, individuals without dyslexia show cueing effects even at long cue-target onset asynchronies, indicating an ability to sustain a narrowed attentional focus. Individuals with dyslexia only show the

---

<sup>1</sup> Any element that was probed could be reported, indicating that all had been attended; in the whole-report condition, because iconic memory decays before participants finish reporting all of the elements, participants cannot report every element.

cueing effect at shorter onset-asynchronies, indicating an inability to sustain a narrower attentional focus (Facoetti, Paganoni, Turatto, Marzola & Mascetti, 2000). There is some suggestion that individuals with dyslexia may have a bias towards a wider attentional focus, which is consistent with the observation of asymmetric processing favouring the right hemisphere (Facoetti, Turatto, Lorusso & Mascetti, 2001; Brosnan, Demetre, Hamill, Robson, Shepherd & Cody, 2002). Because the right hemisphere is involved in holistic processing, asymmetric right-hemispheric processing may make individuals with dyslexia more susceptible to interference from adjacent distracters (Roach & Hogben, 2007; Brosnan, Demetre, Hamill, Robson, Shepherd & Cody, 2002<sup>2</sup>). The dorsal stream is responsible for narrowing attentional focus, partly in concert with left hemispheric pre-frontal regions. The observation that dyslexic individuals must expend additional processing resources to narrow their attentional focus is consistent with Brosnan et al's (Brosnan, Demetre, Hamill, Robson, Shepherd & Cody, 2002) finding of a selective deficit in visual attentive processes to which the left prefrontal cortex contributes.

### ***Visual Search***

Visual search tasks measure an individual's ability to identify a visual target amongst distracters. Visual search can be parallel or serial, depending on the features that distinguish targets and distracters (Treisman & Gelade, 1980). Serial search speeds increase with the number of distracters (set size); parallel search does not. The dorsal stream plays a role in serial but not parallel search (Vidyasagar & Pammer, 1999). Serial search is serial because identifying the features that distinguish targets from distracters requires focused attention; consequently, individuals can only process one element at a time. The dorsal stream is responsible for shifting attention between visual elements (Vidyasagar & Pammer, 1999). Jones et al found that individuals with dyslexia were selectively impaired in a "cued" conjunction search, which requires the dorsal stream to shift attention to the cued location. Performance on the cued conjunction search correlated performance on a task of letter-position encoding, the ability to decode

---

<sup>2</sup> Although a wider attentional focus seems inconsistent with a reduced visual attention span, the former might account for the latter. Attention span indexes the number of visual elements an individual can simultaneously process, i.e., to differentiate from adjacent elements, and maintain in memory. A bias towards more holistic processing could reduce performance on assessments of visual attention span by preventing children from encoding a stable representation of any particular element; the adjacent elements would impede its encoding.

phonologically (i.e., to read regular pseudowords) (Jones, Branigan & Kelly, 2007). Similarly, Vidyasagar and Pammer (Vidyasagar & Pammer, 1999) found that dyslexics were poorer than non-dyslexics at a serial conjunction search (detect a blue T amongst red Ts and blue Ls), but not parallel search (detect a blue amongst red Ts).

### ***Relevance to Visual Segmentation of Letters***

Several theorists have articulated how dorsal stream impairments affect letter-sound decoding. The clearest way is by preventing children from visually segmenting letters from text. Lexical visual search (searching for a target amongst distracter letters) is serial, which suggests that processing letters involves the dorsal stream. Accordingly, children with dyslexia are impaired in lexical search (Casco, Tressoldi & Dellantonio, 1998). Reading and spelling require children to map sounds to letters. Mapping letters to sounds requires children to discriminate letters from their neighbors and shift attention between letters (Valdois, Bosse & Tainturier, 2004; Pammer & Vidyasagar, 2005; Vidyasagar & Pammer, 2010). Difficulties disengaging and re-allocating attention would slow decoding speed (Hari & Renvall, 2001). Difficulties narrowing the attentional window would allow adjacent letters to mask the focused letter, resulting in unstable perceptions of letters (Gori & Facoetti, 2015, Roach & Hogben, 2007).

Children cannot read what they cannot see. Interventions for children with dyslexia must make letters (or, more generally, whatever orthographic units are connected to sounds) less susceptible to masking and easier to focus to.

### ***Relevance to Auditory Segmentation of Phonemes***

Mapping letters to sounds requires children to segment individual sounds from speech and connect them to their letters. Children with dyslexia have difficulty segmenting phonemes from continuous speech (Richardson, Thomson, Scott & Goswami, 2004). Analogous to their difficulties in segmenting letters from continuous text, the poor auditory segmentation of dyslexia is supposed to attribute “fuzzier” perceptions of phonemes. That is, individuals with dyslexia present shallower increases in auditory discrimination sensitivity than do controls for pairs of frequencies that span a phoneme boundary, relative to frequencies that do not (Serniclaes, Ventura, Morais & Kolinsky, 2005).

Although the visual attention and phonological deficits are independent (Bosse, Tainturier & Valdois, 2007), some researchers have uncovered correlations between measures of visual attention and phoneme awareness (Vidyasagar & Pammer, 2010). Based on correlations between magnocellular functioning and phonological decoding (Cestnick & Coltheart, 1999), Vidyasagar and Pammer argued that the phonological core deficit can be reduced to a causal visuo-spatial deficit that itself originates in compromised magnocellular neurons (Pammer & Vidyasagar, 2005; Vidyasagar & Pammer, 2005). Although the finding of children who present independent visuo-spatial and phonological deficits but typical magnocellular processing (Wright, Conlon & Dyck, 2010) challenges Vidyasagar and Pammer's strong claim, there is reason to presume that compromised visual processing of letters, regardless of its cause, exacerbates poor phoneme perception.

The key reason is the finding that learning an alphabetic orthography (a system that maps phonemes to visual elements) improves phoneme and phonological awareness (Mann, 1986; Read, Yun-Fei, Hong-Yin, & Bao-Qing, 1986). The strongest demonstration was a study comparing the phonological abilities of Chinese adults who had and had not learned Pinyin. Pinyin is an alphabetic script that the Chinese government introduced to compensate for the difficulties children had in mastering traditional Chinese ideography. Crucially, the adults who did and did not learn Pinyin differed strictly in age. They were equated on IQ, SES and educational achievement. Adults who learned Pinyin showed better phoneme awareness and manipulative skills (e.g., deleting/substituting phonemes in syllables) than adults who did not learn Pinyin. The syllables and phonemes were common to Pinyin and traditional Chinese, so the difference was not explained by differential familiarity with the sounds. In addition, training the adults to read Pinyin produced immediate gains in their phonological abilities (Read, Yun-Fei, Hong-Yin, & Bao-Qing, 1986).

Such findings led theorists to surmise that learning an alphabet script helps children develop phonological awareness and categorical phoneme perception (Pammer & Vidyasagar, 2005). Converging evidence exists. Training to associate auditory features that are difficult to discriminate with visual features that are easy to discriminate

improves auditory discrimination (Kujala, Karma, Ceponiene, Belitz, Turkkila, Tervaniemi, & Näätänen, 2001).

An alleged cognitive mechanism is that visual stimuli provide information to attentional processes that segment the auditory stream. For example, because alphabetic orthographies match letters to phonemes, the number of letters in a word indicates how many unique phonemes the speech stream contains; words with common letters indicate which stimuli constitute common phonemes (e.g., “dog” and “get” share “g”; therefore, the sounds at the end of “dog” and beginning of “got” constitute the *same phoneme*).

There is also a candidate neurological mechanism. The auditory cortex contains multi-modal neurons that increase and decrease their firing rates to specific phonemes upon the presentation of congruent and incongruent letters, respectively (Van Atteveldt, Formisano, Goebel & Blomert, 2004). Accordingly, children with dyslexia present functional disturbances in discriminative regions of auditory cortex (Galaburda, Menard & Rosen, 1994), that audio-visual training could mitigate (Kujala, Karma, Ceponiene, Belitz, Turkkila, Tervaniemi, & Näätänen, 2001).

If matching letters to phonemes helps children develop phoneme awareness, then difficulty perceiving letters would impede phoneme awareness. The visual disturbances that dorsal impairments cause may therefore indirectly prevent children from developing phoneme awareness. Complementarily, interventions that compensate for dorsal visual impairments might improve phoneme awareness.

### ***Relevance to Orthographic Decoding and Spelling***

Children with dyslexia have difficulties segmenting words into multi-letter units that parallel their difficulties segmenting words into letters. Casalis (Casalis, Colé & Sopo, 2004) found that struggling readers were poorer at segmenting words into bases and affixes (e.g., “darkness” becomes “dark” and “ness”) than typical readers.

There are two ways that dorsal stream impairments could impact orthographic decoding: directly (via preventing children from visually noticing orthographic patterns) and indirectly (via impeding phonological decoding). I describe these in turn.

### **Direct (Visual)**

Encoding multi letter units (“dle”, “pre”) or contextual patterns (closed syllable predicts short vowel; /aw/ is spelled with “a” after “w”; “o” after “p”) requires children to encode, maintain and manipulate multiple letters simultaneously. Poorer visual attention span for complex stimuli such as letters could prevent children with dyslexia from encoding the letters that comprise units or patterns and learning their correlations with stable sound patterns (Gori & Facoetti, 2015; Bosse, Tainturier & Valdois, 2007). Likewise, learning contextual patterns (such as syllable types) that predict vowel sounds requires children to notice a visual pattern comprised of single letters (e.g., that every vowel-consonant substring has a short vowel sound). Difficulties focusing to stimuli distinguished by contour details (such as letters) could prevent children from extracting these higher-level patterns, similar to how they prevent children from extracting “embedded figures” from irrelevant visual context (Brosnan, Demetre, Hamill, Robson, Shepherd & Cody, 2002). Dorsal stream impairments that prevent children from encoding and maintaining multiple letters simultaneously would thereby directly impede orthographic decoding.

### **Indirect (Auditory)**

Several theorists argue that phonological skills are pre-requisites for orthographic decoding. Before children recognize a unit as a stable pattern, and decode or spell it “en masse”, children must experience several occasions of decoding it letter-by-letter, i.e., phonologically (Share, 1999). Likewise, even after they are memorized, using multi-letter units to decode words often requires phonological manipulation. Ehri (Ehri & Robbins, 1992) found that children’s ability to use “rime analogies” depended on their facility with phonological decoding. Similarly, Casalis et al (Casalis, Colé & Sopo, 2004) found that segmentations requiring a phonological change from the compounded word (an English example would be segmenting “pressed” into “press” and “ed”) were more difficult than those not involving a change (an English example would be “darkness” into “dark” and

“ness”). Dorsal stream impairments that impede phonological awareness could thereby indirectly impede orthographic decoding.

### **2.3.2. Impairments in Executive Functions That Are Relevant to Reading and Spelling**

Children with dyslexia present impairments in several measures of executive functioning. These are: Stroop colour/word naming (inhibition), rapid automatized switching (task switching) and embedded-figures detection (inhibition) (Berninger, Abbott, Thomson, Wagner, Swanson, Wijsman & Raskind, 2006; Brosnan, Demetre, Hamill, Robson, Shepherd & Cody, 2002). The embedded figures task requires children to inhibit an disruptive visual context (Brosnan, Demetre, Hamill, Robson, Shepherd & Cody, 2002); their poor performance is likely related to their left prefrontal and dorsal stream impairments (ibid.). In rapid automatized switching, children view sequences of letters, sight words or digits and report the name of the foveated symbol.

Both the Stroop colour-word naming and rapid automatized switching tasks require children to retrieve and report the name of a visual feature. In both tasks, at least one of the visual features is a letter-like symbol. Children with dyslexia who are not co-morbid for ADHD or autism present typical task-switching and inhibition when the visual features are not letter-like symbols and children respond via keypress (versus verbal report). For example, Stoet et al (Stoet, Markey & Lopez, 2007) found that children yielded similar switching-costs to those without dyslexia when pressing the left or right key on the basis of colour or shape. Poljac et al (Poljac, Simon, Ringlever, Kalcik, Groen, Buitelaar & Bekkering, 2010) found no additional impairment on the “go/no-go” task for autistic children who also had dyslexia. Similarly, individuals with dyslexia perform typically on tasks assessing executive functions involving strictly cognitive processes (e.g., planning and memory organization), versus those involving the discrimination or manipulation of phonological or visually complex stimuli (Brosnan, Demetre, Hamill, Robson, Shepherd & Cody, 2002).

It is possible that the difficulties that children with dyslexia have in Stroop and rapid automatized naming tasks attribute impairments in a subset of executive functions involving auditory working memory. Visual (visuo-spatial sketchpad) and auditory (the

phonological loop) working memory are two independent subsystems. The subsystems maintain visual or auditory information such that the individual can operate upon it mentally, for example, scanning visual memory for a cued item or reversing a sequence of digits (Berninger, Abbott, Thomson, Wagner, Swanson, Wijsman & Raskind, 2006).

Individuals with dyslexia perform poorly on assessments of auditory working memory, such as digit span (Berninger, Abbott, Thomson, Wagner, Swanson, Wijsman & Raskind, 2006; Brosnan, Demetre, Hamill, Robson, Shepherd & Cody, 2002). Conversely, Brosnan et al (Brosnan, Demetre, Hamill, Robson, Shepherd & Cody, 2002) found that individuals with dyslexia perform typically on assessments of visual working memory, such as memorizing the order of visual events and global pattern recognition. In Brosnan et al's studies, the visual stimuli were coarse shapes (squares, circles) and colours. Although their findings seem inconsistent with the reduced visual attention span of individuals with dyslexia (Bosse, Tainturier & Valdois, 2007), the discrepancy could indicate that the visual attention span deficit is specific to complex contour stimuli, such as letters; if assessed with colours and shapes, visual attention span might be typical. Individuals with dyslexia may also be biased towards visual encoding strategies, even if they have difficulty visually encoding the stimuli. Miller and Kupfermann (Miller & Kupfermann, 2009) found that individuals with dyslexia were more disrupted by visually than phonologically similar distracters on a word memorization task, indicating a bias towards visual (i.e., memorizing the word's global visual features, engaging visual working memory) versus verbal encoding strategies (verbal rehearsal, engaging auditory working memory), which is opposite to individuals without dyslexia.

In addition to lower-level symbol sound correspondences, spelling and reading in English require children to retrieve and apply abstract rules. In some cases, understanding which rule to apply requires children to act upon information in working memory. For example, determining which letter to use in spelling the /k/ sound requires children to consider the position of the letter and the letters that precede it, and then to retrieve the appropriate response, i.e., spelling decision. The default channels for such information are the forms (and corresponding positions) of letters and phonemes, which are difficult for children to, respectively, attend to (section 2.3.1) and mentally manipulate. Wagner and Torgesen (Wagner & Torgesen, 1987) suggest that children's

difficulty maintaining information in verbal working memory impedes spelling and decoding by consuming cognitive resources that would otherwise be directed towards retrieving and applying the relevant rules. In other words, some of the difficulty children with dyslexia have in reading and spelling English words might attribute breakdowns in the links between visual or auditory working memory representations and cognitive reading and spelling rules.

The typical performance of children on tasks requiring them to retrieve alternate rulesets on the basis of coarse visual features, such as colour or shape (Stoet, Markey & Lopez, 2007), combined with their preference for visual encoding (Miller & Kupfermann, 2009), suggests that they might have an easier time retrieving spelling and reading decisions (i.e., task switching) if a more easily processed visual feature encoded the pertinent linguistic stimuli (letters or sounds). Similarly, because children's ability to inhibit irrelevant visual information appears related to dorsal stream impairments (Hansen, Stein, Order, Winter & Talcott, 2001; Brosnan, Demetre, Hamill, Robson, Shepherd & Cody, 2002), encoding information in visual dimensions that the dorsal stream does not process could support children in inhibiting irrelevant information.

### **2.3.3. Implications for Interventions and Design Requirements**

To summarize, there are 3 core challenges that children with dyslexia face which are relevant to literacy and which my intervention targets. I translate the challenges into design requirements:

**Challenge 1:** poor ability to segment visual letters and identify important multi-letter units stemming from compromised dorsal visual systems.

**Requirement 1:** leverage visual systems that are not compromised to develop visual supports for children to discriminate letters or letter-units and their corresponding sounds or sound-units.

**Challenge 2:** poor ability to mentally manipulate letters and sounds, which overloads working memory resources and impedes other decoding sub-processes

**Requirement 2:** leverage visual cues that children have an easier time mentally manipulating to free up working memory resources for rule retrieval and application

**Challenge 3:** poor ability to inhibit attention to irrelevant stimuli and to quickly shift attention between relevant stimuli

**Requirement 3:** leverage visual cues that are easy to focus to and that are robust to interference from spatially and temporal contiguous distracters

My requirements succeed decades of interventions along the guiding framework of *explicating the units and strategies that are statistically predictive of English pronunciations*. My requirements build on previous work by more specifically identifying colour-- a visual feature that is not compromised by the dorsal system, which is robust to lateral and backwards masking, and which might be easier than letters or phonemes for children to mentally manipulate-- as a candidate feature for explicating English multi-letter units and contexts.

In the next section, I review some large-scale interventions for children with dyslexia. My goal is understanding the principles behind their techniques and applying my analysis of the attentional challenges of dyslexia and attentional demands of reading to suggest some ways that contemporary technology- more specifically, a software implementation- could use colour to enhance them.

## **2.4. Interventions for Struggling Readers**

This review has three sub-sections. Sub-section 1 reviews some general principles and specific intervention strategies that were empirically validated. Sub-section 2 describes the techniques of mainstream multisensory Orton-Gillingham (OG) interventions. OG programs, which are the forerunning “evidence-based” curricular intervention (Alexander & Slinger-Constant, 2004), synthesize the empirically validated principles and techniques that I describe in sub-section 1. Section 3 describes a class of interventionist strategies that use colour-coding. These have received little direct empirical support, but can be rationalized in terms of colour’s effects upon attention and the attentional demands of reading and spelling in English. My core theoretical contribution is synthesizing empirically validated interventionist principles and strategies

with the intuitions behind the use of colour-coding into a guiding framework of using colour to support attention to relevant multi-letter units and patterns.

#### **2.4.1. Mainstream Interventions: the need for Explicit Instruction in Phonics and Multi-Letter Units and Contexts**

Alexander and Slinger-Constant reviewed the status of current interventions for children with dyslexia and summarized several experiments that compared different approaches (Alexander & Slinger-Constant, 2004). The success of an approach depended on two factors: first, what level of orthographic unit was the intervention's focus? Second, was instruction *implicit* or *explicit*? Successful interventions, i.e., those which yielded the greatest and more durable improvements to children's reading and spelling, focused on multiple levels of orthographic unit and were explicit. I review these studies here. The goal of my review is identifying some general design features for integration into my framework.

##### ***Lindamood Phonics and Variants (Earobics, FastForWord)***

Lindamood phonics sequencing programs focus on phonemes. They target phoneme awareness and phonological manipulation. Lindamood programs are multisensory. Lindamood programs and variants compensate for children's unstable representations of auditory phonemes by representing phonemes in alternate ways. One way that is relevant to my intervention was representing each phoneme in a word as a uniquely coloured block. For example, the words "dog" and "dot" might correspond to two rows of blocks, the first consisting of a red, yellow and green block; the second, a red, yellow and a blue block. The same supports that underlie the contribution of learning an alphabetic script to phoneme awareness apply here: the coloured blocks inform children of the number and identities of phonemes in words. Coloured blocks support children with dyslexia because the boundaries between them are easy to perceive, through vision (colour perception is typical) and touch (Birsh, 2011).

There are some limitations to Lindamood programs. Lindamood-styled programs (including FastForWord and Earobics) do not explicate the connection of phonemes or sounds to orthography. Pokorni (Pokorni, Worthington & Jamison, 2004) found that while

Lindamood and FastForWord improved children's ability to segment and manipulate phonemes (i.e., they could perform tasks such as "replacing" the first sound of *pig* with the last sound of *dot*), the children became no better at reading. This is consistent with the idea that children require attentional supporting to *connect* sounds to orthography, not simply to segment sounds from speech.

### ***Integrated Picture Mnemonics***

Training children to learn letter-names improves their reading (Ritchey & Speece, 2006). Because learning a letter's name is a similar task to learning the sound a letter represents, interventions that increase the speed with which children can access letter-names might yield insights for how to strengthen the connections between letters and sounds.

Ehri (Ehri, Deffner & Wilce, 1984) used a tactic called "Integrated picture mnemonics" to help children learn letter-names. Ehri assessed 6 letters ("m", "a", "f", "b", "t", "i"). Each letter was assigned to an object whose name began with the sound that the letter represented, e.g., a "mountain" for "m", a "flower" for "f". Ehri created two sets of picture flashcards. In the "integrated" set, the objects were drawn such that their shapes resembled the shapes of their letters (e.g., the object for "m" was "mountain"; in the first set, the mountain in the drawing was a double-peaked mountain, shaped like an "m"). In the "unintegrated" set, the shapes of the objects and the letters were mismatched (a single-peaked mountain). Children who were given the integrated set learnt the letter-names faster than children given the unintegrated set.

Mnemonics work because they associate each of two stimuli that are hard to associate with another stimulus that is easy to associate. Thereafter, learners can "bridge" between the stimuli that were difficult to associate via the stimulus that is easy to associate. Because the connections between letters and sounds or names are arbitrary, they are difficult to associate (Ehri, Deffner & Wilce, 1984; Birsh, 2011). Ehri's integrated pictures sought to bridge letter forms and names by a third stimulus- an object- that was intrinsically connected to both.

The absence of an effect for unintegrated pictures illuminates another requirement: the connections between the third stimulus and the visual letter must be direct, or visually apparent. The unintegrated mnemonics obliged children to forge a strictly cognitive link between the letter (e.g., “m”) and object on the basis of the letters in the object’s name. Wagner and Torgesen pointed out the ineffectiveness of mnemonics requiring such strictly cognitive steps. Cognitive operations such as retrieving the orthographic name of an object, looking up the first letter, retrieving the sound, looking up the first sound, consume valuable working memory resources (Wagner & Torgesen, 1987). The consumption of working memory resources questions the mnemonics’ purpose, given that mnemonics are supposed to *automate* retrieval and thus *conserve* working memory resources.

Ehri’s integrated picture mnemonics suffer this limitation as well. To use them, children must a) retrieve the object’s name b) segment the first sound c) retrieve the written name and d) match the first sound to the first letter. Ehri’s experiment only assessed children’s abilities to retrieve names of letters shown in isolation, and the children were not dyslexic. Consequently, the demands on children’s working memories were low. It seems unlikely that children would be much assisted by these mnemonics when reading continuous text, where the demands on working memory are high, or when children are dyslexic, and the capacity of working memory is low.

What is needed is a mediating third-feature that requires fewer resources to process or represent than Ehri’s objects, or that can be physically projected on the letters and therefore does not require mental representation. In later sections I argue that colour could play this role.

### ***RAVE-O and Other Multi-Letter Unit Approaches***

Another limitation of integrated-picture mnemonics, which use objects to bridge between letters and sounds, is teaching the relations between multi-letter units and sounds. As I explained in section 2.2, explicit instruction in multi-letter units is needed to support Anglophone children in achieving literacy at an acceptable rate. This section provides further evidence for my assertion and identifies some strategies for explicating

multi-letter units via describing some experimental approaches to teaching multi-letter unit strategies and their successes over and above single letter strategies.

Wolf et al's (Wolf, Miller & Donnelly, 2000) RAVE-O program (Retrieval, Automaticity, Vocabulary-Elaboration, Orthography) is a supplement to basic phonics that teaches children to automatically recognize and use high-frequency multi-letter patterns. RAVE-O, like Berninger's experimental approach, uses visual supports (such as colour-coding units of words that are the focus of attention, e.g., "tion" in "caption"). Wolf compared children's reading performance after experiencing a Lindamood style analytic-phonics with letters intervention with and without additional RAVE-O; children who experienced RAVE-O were faster and more accurate readers and spellers.

RAVE-O targets all (affixes, rimes, syllables) multi-letter units. Researchers exploring specific multi-letter units have also yielded gains over traditional phonics. Nunes, Bryant and Bindman (Nunes, Bryant & Bindman, 1997) taught British children to memorize and read and spell via detecting affixes. Nunes et al used various visual supports (boxes, colour-highlighting) to explicate affixes in words and told children to treat them like "single letters". Crucially, Nunes et al's intervention targeted children below the age at which children spontaneously demonstrate affix-based strategies. Nunes et al's intervention enabled children to use affix-based decoding. These results suggested that incorporating multi-letter units in early literacy instruction might help Anglophone children become literate at comparable rates to peers learning transparent languages.

Berninger et al (Berninger, Abbott, Brooksher, Lemos, Ogier, Zook & Mostafapour, 2000) applied their connectionist framework to develop a three-layered intervention aimed at helping children recognize predictable multi-letter-sound combinations. Each layer explicated the connections between a different 'grain' of letter-unit and sound. In first layer training, children memorized letter-sound correspondences out of the context of words. In second layer training, children rehearsed them within the context of words. In this case, Berninger exposed children to flashcards wherein the relevant letter-units (single letters and digraphs or rimes and onsets) were each differently coloured. In the third layer, children consolidated their knowledge by reading

connected text. Relative to children who received analytic phonics (Lindamood styled) and orthographic awareness training (tutors named a letter and children reported the letter that preceded or succeeded it), the children who received explicit training in the connections between phonics and orthography performed better on post-tests of reading and spelling, even for words not encountered in instruction.

Cunningham (Cunningham, 1998) and Archer et al (Archer, Gleason & Vachon, 2003) advocate syllable-based strategies. Like affix-based strategies, syllable-based strategies involve training children to memorize some high-frequency units (stable syllables, such as consonant-le), however, they include a more generalizable algorithm of partitioning words around vowels, and assigning consonants by seeking out consonant blends.

Bhattacharya and Ehri (Bhattacharya & Ehri, 2004) assessed a “flexible” syllable analysis approach in which they deferred teaching children “dictionary rules” for syllable division and instead taught children only the rule that each syllable has “exactly one vowel”. Children were repeatedly exposed to a set of 100 words, which they syllabicated by flexibly assigning consonants to vowels, and then physically mapped to the syllable sounds by pronouncing each syllable whilst covering the syllables that were not being pronounced. Relative to children who read the whole word without analyzing them into syllables, children who received syllable-awareness training read instructed words and uninstructed words containing instructed syllables more accurately, and remembered how to spell words they had previously seen. Berninger et al (Berninger, Vaughn, Abbott, Brooks, Begayis, Curtin & Grahm, 2000) also explored whether teaching children to categorize words into syllable types benefitted them over and above training to recognize letter-sound, onset-rime and whole word units. Their results yielded little effect of categorizing words into syllable types, which is consistent with Archer (Archer, Gleason & Vachon, 2003) and Bhattacharya’s (Bhattacharya & Ehri, 2004) recommendations to supplement “dictionary” syllable knowledge with more actionable practice in using syllabication to memorize and identify frequently occurring syllable types.

## **Summary**

Successful approaches to help children read involve explicating the connections between single and multi-letter units and sounds and providing students large amounts of supported practice. Visual cues are one way of explicating multi-letter units that help children match them to sounds.

Although these principles apply to children with and without dyslexia, interventions for children with dyslexia may require additional or more powerful attentional supports. The next subsection reviews Orton-Gillingham multisensory curricula, which extend the principles and tactics of general literacy interventions to satisfy the unique requirements of children with dyslexia.

### **2.4.2. Orton-Gillingham Multisensory Curricula**

Orton-Gillingham (OG) multisensory programs are structured, sequential multisensory interventions for children with reading and spelling difficulties (Gillingham & Stillman, 1946; Henry, 1998; Cox, 1974; Cox, 1985; Birsh, 2011). Our software system was based upon similar theoretical principles as OG curricula, namely, that alternate sensory channels (touch) and features (colour) could be recruited to better explicate linguistic rules. Accordingly, I based much of the design of the colour codes and intervention on OG curricular activities. Another reason why we founded our design on OG programs is that they are the only current, commercially available intervention that explicitly draws upon empirical research (Alexander & Slinger-Constant, 2004). OG interventions have also received some empirical validation (Oakland, Black, Stanford, Nussbaum & Balise, 1998; Ritchey & Goeke, 2006). In addition to seeming theoretically sound, OG-styled programs are mainstream (Alexander & Slinger-Constant, 2004; Birsh, 2011). We therefore expected that using OG principles to design our software, which would ensure its compatibility with OG school curricula, would increase our software's user base.

Although OG programs are relatively successful, many children, even after several years of instruction, fail to reach even the 30<sup>th</sup> percentile of reading and spelling achievement (Alexander & Slinger-Constant, 2004). Researchers and educators have

identified *poor attention* as the main impediment to children's profiting from intervention (Alexander & Slinger-Constant, 2004). Accordingly, my main design goal was increasing the effectiveness with which Orton-Gillingham techniques capture and direct children's attention towards learning and transferring letter-sound relations.

The software that I and my colleagues designed, which I use as a research instrument, implements Orton-Gillingham strategies. Consequently, I designed my colour-codes to support their objectives. To justify my use of colour and my specific colour coding schemes, I here describe the two Orton-Gillingham styled approaches that I considered in design: guided discovery and multisensory engagement. Both support children in overcoming the attentional challenges that prevent them from acquiring and transferring letter-sound relations. I describe some limitations in the current paper and pencil implementations of guided discovery and multisensory engagement, and I conclude this section by describing how software could help.

### ***Guided Discovery***

Guided Discovery is an instructional technique of constructivist learning, which Mayer (2004) summarized: constructivist theorists maintain that "meaningful learning occurs when learners actively engage in selecting, organizing and integrating" new with outstanding knowledge about some domain, with the goal of learning the domain's premises, generalizations and transferable skills (for example, solving conservation or logic problems). Constructivism contrasts passive approaches wherein teachers, books or other media provide children these premises, generalizations and procedures. Extreme interpretations of constructivism advocate pure discovery learning- that learners should receive no explicit guidance in their exploration of a learning domain. Pure discovery learning approaches have received scrutiny for their failure to yield transferable gains (Mayer, 2004). Guided discovery is a more moderate interpretation. Guided discovery maintains the constructivist emphasis on the child's active engagement with and elaboration of principles about a learning domain, but differs from pure discovery learning by advocating some instruction (or attentional guidance) to explicate the learning goal (e.g., figure out a generalizable strategy for solving 'odd one out' type word problems) and support learners in selecting domain information that is relevant to the learning goal. Mayer (Mayer, 2004) reviewed three classes of

experiments suggesting that guided discovery was superior to both pure discovery and expository (wherein teachers tell students the solutions to problems) learning in supporting children's retention and transference of learning material. Mayer (ibid) summarized the review as an endorsement of the guided discovery approach to helping children construct knowledge about various domains, and emphasized that the key question for educators- and the designers of interactive experiences- is how, and by which kinds of cues, to guide children to notice the domain's relevant information and to explicate the learning goals? Mayer also points out that- for meaningful learning to result from a guided discovery experience- learners must be intrinsically motivated to generate the desired knowledge representations. Noticing, organizing and integrating novel and outstanding information involves considerable cognitive effort, much more than “passive” learning. If students are unwilling to expend the requisite cognitive effort, guided discovery may be no more effective than expository or pure discovery approaches.

Orton-Gillingham multisensory curricula apply constructivist guided discovery to help children appreciate English spelling and pronunciation rules. Tutors immerse children in examples of words that involve a linguistic rule. Tutors guide children to 'discover' the rule by drawing their attention to relevant linguistic information. For example, to guide a child to discover the rule, “vowels in vowel-consonant-e syllables sound long”, the tutor would expose children to a list of words, each of which involves the vowel-consonant-e pattern, and each of which contained a long vowel. (For example, “fade”, “mode”, “fine”). The tutor would read each word and emphasize the sound qualities that are similar, in this case, the long vowel. The tutor would prompt the child to identify “what letters and sounds” the examples have in common. After the child identifies the pattern the tutor guides them to articulate a general rule (“when there is an e on the end of a vowel-consonant syllable, the vowel says its name”). Tutors guide children to notice the relevant linguistic variables (short versus long vowel; presence/absence of silent e) by visually annotating the words. Tutors mark “short” vowels by a breve; long vowels by a macron, and silent letters by a slash. These marks play a dual role of identifying important letters (vowels and the silent e), and visually representing an acoustic contrast (short versus long versus silent).

There are some limitations to this approach. Guided discovery is effective to the extent that learners notice and understand the relevance of the attentional guides, such that learners “activate” the necessary background information (in this case, the sounds of the vowels). Annotations may be no easier for children to notice and focus on than the letter forms themselves; both would involve the compromised posterior-parietal cortex. Based on Ehri’s (Ehri, Deffner & Wilce, 1984) findings for integrated versus disassociated picture mnemonics, marks such as “breves” and “macrons”- which bear little natural connection to the sound categories- long and short- would be ineffective visual mnemonics. If anything, children might have a difficult enough time remembering what sounds breves and macrons corresponded to that using them would place an additional tax on their limited working memories.

Paper and pencil has additional limitations as a medium for guided discovery of linguistic rules. Linguistic rules involve relationships between sounds and orthography. Children must therefore understand the impact of certain orthographic differences (e.g., “fad” versus “fade”). One tenet of constructivism is that active engagement and experimentation with the domain variables supports learners in comprehending the relationships between domain variables (Mayer, 2004). Applied to this context, guided discovery would involve children’s active modification of the words and apprehension of the changes (in sound) that result. To guide children to apprehend the changes in sound ensuing from, for example, removing the “e” from “fade”, tutors would have to physically alter the markings on the words. The window for inter-stimulus binding (i.e. between an orthographic and a phonological change) is short (Powers, Hillock & Wallace, 2009); tutors might not mark the change quickly enough for children to develop lower-level connections between the orthographic and phonological changes. Likewise, children would have to “hold” the change (which involves remembering how the word appeared before the change) in memory until tutors mark the new sound. Assuming that knowledge construction requires considerable cognitive effort (Mayer, 2004), the tax on children’s working memories might prevent them from realizing the orthographic relationship. In a software implementation, conversely, such changes could be immediate. Learning could leverage lower-level binding mechanisms, and children would be relieved of the need to hold word-forms in working memory.

## ***Multisensory Engagement***

The second strategy by which tutors support children with dyslexia in acquiring and transferring letter-sound relations is multisensory engagement. Orton and Gillingham focused on the visual, auditory and kinesthetic modalities. Children *heard* a letter-unit's pronunciation whilst *seeing* the letters and physically tracing them. The argument was: each letter-sound unit is a binding of visual (letter-form), auditory (sound) and kinesthetic (how to write it) representations; much of children's reading and spelling difficulties attributes breakdowns between these representations; experiencing the representations as a single connected event would help children mentally bind them. Some of these intuitions have been borne out experimentally (Hulme, Monk, & Ives, 1987; Fredembach, Boisferon & Gentaz, 2009).

Dual Coding Theory generalizes the intuition: learners may understand and remember information better if multiple modalities represent it: the more representations that encode the event, the more pathways children have to recall it (Clark & Paivo, 1991; Sadoski & Paivio, 2004). The insight of dual-coding theory is that the representations need not be canonical. OG methods focus on canonical representations (letter forms, sounds and gestures for writing). An example of a non-canonical representation would be a colour or shape that correlates a relevant property, such as the letter's category (vowel or consonant), position or sounds. Designers could represent letter properties with non-canonical representations that are easier for children with dyslexia to process. Kast et al (Kast, Meyer, Vogeli, Gross & Jancke, 2007; Kast, Baschera, Gross, Jancke & Meyer, 2010) articulated these ideas in the rationale for their software system, DyBuster (described in section 2.4.3.2): software mediums would enable designers to use a variety of representations and adjust them dynamically, as children change the letters and sounds of words. Children could therefore recall the connections between letters and sounds via visual representations (colour, shape) that are easier for them to encode.

To summarize, contemporary multi-sensory approaches use various techniques to support children with dyslexia in acquiring and transferring letter-sound relations. Each strategy supports children's attention. There might be room for improvement. Attentional challenges remain the core obstacle to children's succeeding through multi-

sensory instruction, but an analysis of their strategies reveals some ways that alternate features and software could help.

Software designers have recognized the applicability of dynamic multimedia learning environments to remedial literacy training (for a review, see Blok, Oostdam, Otter & Overmaat, 2002). Consistent with my suggestions, most programs use the software medium to provide children practice in learning letter-sound correspondences or analytic phonics using various visual and auditory cues, for example, Magnan et al (Magnan & Ecalle, 2006; Magnan, Ecalle, Veuillet & Collet, 2004) developed an application wherein learners match grapho-syllable units (“pa”/”ba”) to sounds (/ba/ or /pa/) via keypress.

To date, however, most applications have focused on learning simple letter or multi-letter unit and sound combinations, or analytic phonics or alphabetic knowledge. Applications do not focus on learning generalizable strategies or rules (such as syllable types, syllable division or context-sensitive spelling rules) that impede older learners (Cunningham 1998; Archer, Gleason & Vachon, 2003). OG programs focus on these rules as well, using the same principles of Guided Discovery and Multisensory engagement. As with simple letter sound correspondences, limitations of visual features that OG approaches currently use, and of the paper and pencil medium, open a potential role for alternate visual supports and software implementation.

### **2.4.3. Colour-Coding and Literacy**

Throughout my review I have alluded to experimental interventions that used colour as an attentional support, typically for highlighting important letter units, and I have spoken of colour as a candidate substitute for visual contour in conveying the sounds that certain letters have. Here, I develop these ideas more thoroughly.

This section advances colour as a means of supporting children’s acquisition of letter-sound correspondences. I argue that colour could be uniquely effective when implemented in software. Articulating a unique role for software is an important design goal: children with dyslexia are more motivated to use software that their teachers and

parents endorse (Mioduser, Tur-Kaspa & Leitner, 2000); without a clear sense of a software application's benefits, tutors are reluctant use it.

This section has three subsections. First, I articulate a general rationale for colour that is based in colour's attentional effects and its uses in non-literacy learning and analysis contexts. Second, I narrow my rationale by describing the typical colour perception of children with dyslexia. Third, I discuss previous attempts to use colour in literacy acquisition.

### ***A Rationale for Colour-Coding***

This subsection describes some general attentional properties of colour and some studies that have exploited these properties in non-literacy instructional contexts. Colour affects attention, memory and information integration in ways that might support learning, remembering and transferring information. I indicate in parentheses the design requirement (2.3.4) that each colour property satisfies.

#### **Attention**

Colour contrasts are salient and processed pre-attentively (i.e., without focused attention) (**R1**). This is why search for coloured-targets yields parallel slopes. Serial slopes (time to detect a target increases with the number of distracters) occur because focused attention to an item is needed to process it. Parallel slopes indicate that serial attention is unnecessary (Treisman, 1982). Serial search involves the dorsal stream; parallel search does not (Vidyasagar & Pammer, 1999).

Colour targets are easier to focus to than targets defined by shape. This observation has particular relevance for the executive functions that predict literacy skill. Banich et al (Banich, Milham, Atchley, Webb, Wszalek & Brown, 2000) compared the "switching costs" for re-focusing to words, objects or their colours. Although switching always exerts a cost, switching to a colour causes less slowdown than switching to a word or an object (**R3**).

Stimuli that are coloured differently also produce less inter-stimulus interference than stimuli that are coloured the same. Visual crowding (adjacent distracters prevent

participants from identifying a foveal target) is especially strong in individuals with dyslexia and seems to mimic natural reading (Gori & Facoetti, 2015); colour contrasts reduce inter-stimulus crowding (Poder, 2007). Similarly, backwards masking (distracters that follow a target in time prevent participants from identifying the target) contributes to the attentional blink, which again mimics natural reading, and to which individuals with dyslexia are more susceptible (Hari & Renvall, 2001); colour contrasts reduce backwards masking (Gellatly, Pilling, Cole & Skarratt, 2006).

Zentall et al (Zentall, Grskovic, Javorsky, & Hall, 2000) exploited some of these properties to support children with ADHD in focusing on and learning informative text. Children read one of two documents. In the first document, text was uncoloured. In the second document, some lines at the end of the document were highlighted in contrasting colours. Children recalled more information from the colour-highlighted than the colourless document.

## **Memory**

Colour plays a role in memory and learning. Colours that are strongly associated with objects (e.g., bananas are yellow) speed individuals' ability to recognize those objects, i.e., to access their names or to remember if the object was previously observed (Hanna & Remington, 1996) (R2).

## **Information Integration**

Colour is a powerful cue for “perceptual grouping”: seeing multiple visual elements as part of the same unit or category (Treisman, 1982). Although grouping is a perceptual effect, grouping can have cognitive consequences too (Christ, 1975; Ware, 2012). Ozcelik et al (Ozcelik, Karakus, Kursun, & Cagiltay, 2009) explored how common colours that support perceptual grouping might also support *cognitive* integration. Ozcelik et al exposed college students to two different learning displays. Both displays required students to integrate information conveyed by a picture with information conveyed by text (for example, a diagram of an axon potential and a paragraph describing an axon potential). Text was colour-highlighted; pictures had coloured borders. In one condition, each text-picture pair used a unique common colour; in the other, all text-picture pairs used the same colour (i.e., colour did or did not “code” a text-

picture pairing). Students who studied with the colour-coded displays performed better on recall and comprehension tests than students who studied uniformly coloured displays. The comprehension questions required students to combine information that was unique to the text or picture of a pair and therefore tested successful integration.

Similarly, information visualization designers exploit colour contrasts and colour-based grouping to support analysts in detecting correlations between categorical and other kinds of information. (For example, a visualization showing the distribution of species in Canadian forests might code species with colour and geographical location with spatial location) (Christ, 1975). Information Visualization researchers have assessed the usefulness of each visual dimension (colour, size, saturation), for conveying different types of information (categorical, spatial, continuous). Users consistently perform best on analytical tasks involving categorical information when colour contrasts code them (Christ, 1975; Ware 2012).

## **The Potential Application to Literacy Interventions**

### **Attention**

Poor attention, which presents as visual, executive and associative disturbances, is the main impediment to children's benefitting from OG literacy interventions. Colouring relevant multi-letter units might help children to focus on them and ignore irrelevant information, which could be uncoloured:

Cake mistake fake

drink crack

**Figure 2-1. How colour might draw attention to the rime unit (top), or the presence and positional restrictions on front and end blends (bottom).**

## Memory

Similar temporal associative regions may be involved in binding visual objects and letters to colours (Wolf, Bally & Morris, 1986). Repeatedly pairing letter-sound units with particular colours could develop strong associations between them and the colours (Colizoli, Murre & Rouw, 2012). Subsequently, the colours might help children recall additional information (positional restrictions, sounds, etc.) that were associated with the letter-sound unit.

## Information Integration

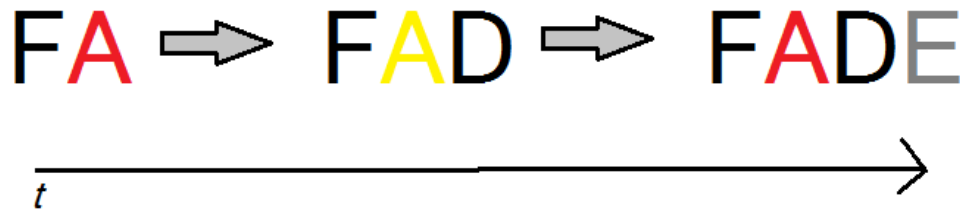
Because letter sounds are categorical, colour could effectively code them. Letter sounds correlate other information, such as the letters' positions in a word, or the letters that surround it. Because these are also visual properties, colour-coding sound could enable children to learn their correlations with sound. Children might have an easier time appreciating correlations between letter-unit and position (for example, the stable syllable “consonant-le” only occurs *at the end of a word*) if the units were uniquely coloured, so that that colour only ever appeared at the end of a word. Likewise, colour-coding categories (such as consonant and short vowel) the same might help children appreciate that, for example, the words “cat”, “pop”, “din” are all instances of the same category (closed syllable), and could be decoded by applying a general rule (vowels in closed syllables sound short):

stable muffle fled blemish  
hug huge bit bite fad fade

**Figure 2-2.** How colour might support learners in noticing the rule that consonant-le syllables only appear at the end of the word (top), or support them in “grouping” (noticing the relationship between) all closed and all vowel-consonant-e syllables, respectively.

## A Unique Role for Software

Colour-coding introduces a unique role for software. Software mediums can change letters' colours. Colour changes can quickly re-focus attention and executive functions can more readily re-focus to colour than other visual targets. In section 2.4.2 I argued that software mediums could improve upon paper and pencil by providing children dynamic visual representations of phonological variables, enabling them to manipulate a word's orthography and "see" the changes in sound. Software thus has the potential to leverage the attentional affordances of colour changes. In decoding words involving large-unit and phonological strategies, colour changes could cue children's attention dynamically to different units of the word as per the decoding step (i.e., syllabification or retrieving letter-sound correspondences). In modifying words in ways that modify their sounds (e.g. adding "e" to "fad"), colours that represent vowel sound category would immediately change, helping children appreciate the correspondence between orthography and phonology:



**Figure 2-3.** A software application that enables children to spell words could leverage dynamic colours to draw children's attention to the changes in sound that correlate key changes the child makes to the word's letters.

But is there any reason to suppose that children with dyslexia would benefit from colour, over and above their proficiency with shape? Reader, there is.

### *The Typical Colour Perception of Children With Dyslexia*

Children with dyslexia present dorsal stream impairments. Some children present additional impairments in the magnocellular visual system (Stein & Walsch,

1997). Neither of these systems processes colour. Colour is detected by parvocellular neurons (Smith & Pokorny, 1975) and processed further by the temporal-associative ventral stream, which is independent of the dorsal stream (Shmuelof & Zohary, 2005). Children with dyslexia have typical colour perception (Dautrich, 1993). Colour perception in their visual peripheries may be superior to that of controls (ibid.), suggesting that peripheral colour changes may be a useful attentional cue.

Consistent with the anatomical typicality of colour-vision systems in dyslexia, cognitive and perceptual colour functionality are typical in dyslexia too. Children with dyslexia show typical parallel slopes when searching for coloured targets (Vidyasagar & Pammer, 1999). This implies that children with dyslexia can i) rapidly focus to colour and ii) ignore 'distractions' that are distinguished by an irrelevant colour (**R1**). Brosnan et al (Brosnan, Demetre, Hamill, Robson, Shepherd & Cody, 2002) exposed children with and without dyslexia to a task that required them to view global patterns composed of differently coloured blocks and to determine which of an array of subsequently presented patterns matched the one they had previously seen. In many cases, the target pattern differed from distracters in the location (versus presence) of a specific coloured element. Individuals with dyslexia performed the same or better than controls. This suggests that individuals with dyslexia can i) quickly encode patterns composed of variously coloured elements ii) recognize differences in the presence or location of a coloured element. Such capacities contrast the impairments that Berninger (Berninger, Abbott, Thomson, Wagner, Swanson, Wijsman & Raskind, 2006) and Battacharyna et al (Bhattacharya & Ehri, 2004) found for the encoding and detection of single letter-changes to whole word orthographic patterns, and the impairments that Bosse et al (Bosse, Tainturier & Valdois, 2007) uncovered for encoding the identities and positions of arrays of letters. Neither the impaired visual attention span nor impaired spatial attention seem to impact global colour-pattern perception as much as they impact letter and letter-pattern (i.e., word) perception (**R1, R2, R3**).

Typical global colour-pattern perception could be leveraged to help children notice relevant orthographic patterns and changes to those patterns (i.e., "fad" versus "fade") that correlate changes in sound. Typical colour perception means that all typical attentional benefits of colour (attentional capture, discrimination and prevention of

interference) would apply to children with dyslexia. Typical visual working memories (Brosnan, Demetre, Hamill, Robson, Shepherd & Cody, 2002) and a preference for visual encoding (Miller & Kupfermann, 2009)- combined with the usefulness of colour as a memory cue (Hanna & Remington, 1996)- open the possibility of using colour to teach and support children in retrieving cognitive decoding and spelling rules **(R2)**.

Researchers are aware of such possibilities and have explored them, but a principled framework for understanding colour has yet to develop. The next section describes some attempts to use colour to mitigate the higher-level attentional and cognitive challenges of dyslexia.

### ***Attempts to Use Colour In Literacy Acquisition***

The next few sub-sub-sections describe the major attempts to use colour to support children with dyslexia in attentional and cognitive aspects of literacy. Although these interventions are promising, they suffer the absence of a guiding framework for designing or assessing different colour-coding schemes.

My review motivates my research goal of developing a more principled understanding of *what* specific aspects of literacy acquisition (attention, association, remembering; for large units, small units, or abstract category relations) colour could support, and how to design schemes that support each of these goals.

For each intervention, I identify: a) *what* task the researchers used colour to support b) the *mechanism* by which colour was presumed to help and (when applicable) c) limitations of the approach or experiment and their implications for design.

### **Reading with Words in Colour**

Caleb Gattegno (Gattegno, 2000) used colour to help children without dyslexia learn orthographic rules. Gattegno mapped each English speech sound (all 46 phonemes, plus blends and stable units, totalling 96 sounds) to a unique colour. He then exposed children to charts of words whose letters were coloured according to their sounds. Colour was supposed to help children learn the relations between certain English orthographic and phonological contexts by enabling them to infer the

pronunciations of unfamiliar words via visual comparison to familiar words. For example, Gattegno would teach children the rule, “c sounds hard (*like k or ck*) before a, o and u, but soft (*like s*) before i, e and y” by exposing them to the words: cite, sit, cake and pack. Because they sound similar, “c” in “cite” would have the same colour as “s”; “c” in “cake” would have the same colour as “ck”. Children would therefore be capable of pronouncing words they did not know, such as “cite” and “cake”, via matching colours from words they did know, such as “sit” and “pack”. Gattegno believed that repeated exposure to coloured words and the experience of decoding new words by comparison to familiar ones would consolidate children’s mental representations of the rules and support their abstraction and transfer to uncoloured text, better than if teachers were to tell children the rules explicitly. Gattegno’s principles are shared by OG Guided Discovery, which de-emphasizes explicitly telling children rules, and instead guides children to discover patterns themselves.

I was unable to find an empirical assessment of Gattegno's scheme and so may only critique it on theoretical grounds. My criticism of Gattegno’s scheme is that the large number of colours and correspondingly fine-grain of phonological information might be inappropriate for highlighting orthographic patterns involving larger categories.

For example, one pattern that Gattegno intended students to notice is that the vowel in vowel consonant-e syllables sounds long. Because the pattern holds for every particular vowel and consonant, it is a relationship between categories of letter (consonant or vowel) and vowel sound (long or short). Gattegno’s scheme assigned distinct colours to each particular vowel and consonant sound. Humans tend to assume that similarly coloured elements form a group; they assume that differently coloured elements are distinct (Gellatly, Pilling, Cole & Skarratt, 2006). Gattegno’s rationale for exposing children to multiple examples of a rule is that children should notice the examples’ common features (in this case, that *vowel-consonant-e* always correlates a long vowel) and abstract a general rule. Assigning different colours to the letters in, for example, the words *fade*, *bite* and *mope*, might have prevented children from appreciating the –groups- of letters (vowels, {a, i, o}; consonants, {d, t, p}, and e) between which the relationship holds, and consequently from abstracting the syllable pattern. A scheme that distinguished only vowel from consonant and short from long

vowel sound, the same way that OG tutors use breves and macrons to distinguish short and long vowel sound, might better support children in noticing that what the examples share is their *pattern* of consonant and vowel and vowel sound category.

## Dybuster

Kast et al (Kast, Meyer, Vogeli, Gross & Jancke, 2007; Kast, Baschera, Gross, Jancke & Meyer, 2010) used colour as part of a multimedia audio-visual software application (DyBuster) that sought to improve the spelling of children with dyslexia by training them to re-code textual representations of words as unique audio-visual patterns. Visual properties (shape and colour) represented particular letters or letter identities (e.g., capital, accented, lower-case). The rationale was based on Paivo's application of dual-coding theory to reading (Sadoski & Paivio, 2004)-- that words can be represented in multiple forms, each of a different modality, and that forms of one or another modality can "trigger" activation of the others-- and the observation that children with dyslexia prefer visual to verbal encoding strategies (Miller & Kupfermann, 2009), such that visual representations might be easier for children to encode and retrieve than verbal or textual representations<sup>3</sup>.

Kast et al surmised that associating the word's letter-sound correspondences with audio-visual conjunctions- a colour and shape at a specific location- would produce stronger memory traces than the textual letter alone, and help children retrieve the textual letter when presented with the word's sounds (i.e., spelling by dictation), and its

---

<sup>3</sup> Although *attending to colour* may involve different visual channels than attending to shape, texture or form, *maintaining colour information in visual working memory* (e.g., remembering the locations of differently coloured items, their sequence or arrangement) seems to involve the same processing channels as other information, i.e., the "visuo-spatial sketchpad". That colour and form involve the same visual working memory channel is suggested by the observation that capacity limits on visual working memory (recent estimates being four elements) apply to colour and shape (i.e., observers can maintain four distinct colours *or* shapes, or any combination thereof) (Vogel, Woodman & Luck, 2001). If colour and shape involved different working memory channels then we should be able to remember four colours *and* four shapes. The notion that colour and shape involve the same visual working memory system is also consistent with the notion that visual working memory is responsible for "sensory binding"- *integrating co-located colour and shapes into singular visual objects (or conjunctions)* (Wheeler & Treisman, 2002). Indeed, observers can remember four colours and four shapes *provided that* each colour and shape is paired, i.e., into four mutually exclusive conjunctions (Vogel, Woodman & Luck, 2001).

corresponding audio-visual code. Kast's premise is consistent with OG multisensory principles.

In contrast to Gattegno, Kast et al used eight colours. Therefore, some letters had the same colour. Colour's primary role was reducing spelling errors that attributed inter-letter confusion, resulting from either visual (e.g., mistaking “b” and “d”) or positional (e.g., mistaking “t” for “p” because they are both common first letters) similarities. Although certain letters had the same colour or shape, each word had a unique arrangement of colour-shape conjunctions.

Kast et al trained children to associate letters to colours, to segment words into syllables and letters, and to re-code the audio-visual representations as textual representations as words (i.e., to spell). Kast et al found that children who used Dybuster improved more in their spellings of instructed and uninstructed words to a greater extent than controls (Kast, Meyer, Vogeli, Gross & Jancke, 2007).

That children's spelling improvements generalized to uninstructed words suggests that DyBuster provided children some generalizable spelling knowledge. It is possible that DyBuster functioned similarly to Berninger's intervention, but applied colour in the manner that Gattegno intended. In German (as well as in English), certain multi-letter and sound combinations frequently occur (Kast, Baschera, Gross, Jancke & Meyer, 2010). Mapping colour to letters, matching them to sounds, and flooding children with many example words might have enabled children to learn correlations between multi-letter units and sounds as correlations between certain patterns of colours and sounds (e.g., a common letter digraph would become a common colour-pair). Given that information designers recommend a maximum of six-eight colours (Ware, 2012), the small set of colours that DyBuster had would have made it easier for children to notice global colour-sound patterns.

Because colour was one of several alternate visual codes, and because children were explicitly taught a syllable segmentation strategy that is also supposed to benefit spelling and decoding (Berninger, Abbott, Brooksher, Lemos, Ogier, Zook & Mostafapour, 2000; Bhattacharya & Ehri, 2004), Kast et al's experiments do not clarify what role colour played. Furthermore, it is unclear how Kast's approach would benefit

learning orthographies like English, wherein a goal is appreciating that many specific words are instances of a common category. Kast's scheme would assign distinct colours to letters appearing in the same position; letters playing similar roles (e.g., vowel in closed syllable) often appear in the same position.

### **The Colour Vowel Chart**

Madeline Wrembel (Wrembel, 2007; Wrembel, 2009) used colour to help English and Polish adults learn vowel sound-letter correspondences during second language (L2) learning. Like Gattegno, Wrembel coloured letters according to their sound. Unlike Gattegno, Wrembel focused her scheme on sounds (vowels) that humans innately associate to particular colours.

In several earlier experiments, Wrembel discovered that adults associated categories of vowel-sound with categories of colour: front vowels (/i/, /a/) with warm colours (red, orange, yellow), central vowels (e) with green or cyan, and back vowels (/o/, /u/) with cool colours (blue or purple). Because Polish and English speakers yielded similar mappings, Wrembel assumed that the mappings attributed intrinsic properties of the vowel sounds, versus learned (language-specific) associations (Wrembel, 2009). Wrembel's idea, which was never empirically tested, was to present second language learners with text wherein the vowels were coloured according to their sounds. Colour was supposed to help second-language learners memorize new symbol-sound correspondences via activating multiple “emotional and sensory pathways” and thereby create a more sensorially and emotionally involved experience (Wrembel, 2007). Emotionally and sensorially complex experiences are easier to remember than unemotional plain experiences (MacKay & Ahmetzanov, 2005).

Wrembel did not assess young children so it remains unclear whether these associations are truly innate or attribute learned, third associations that are sufficiently general to cut across languages (e.g., humans tend to produce front vowels when they are excited or upset; most cultures use warm colours for alerts). Either way, Wrembel's experiments suggest a readiness to associate colours and sounds that exceeds that of associating letter-forms to sounds.

Although individuals with dyslexia present impairments learning new associations between non-linguistic visual and auditory stimuli (Hulme, Monk & Ives, 1986), and are no better at retrieving the names of colours than of letters or numbers (Wolf, Bally & Morris, 1986), no one has confirmed that they do not share innate biases to associate certain visual and auditory stimuli, such as Wrembel's vowel colours. Indeed, if they share in innate biases to associate colours and vowels, individuals with dyslexia could leverage the bias to associate letters and vowels: colour could serve as “visual mnemonics” that bridge between letters and sounds. Because colours can be physically projected upon letters, they would evade the criticism that mnemonics consume more resources than they save (Wagner & Torgesen, 1981). Wagner and Torgesen's criticism applies to mnemonics that require learners to visualize and manipulate a third mental feature, such as Ehri's objects. If there are innate colour-sound associations then seeing letters coloured a particular way should evoke the sound automatically. Using them would require little mental visualization or manipulation.

Innate colour-sound associations could even help children read *uncoloured* text. Strong associations between stimuli (A and B) can transfer to a third stimulus (C) if learners associate A and C. This effect is called “associative transitivity”. Colours might support associative transitivity. Colour-grapheme synaesthetes showed associative transitivity between colours, letters and numbers (Watson, Blair, Kozik, Akins, & Enns, 2012). Repeated exposure to coloured text can develop some “pseudo-synesthetic” effects in non-synaesthetes (Colizoli, Murre & Rouw, 2012). As with associating colours to sounds, humans seem to readily associate colours to letters. Therefore, if associative transitivity characterized the associations between colours and sounds, then learning associations between colours and letters might- without much additional effort- teach children the associations between letters and sounds.

Wrembel did not empirically test her scheme, so its effectiveness is unclear. Wrembel also restricted her scheme to colours and sounds that were innately associated, which limits its application to literacy learning (consonants are excluded) and exposes it to the same criticism I posed to Gattegno: the associations hold between *particular* sounds, not categories of sound, but many useful orthographic patterns involve categories of sound. Nevertheless, Wrembel's experiments further validate the

idea of using colour to help children associate letters to sound, and suggest a design guideline of mapping sounds to colours on the basis of either Wrembel's innate associations or general learned associations.

### **Colour to Discriminate Letters**

Goodman and Cundick (Goodman & Cundick, 1976) investigated the notion of using colour to help language learners associate symbols to sounds. Their work was similar to Wrembel's, with three main differences. First, Goodman and Cundick's colour-symbol mappings were arbitrary. Second, participants were children. Third, children associated symbols to syllables, not their true sounds (phonemes). The symbols and syllables were unfamiliar. Goodman and Cundick presented children with displays containing all symbols and played children a syllable. Children picked the symbol that matched the syllable. For half the children, each symbol appeared in a unique colour. For the remaining children, symbols were black. On the basis of Gattegno's research, Goodman and Cundick predicted that colour would help children learn symbol sound relations, i.e., they would achieve 100% matching accuracy in fewer trials than children using colourless symbols, and would transfer performance to uncoloured letters. Although children using coloured symbols achieved 100% matching accuracy quicker than children using colourless symbols, their performance did not transfer to colourless symbols.

The colour effect was greatest for pairs of symbols that were distinguished by a small contour detail (e.g., presence or absence of a crossbar). Goodman and Cundick interpreted the results to suggest that colour helped children *discriminate* the Hebrew symbols. Specifically, Goodman and Cundick surmised that the salient colour differences provided children an easier discriminating feature than the small contour details that otherwise differentiated the letters.

Colouring letters differently would have provided children an opportunity to focus on and process the contour differences between the symbols. Had children done this, their greater matching accuracy should have transferred to colourless letters. Therefore, the fact that children's performance did *not* transfer suggests that children did not spontaneously exploit the attentional focusing of colour to encode the letters' forms. This

is consistent with Hanna and Remington's observation that users do not remember contour-based scene details unless the experimenter requests it (Hanna & Remington, 1996).

The ease of processing salient colour differences may in fact deter children from processing letter-form. Such considerations motivate a design guideline of using colour to focus children's attention to letters (that are connected to particular sounds) but then explicitly encouraging children to encode the letter-forms.

### **Colour to Highlight Multi-Letter Units**

Hines (Hines, 2007) used colour to help struggling readers learn and transfer two types of "rimes". Rimes (2.2.3) are multi-letter units that comprise a word's final vowel and consonants. Hines assessed rimes with short-e and short-a vowels (e.g., "pet", "bat"). Hines extended Berninger's approach of colour-highlighting important multi-letter units by using a colour contrast to draw children's attention to the key phonetic difference between her rimes: short-a rimes were blue; short-e rimes were green. Colour was supposed to help children a) identify and memorize the rime units and b) remember which vowel sound featured in each unit. Like Gattegno, Hines also intended the colour to help children abstract a *general* strategy of rime based decoding, versus the specific example rimes that Hines taught.

Hines empirically tested her scheme. Hines exposed seven children to storybooks consisting of short sentences, the final word of which involved an instructional rime (e.g., "the new pet"). Hines used a repeated measures design wherein children alternated practice sessions reading storybooks with uncoloured and coloured words. After each session, children read three kinds of words: instructional words, 'near' transfer words and 'far' transfer words.

Instructional words were used in the storybooks. Near transfer words were new, but had the same final consonants (i.e., the same rimes) as instructional words (e.g., "bet" instead of "pet"). Far transfer words were new, but had different initial and final consonants (i.e., different rimes) than instructional words (e.g., "ben" instead of "pet").

Far transfer words therefore tested whether children had acquired *the general strategy of rime based decoding*, versus the specific example rimes that Hines taught.

Teaching children to acquire *general* purpose skills is considered a more effective approach than teaching them sets of specific examples (Treiman & Kessler, 2006). One purpose of teaching specific rime analogies to children is encouraging them to adopt a general strategy of attending to and memorizing the rime units of new words they encounter. On the assumption that most English rimes are profligate and useful, children who attend to rimes will be better equipped to leverage their vocabulary to read additional new words than children who do not (Ziegler & Goswami, 2005).

The key impediment to children with dyslexia in using these strategies is *noticing the relevant patterns*. Hines designed her scheme to help children *notice the rime units in instructional words*. That children's performance on instructional and near transfer words improved suggests that Hines' scheme helped them memorize the specific rime units to which Hines exposed them, but the lack of improvement for far transfer words suggests that Hines' scheme did not teach children the general strategy of attending to rimes. If it had, children might have attended to rimes in words they encountered outside of Hines' intervention. Assuming that Hines' far transfer words had the same rimes, children would have performed better on Hines' far transfer words.

Features of Hines' colour-coding scheme may have prevented children from developing a general strategy of attending to rimes. Where Gattegno's scheme may have been too fine-grained; Hines' may have been too coarse. Her colours distinguished short "e" from short "a" rimes, but they did not distinguish vowel from consonant. Failing to distinguish vowel from consonant may have prevented children from appreciating the definition of a rime: the final vowel and consonants of words. Hines' assignment of distinct colours to short "e" and short "a" rimes warrants the same criticism I posed to Gattegno. Although Hines' colour contrast highlighted the information that rime units provide (the vowel sound), it might have prevented children from mentally grouping the rimes. Mentally grouping the short "e" and short "a" rimes into the abstract category, rime unit that communicates vowel sound, might be a pre-requisite to developing the general strategy of attending to rimes.

## **Colour to Highlight Abstract Categories**


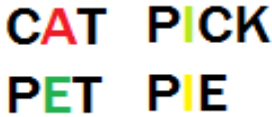


Colour can highlight abstract linguistic categories. Alphabetic Phonics are extensions of OG interventions that feature a wider variety of multisensory cues (Cox, 1985). One extension is using colour contrasts to direct children's attention to various kinds of abstract linguistic information. One category colour distinguishes is consonant versus vowel. Children learn letters with flashcards; vowel flashcards are salmon; consonant flashcards are white. Another category is position. Prefixes, which appear at the beginning of words, are green; suffixes, which appear at the end, are red. The rationale here is similar to that for other multisensory techniques. The more channels that are recruited to convey information (e.g., "a" is a vowel) the more likely that children will learn it.

Although Alphabetic Phonics curricula have been examined overall, specific aspects of their approach- including colour coding- have not. On theoretical grounds I question the effectiveness of how these curricula apply their colour-codes. Memorizing vowel and consonant sets is merely a foundation for learning other reading and spelling rules, such as rime analogy and syllable patterns. My reading of Alphabetic Phonics curricula suggests that tutors do not leverage the colour-category associations to help children learn new information, such as syllable analysis. In syllable analysis lessons, for examples, tutors show children syllable boundaries with colourless words that are spatially split into syllables (e.g., SCI ENCE). The attentional and mnemonic properties of colour suggest that its efficacy might lie in helping children identify facts that are relevant to syllable analysis, such as "every syllable has exactly one voiced vowel" (perceptually: every syllable has one salmon letter). If colour has mnemonic effects, then another design guideline is using consistent colour codes throughout children's curricula. Colour codes that support early learning (e.g., categorizing letters as consonants or vowels) might then serve later learning, via triggering students' memories of prior information, on which the new knowledge builds.

#### 2.4.4. The Need for a Principled Framework for Using Colour in Literacy Interventions

Researchers and educators have used colour to highlight letter units for attention and to support visual discrimination. The possibility of intrinsic associations between some colours and speech sounds invites the possibility of using colour as an associative bridge between letters and sounds. Stimuli that have the same colour tend to be grouped, both perceptually (e.g., when the stimuli appear in the same scene), and conceptually. Common colours may also help individuals classify new objects into a single category. For these reasons, colour that distinguishes abstract linguistic categories (consonant versus vowel; short versus long vowel) might support children in abstracting more general descriptions of useful linguistic units, which they can apply to derive greater benefit from self-directed reading.

**Table 2-1. A summary of some previous attempts to apply colour to literacy.**

Name of designer or system	Illustration <sup>a</sup>	Design Goal
Gattegno (2000)		<ul style="list-style-type: none"> <li>• Infer pronunciations of new words</li> <li>• Discover contextual rules (e.g. vowel-consonant-e)</li> </ul>
Wrembel (2007; 2009)		<ul style="list-style-type: none"> <li>• Memorize vowel sounds; associate to letters</li> <li>• Adults; second-language</li> </ul>
DyBuster (Kast et al, 2007; 2010)		<ul style="list-style-type: none"> <li>• Mentally distinguish letters that appear similar or in similar positions <sup>b</sup></li> </ul>
Hines (2007)		<ul style="list-style-type: none"> <li>• Recognize certain rimes</li> <li>• Distinguish short "e" from short "a" rimes</li> <li>• Develop general strategy of rime-based decoding</li> </ul>

<sup>b</sup> Illustrations are my renditions. They are not originally sourced from the researchers.

<sup>b</sup> Designed for German readers. My illustration applies the scheme to English but Kast et al did not.

Table 2-1 summarizes my survey of previous attempts to apply colour to literacy. As table 2-1 shows, there is considerable variability in the design (how many colours the researchers used, which units were highlighted) and objectives (which rule or strategy did they aim to teach?) of previous attempts to apply colour to literacy acquisition. Although the theory and some preliminary assessment of these attempts holds promise, research and design have been stymied by the lack of a principled framework that articulates the key variables by which the schemes differ, and relates these variables to outcomes in a manner that could guide design.

A framework for designing colour codes would provide designers one or more general principles, which would outline a general design requirement, and a set of more specific design elements, which would identify the key design choices, alternatives and trade-offs associated with each alternative. A comprehensive framework for designing colour codes presupposes a) meta-analysis of previous approaches to identify key factors differentiating the colour codes, which translate into design decisions, b) systematic experimental comparisons of the impact of different design decisions (i.e., values on the variables) on various outcome measures and c) identification of and agreement upon desirable outcome measures.

I do not claim to provide or to test a comprehensive framework. The necessary research is outside of my scope. Here, I take the first steps towards a comprehensive framework. On the basis of my literature review, I propose two general principles and identify four design elements, each centered around a core variable (cum design decision), that might impact a code's effects. To help designers narrow the design space, and inform their choices within the space, I summarize the main design alternatives associated with each decision, the trade-offs associated with each alternative, and the factors that designers might consider in making their decisions.

I articulated my framework as a means of rationalizing and streamlining my process of designing the colour codes for PhonoBlocks. The variables were informed by my reading and are the major design decisions that I encountered. My framework documents how I rationalized my design choices, and it helped to structure my

observations of PhonoBlocks' users. Although other researchers could use my framework as a springboard for designing their systems, I do not claim to prescribe the solution for every design problem or context. One additional use of my tentative framework is a source of future research hypotheses. To support research, I have striven to articulate my principles, variables and recommendations as testable predictions. Designers and researchers should therefore consider my tentative framework as a set of premises that are open to revision and refutation.

### ***My Framework for Designing Colour-Coding Schemes***

My framework encompasses two interrelated general principles and four design elements. The general principles outline a general design goal. The general design goal is using colour to sculpt children's attention to the information that is needed for them to learn a linguistic rule (e.g, the relation between syllable types and vowel sounds) or master a decoding or spelling strategy (e.g., reading by example, as in Gattegno's approach). The design elements clarify how to apply the general principles. Each element focuses around a "design choice". Each design choice involves a variable that I identified as relevant in differentiating colour codes, on the basis of my literature review. The design elements do not prescribe specific courses of action but seek to identify the possible trade-offs associated with design alternatives and what designers should consider in making their selections.

My framework involves many interrelated principles, predictions and open research questions. A comprehensive empirical validation of my framework was beyond the scope of my thesis. In this thesis, I aimed to use my framework to develop two colour codes that could be implemented in our working software prototype, and explore their use.

#### **General Principles:**

First, I state my general principles. On the basis of my review of the uses of colour coding in information displays and multimedia learning, and its general attentional properties, I propose:

A) Colour codes should be designed around one decoding or spelling strategy or rule (e.g., a colour code for learning syllable types, a colour code for learning decoding/spelling by example)

B) Colour codes will be useful for learning a decoding or spelling rule insofar as they highlight all and only the distinctions that are relevant for a given decoding or spelling rule.

### **Design Elements:**

My two principles refer to four design elements that must be considered. For each, I identify the element, provide guidance and suggest how to deal with trade-offs. I supplement my descriptions with a use case.

#### **DE1: Identifying the “relevant” distinctions in a reading or spelling rule**

Distinctions are differences on one or more properties that an individual must consider in deciding how to spell or pronounce a word, or deciding whether or not an example word fits a definition. Properties are those of letters. Some examples are: the letter's position, identity as vowel or consonant, sound, sound category, membership in a letter-unit, role in a letter unit, etc. The choice of what to highlight is therefore not simply which letters to highlight, but which property to highlight.

Some properties can be ranked in terms of informativeness. For example, a code that conveys a letters' role in a unit codes the letter's presence in a unit, but a code that conveys presence in a unit does not necessarily code its role in a unit. Information visualization designers acknowledge that attention is a limited capacity resource, even when using an easily processed feature (such as colour): humans can maintain about eight colour-category associations at any given time (Ware, 2014). Because more informative codes imply more colours, there is a potential trade-off between informativeness and sensory confusion. Minimizing the number of colours is a reasonable design goal.

Designers can approach a learning task by asking: what specific properties do children need to “see” in order to learn the rule/master the strategy? For example, when the goal is learning the general definition of consonant-le syllables, children might benefit from a colour-code that identifies each letter's role in the unit, (as well as presence in a

unit). A good colour code would highlight letters appearing in consonant-le syllables, but additionally would assign different colours to each of the three roles: initial consonant, l, and e. Conversely, when children are learning syllabification, (and the rule that consonant-le is a “pivot” around which one divides a word), children need only to know the -presence- of each letter in the consonant-le unit. A good colour code would highlight all the letters appearing in consonant-le, but would assign the same colour to each letter in the unit (i.e., would not distinguish between roles).

A possible trade-off is that children may become confused by changes in how specific mental units are coded, and changing the way a unit (such as consonant-le) is colour-coded might nullify the colours' capacity as a memory aide. That is, if consonant-le had different colours when learning its definition and applying it to syllabify words, then the colours might not help children retrieve previously learned information about consonant-le. One avenue for future research is assessing the performance trade-off between the lesser attentional sensory confusion of coding the same units with different informational grains, and the greater mnemonic potential of consistent colour-codes.

## **DE2: Identifying what to highlight with a unique colour**

“Highlighting” is using colour to draw attention to a relevant distinction. As described in (1), the limitations on human visual working memory suggest that- when possible- designers should reduce the number of distinct colours. Some distinctions (such as a presence versus absence of a letter at a given location) might be highlighted by “emergent colour patterns”, rendering an additional colour unnecessary. For example, suppose a colour code distinguished vowel from consonant (vowels red, consonants blue) and the child was learning to divide words into syllables. It is relevant whether a consonant appears before or after the vowel, because a consonant appearing after a vowel changes the syllable type (and hence the vowel sound). Despite the importance of the before/after vowel property, assigning different colours to consonants appearing before and after a vowel might be redundant: there is already an emergent visual difference- red then blue versus blue then red- which children could associate to the differences in syllable type.

A possible trade-off is that patterns of colours demand more resources to remember than single colours, and the differences in patterns of colours are less salient than differences in pure colours. Patterns should be limited to simple two-colour examples, such as that I described, where the members of the pattern are adjacent. Another avenue for future research is exploring the variables that determine the point at which the cost in sensory load (from multiple unique colours) exceeds that of discriminating emergent multi-colour patterns.

### **DE3: Identifying when and how to make colour dynamic**

“Dynamic” means that the colour of a letter at a given position might change, throughout the course of an activity, and that the change is visible to children. Software mediums, which can easily change the appearance on onscreen letters, therefore wield an additional visual dimension: change, and the variables (temporality, directionality and reliability) that characterize change.

Linguistic rules involve correlations between orthographic and auditory features (e.g., the closed syllable pattern correlates to a short vowel sound). I describe the relationships as correlational (versus “causal” or deterministic) because virtually every English reading or spelling “rule” includes “edge cases” for which the rule does not hold. The fact that reading and spelling rules do not always hold is relevant information for learners, because it predisposes them to seek and memorize exceptional cases (Archer, Gleason & Vachon, 2003).

Orthographic analyses, such as those by which Goswami (Goswami, Ziegler, Dalton & Schneider, 2003) justified her emphasis on rime-based decoding, can indicate a rule's reliability. Dynamic mediums can use visual change to communicate the reliability of a linguistic rule. For rules that are less reliable, changes could occur within random temporal periods of one another, the colours could flicker intermittently, and the letters could alternate which changes first.

On the other hand, random, intermittent or otherwise unpredictable changes may be less conducive to lower-level sensory binding, which is strongest when the to-be-associated stimuli occur simultaneously and reliably (Powers, Hillock & Wallace, 2009).

One way to balance the trade-off would be to implement a trajectory similar to that of OG curricula. An application can introduce a generalization in a “toy context” of words for which the rule holds. In this context, it is appropriate to present the rule as “perfect correlation”, so the coupled changes in stimuli can be immediate and reliable. After the child comprehends the rule, the application can refine children's understanding via the introduction of “exception words” (for which the rule does not hold). In the expanded context, delaying or flickering the changes in colour between related stimuli could communicate to learners that, before applying a rule, they must determine whether the word is exceptional.

#### **DE4: Identifying which colours to use**

A final question is how to assign colours to different pieces of information, again with the goal of helping children learn a linguistic rule or strategy. Two key processes are involved in learning: attention, and recall. “Attention” concerns children's ability to process the colour codes during learning; “memory” concerns children's ability to recall what they learned, either with coloured or uncoloured letters.

General colour principles predict different attentional effects. For example, the contrasts between colour complements (red-green, blue-orange) are more salient than contrasts between non-complements (red-yellow, blue-green). Colours fall into certain natural sets, for example, warm and cool. Although identically coloured elements form the strongest perceptual groups, elements with colours within similar sets (particularly the warm and cool set) may be more readily grouped than elements with colours from different sets. Finally, warm colours may be more attentionally salient than cool colours. Designers should consider these factors, in tandem with their activity's goals, in assigning colours to elements. For example, a rule might require children to distinguish both vowel from consonant but also to distinguish different particular vowels. To support children 'grouping' all vowels and all consonants, the designer could use one set (warm or cool) for vowels, the other for consonants.

Designers should also be aware that certain user groups may be already accustomed to associating certain colours and pieces of information (for example, a classroom alphabet chart that presents “c” with the image of an “orange cat” may have

biased children to associate “c” and orange). In these cases, it might be advisable for the designer to respect the associations that users already have, versus requiring them to learn new ones. Such decisions need to be weighed against the cost to other aspects of the code (e.g., a rule to colour all consonants cool). The decision should minimize the accumulated processing cost, so if the code picks out all consonants (versus just c), and if all consonants (except c) are better coded with cool colours, then if a goal is grouping “c” into a set of consonants, the best choice is probably to assign “c”, along with all other consonants, a cool colour.

One issue in using colour codes is that learning them imposes additional cognitive demands, which possibly translates into greater working memory load. Greater working memory load would leave fewer resources available for learning the rule. Conversely, associations that are more intuitive might be easier to learn, involving fewer working memory resources, and easier to remember. If they exist, then, designers should exploit intuitive associations between colours and sounds or letters.

Wrembel uncovered some potentially innate associations between colours and linguistic information, though the information was limited to vowel sounds, and her studies involved adults. Although Wrembel conjectured that the associations were innate, innateness is not equivalent to intuitiveness: certain mappings between colours and letters or sounds may be more “intuitive” (or easy to understand) than others, on the basis of some explicit or implicit third association. For example, there is a class of consonants called “fricatives”, the production of which involves air blowing over the lips. The association between air and light blue might bias individuals to associate fricative sounds with light blue more readily than with other colours. Certain colour-letter associations may be more intuitive on the basis of colour-words, in which the letter is first. (E.g., “g”-green; “b”-blue). One task for researchers is seeking out and assessing such proposed associations.

A final consideration in choosing colours is that inter-colour relationships might, along with the dynamic properties discussed in DE3, convey correlation. For example, the “r” in r-controlled syllables is associated with (or “makes”) the vowel sound a particular way (“r influenced”). A dynamic application could leverage visual change to

convey the relation between the presence of “r” (versus another consonant) and the vowel sound, but static colour choices could encode this as well. Suppose the “r” is blue. Then if vowel sounds were warm colours, r-controlled vowel could be magenta: warm, but “tinged” with blue (i.e., “r influenced”). One might also consider assigning the exact same colour to elements that influence one another, (e.g., colour r and r-controlled vowels blue), but this scheme fails to communicate- in colour- which classes of properties are involved (i.e., the sound category of vowel versus the identity of the consonant).

### **Using the Framework:**

As an example of how designers would use the framework, I here describe briefly how it would apply to designing colour-codes for learning the rule, “c” sounds hard before “a”, “u” and “o”, but soft before “i, e” and “y”. The example shows how a designer might use the overarching principle and the four elements to guide their design, in the sense of identifying the key design decisions, the alternatives and trade-offs associated with either approach.

#### ***Use Case: colour-coding the rule, “c sounds hard before a, o and u, but soft before e, i and y”***

The relevant information is whatever might change a child's decision to pronounce a “c” as hard or as “soft”. We assume that “c” appears in the word. The first piece of information is: does the “c” precede a vowel? If yes, the rule applies. Otherwise, it does not. Therefore, “cs” appearing before vowels should be coloured differently than “cs” appearing before consonants. Cs appearing before consonants must be in a digraph or blend. Because determining the digraph or blend to which bound “c”s belong is irrelevant to determining the sound of “cs” that appear before vowels, bound cs should be uncoloured (white).

The next piece of information is which vowel follows c? Suppose the vowel is “a”. If the vowel changed from “a” to “o” or to “u”, the child's decision (how to pronounce “c”) should not change, though the child's decision should change if “a” changed to “i”, “e” or “y”. Therefore, a child does not need to attend to differences between “a”, “o”, “u”, or between “e”, “i” or “y”. The child only needs to attend to differences in identity that span

the sets. To support this, the letters of the set {a, o, u} should receive a unique colour; the members of {i, e, y} a different colour. For the time being, the designer defers selecting specific colours for the vowel sets. She plans to calibrate them to the colours of hard and soft “c”, which she suspects will involve more pre-existing constraints. Because the vowel's set membership is relevant if and only if it follows a c in a word, vowels in the word that do not follow “c” should be uncoloured.

Assuming that the codes are implemented in software, designers can visually communicate the relationship between the vowel's set and c's sound. To do so, “c” must adopt different colours when it sounds hard versus soft. In this case, because “c” is the only consonant that needs to be coloured, the designer decides to leverage a colour-mapping to which her users are accustomed. We will assume that the children are accustomed to associating “c” with orange, for “orange cat”, and that the designer therefore colours hard “c” orange. On the other hand, soft “c” sounds like “s”. We assume that the designer's users are accustomed to associating “s” with silver (for “silver snake”). To highlight the identity between the sounds of soft “c” and “s” (i.e., in the same manner as Gattegno might), the designer colours soft “c” silver. The designer considered colouring soft c a different shade of orange, such as to respect the children's pre-existing associations, but determined that the need for a strong visual contrast between hard and soft c over-weighed the potential cost of colouring soft “c” differently than children might have expected. Having fixed the colours of hard and soft “c”, The designer can then select colours for the vowel sets. Knowing that the “c” and vowels will appear adjacent, and that a bold visual contrast will attract children's attention, she assigns to each vowel set a colour that complements that of the “c” sound that matches it (blue, to the vowels



matched to hard “c”; yellow, to the vowels matched to soft “c”).

**Figure 2-4. An example of a solution to the design problem of communicating the hard versus soft “c” generalization, developed by applying my framework for designing colour-codes.**

Although the focus of my thesis is design exploration, not an empirical experiment, I elected to take the first steps towards not only *using* but *assessing* my

framework. One assesses a framework the same as one assesses a theory: by deriving and testing falsifiable predictions. A core prediction of my framework is that a colour code should “discount” distinctions that are not relevant to the immediate decoding/spelling task. A more specific phrasing is that children would be better supported in learning rules involving fine grain linguistic information, such as particular vowel sounds, by a colour-coding scheme that assigns different colours to different particular sounds. Conversely, children should be better supported in learning rules involving large-grain linguistic information, such as the relation between syllable and vowel sound *categories*, by a colour-coding scheme that assigns different colours to different sound categories, but the same colour to certain particular sounds. The necessary experiment must design two colour codes, one for fine-grained and one for coarse-grained information, and compare interactions between learning gain, literacy rule and colour code. The framework predicts that children should learn the fine grained rule better with the fine grained scheme; the coarse grained rule better with the coarse grained scheme. I used this experimental goal to select the rules and design the colour codes that I used in my study.

Any framework proposing a novel interface strategy for learning or consolidating material faces another important question, which is whether performance gains resulting from the novel interface transfer to traditional interfaces (i.e., paper and pencil). In this case, the question is whether performance gains resulting from colour-coded words transfer to uncoloured words. “Performance” refers to accuracy on spelling and reading tasks involving the same or similar words to those used in instruction.

Although there has been a lack of systematic assessment, so far, experiments have yielded conflicting results on the viability of colour for supporting transferrable learning. From my analysis of Goodman and Cundick’s results (Goodman & Cundick, 1976), I premise that transfer depends partly on the emphasis that the instructional context places on leveraging colour to *process* letter form, versus attending to colour alone. I incorporate these considerations into my designs of both the codes themselves and the activity contexts in which I implement them. To explore whether my method of designing colour codes supports children in developing *transferrable* knowledge, in my overall assessment of children’s performance gains, I include assessments that measure

transfer of performance to a) uninstructed (but coloured) words, and b) uncoloured words (both instructed and uninstructed).

## **Chapter 3. Methodology**

This chapter describes the methods by which I designed and assessed my colour coding schemes. I divide this chapter into three sections. First, I outline my research questions, hypotheses and study design. Second, I describe the research prototype, PhonoBlocks, in which I implemented my codes. This section includes descriptions of my chosen literacy concepts and how I applied my framework to design colour codes for them, and the tutor design sessions by which I validated my literacy concepts and colour schemes and finalized the learning activities. In the third section, I describe the study participants, procedure and data collection.

### **3.1. Research Design, Questions and Hypotheses**

To assess the validity of my framework, and better understand the mechanism by which colour might help, I used a pre-post-test multiple-case studies design. Before the intervention, children completed a pre-test of their ability to spell words that were subsequently used in the intervention training activities. Children spelled the words using the research prototype (3.2). One set of words were coloured (by the schemes that I developed); another was uncoloured. Children then underwent a four-week intervention during which they used the system and colour codes to spell words that were equivalent to or that involved the same rules as those in the pre-test. After the four-week intervention, children completed a post-test. The post-test was identical to the pre-test, except that half of the words were new.

I applied my framework to design two colour codes for two different literacy activities. Half of the children in my study experienced one colour-coding scheme; half experienced another. Each child used my software system to practice both types of literacy activity. I predicted that children would experience greater pre-post test gains for the activity that matched their coding scheme.

My first overarching research question was whether children who practice a spelling rule given a colour-coding scheme that highlights attention to all and only linguistic variables that are relevant to the rule would show greater improvements in their pre and post practice spelling accuracies for words involving the rule than children who practice with a scheme that highlights irrelevant linguistic variables.

The two rules I focused on were consonant-gemination and vowel discrimination. I measured them using OG-styled spelling activities that required children to complete either a consonant-le word, given the onset of the first syllable, or a short vowel word, given all letters but the vowel. Consonant-le spelling requires attention to vowel sound category but not particular vowel sound; vowel discrimination for monosyllable words requires attention to particular vowel sound but not vowel sound category. I detail and justify these rules and activities in sections 3.2.2 and 3.2.3. I introduce them here to articulate my specific research questions:

**RQ1:** will children who experience a colour-coding scheme that highlights vowel sound category show greater pre and post-test improvement in spelling accuracy for consonant-le words than children who experience a colour-coding scheme that highlights particular vowel sounds?

**RQ2:** will children who experience a colour-coding scheme that highlights particular vowel sounds show greater pre and post-test improvement in spelling accuracy for monosyllable short-vowel words than children who experience a colour-coding scheme that highlights vowel sound category?

Based on my framework, I hypothesized (**RQ1**) children with the categorical scheme would show greater improvement in spelling accuracy for consonant-le words than children with the particular scheme but (**RQ2**) children with the particulars scheme would show greater improvements in their spelling accuracy for monosyllable short vowel words.

My second overarching research question was whether children's pre and post gains transfer to uncoloured or uninstructed words. From my review, I surmised that the pattern of transfer would reflect the mechanisms of colour's benefits. Understanding how colour benefits children can help designers to better exploit colour. Although the specific mechanism by which colour would help was unclear, from my review I derived three

alternatives. For each mechanism I hypothesized a different pattern of transfer; the failure to observe certain patterns can therefore eliminate some alternatives, and increase our understanding of how colour helps. I index my hypotheses CH (for colour-hypothesis) [1...3].

**Colour promotes deep encoding, reflection and information-integration:** if colour promotes deeper encoding and reflection on the connections between letters and sounds, then (**CH1**) for the words involving the rule that their colour-scheme matched, children should yield better spelling accuracies at post than pre-test for new or old and coloured or uncoloured words.

**Colour has transient attentional effects:** if the benefits of colour are limited to transient attentional effects (i.e., supporting children in inhibiting irrelevant information, and focusing on task relevant information), then (**CH2**): for the words involving the rule that their colour-scheme matched, children should yield better spelling accuracies at post than pre-test for new or old coloured words, but not uncoloured words.

**Colour promotes recall:** if colour helps children remember words that they learned in instruction (in addition to or without helping children abstract the underlying rules or more general structures (syllable types, vowel sounds)), then (**CH3**) for the words involving the rule that their colour-scheme matched, children should yield better spelling accuracies at post than pre-test for old coloured words than either uncoloured or new words.

I was present throughout the children's intervention sessions. To supplement my quantitative analysis, I collected audio recordings, informal written observations, interviews with the children's tutors and profiles of each child. These data were used in a qualitative analysis of how the characteristics of individual children affected their use of the system and the colour codes.

The next sections detail the research instrument (the software system PhonoBlocks), literacy activities and colour coding schemes. It also describes the design sessions I conducted with tutors to validate and refine my colour codes, rules, activities, research questions and methodology.

## **3.2. Research Instrument Design**

### **3.2.1. PhonoBlocks**

PhonoBlocks is a tangible software system that my lab and I developed in concert with a team of expert tutors from the Kenneth Gordon Maplewood School program for children with dyslexia. My colleagues and I authored a short paper that describes PhonoBlocks' tangible features in greater detail (Antle, Fan & Cramer, 2015). PhonoBlocks' tangible features are not the focus of my thesis. Here, I briefly describe them, but shall not detail their rationale.

#### ***Hardware and Physical Interface***

PhonoBlocks is a software system that uses tangibility and dynamic colour codes to support tutor-driven Guided Discovery and student-driven practice of Orton Gillingham styled multisensory learning activities (section 2.4.5). PhonoBlocks consists of 46 clear acrylic letters and a platform. The platform has seven slots in which the letters can be placed. Each letter sits on a base. Each base contains a unique combination of POGO pins which serves as the letter's ID.

Users interact with PhonoBlocks by placing letters on the platform. The platform houses an Arduino mega microcontroller which senses the placement of the letters in the slots and identifies the letter given the pattern of pin activation. The microcontroller communicates with a custom software application that I developed for PhonoBlocks and implemented in C# in the Unity game engine. The application displays the letters on-screen, determines their sounds, and colours them accordingly.

Each tangible letter contains an RGB LED strip. The LED strips supported 6 colours: red, yellow, green, blue, cyan and magenta. By changing the energies at each of the RGB channels, the software application causes the letters to glow the same colour as their onscreen counterparts.



**Figure 3-1. The integrated screen and tangible interfaces.**

### ***Software and Digital Interface***

I wrote an algorithm that recursively<sup>4</sup> partitions the onscreen word into syllables<sup>5</sup>. It infers the sounds of vowels by the syllable in which they appear. It identifies whether the last three letters are a consonant-le syllable. Each time the user changes the onscreen letters, the algorithm re-interprets their sounds. The colours of the letters depend on their sounds. The software enables administrative users to set the active colour-coding scheme. The software re-colours the onscreen and tangible letters according to their sound and the active colour-coding scheme.

<sup>4</sup> Based off Tony Hoare's "quick-sort": <http://en.wikipedia.org/wiki/Quicksort>

<sup>5</sup> The algorithm applies the most frequent syllable division patterns, as tutors teach young students to do. It functions as a research prototype for the specific set of words that we used, which follow the most frequent patterns, but it cannot handle exceptions.

The software supports some additional multi-touch interactions that I used (described in section 3.3.4) to support children's learning during the experimental sessions. Left-swiping onscreen letters de-activates them. Right-swiping re-activates them. The algorithm discounts de-activated letters when it calculates the sounds. It re-calculates the letters' sounds and colours after each de-activation/re-activation.

The software uses open-sound control to communicate with google text to speech, which enables it to "read" character strings. Tapping any active letter causes the system to read all active letters as though they were a word.

The screen shows a "submit" button and a vertical "Word History" widget. Children tap the submit button to send the current word to PhonoBlocks. Words are then displayed in their colours in the Word History. Tapping a word in the history causes PhonoBlocks to read it.



**Figure 3-2.** The screen interface. The word the children spell appears in the middle of the screen. Tapping the "check" button submits the word. Completed words appear in the word history (upper right).

### **3.2.2. Literacy Rules and Rationale**

Testing my framework required two different literacy rules, each of which requires attention to different grains of information. The tutors that I consulted to validate and

refine my colour codes and study (3.2.4) provided me with a list of the concepts they taught. I selected my literacy rules from this list.

One task that children with dyslexia struggle with is discriminating short vowel sounds. For example, that the vowels in “pet” and “pat” or “pin” and “pen” differ. Short vowel discrimination involves attention to *fine-grain* information: the sounds of *particular* vowels.

Another task that children with dyslexia struggle with is mastering the consonant gemination rule when spelling words involving the consonant-le syllable. Literate Anglophones apply the rule unconsciously, but cannot articulate it. The rule applies to words involving consonant-le syllables, and involves a single spelling (or decoding) decision. Phrased as a spelling rule, it is:

One geminates (doubles) the consonant of the consonant-le syllable *if and only if* the vowel in the word sounds short.

For example, the “u” in “bubble” sounds short; there is a geminated “b”. The “a” in “stable” sounds long; there is a single “b”. We can rephrase it as a reading rule:

One pronounces the vowel short *if and only if* there is a geminated consonant.

For example, there is a geminated “d” in “cuddle”; we pronounce the “u” as *uh*. There is a single “g” in “bugle”; we pronounce the “u” as *you*.

Applying the rule requires attention to *large-grain* information: the vowel sound category, i.e. long or short. The vowel’s *particular* sound, i.e., whether *a*, *e*, *i* or *u*, is irrelevant.

Vowel discrimination and consonant gemination were therefore appropriate for my experimental comparison because they differed in one main respect, and that was the requisite grain of vowel-sound information.

## ***Spelling Or Reading?***

Both consonant gemination in consonant-le words and short vowel discrimination can be expressed as spelling or as reading rules, both of which involve attention to and manipulation of, respectively, vowel sound category and presence of an extra consonant, and particular medial vowel sound and identity. Reading and spelling differ in which piece of information is attended versus manipulated: reading involves attention to orthography and manipulation of sound; spelling involves attention to sound and manipulation of orthography.

I have no theoretical reason to suppose that colour would be differentially effective for supporting the acquisition of rules for reading versus spelling. Ideally, I would have included both measures, and explored this question myself. Unfortunately, I was limited in the amount of time my sample could dedicate to the study, and therefore in the number of practice and assessment words I could administer. Each word is “measures” a child's abilities. The reliability of any statistic depends on the number of measurements upon which the statistic is based. I therefore elected to devote all measurements to one ability, rather than split them between two, and obtain one more reliable than two less reliable statistics.

Although the bulk of my literature review focused on reading, I decided to measure children's spelling instead of reading. The tutors at KGMS suggested that their students' difficulties in reading manifested as high reading speed, but their difficulties in spelling manifested as low spelling accuracies. Increasing speed (fluency) is a distinct problem from increasing reading accuracy (Lovett, 1987), and many of the interventions I surveyed aimed at improving accuracy. Few interventions substantially impacted speed. I therefore had precedent to assume that my intervention would influence accuracy to a greater extent than speed and that measurements of accuracy would be more sensitive than measures of speed. Accordingly, because my sample was less accurate in spelling than reading, I predicted that assessments of spelling accuracy would be more sensitive than assessments of reading accuracy, (spelling accuracy had “more room” to vary), and thus provide a more valid indication of my intervention's effectiveness.

### **3.2.3. Colour-Coding Schemes and Rationale**

This section describes my colour coding schemes, which I designed with respect to three core cognitive requirements (2.3.4): support discrimination and segmentation of relevant multi-letter units, support associating letter-units and their sounds and support attentional focusing to relevant information and inhibition of irrelevant information, and the considerations (DE-4) that I outlined in my framework. Based on my framework, a coding scheme should highlight all and only information that is relevant to a specific literacy activity. Pursuing this goal obliged me to address my four design elements: which information is relevant, when to allow emergent patterns (versus unique colours) to highlight information, what dynamism should communicate and which specific colours to use. I rationalize my choices with respect to the considerations I outlined in my framework.

My chosen activities (3.2.2) required attention to particular vowel sound and vowel sound category, respectively. My coding schemes therefore differed in how they coloured the vowels. For both schemes, “free” consonants (consonants that were not involved in consonant-le syllables and therefore contained no task-relevant information) were white. As I noted in DE2, literature on the uses of colour to attract the attention of children with ADHD suggested that too much colour can overwhelm children (Zentall, Grskovic, Javorsky, & Hall, 2000). Colouring irrelevant consonants white was therefore expected to help children focus on relevant letters.

For both schemes, I addressed DE3 by implementing (for consonant-le words) simultaneous changes in the colours of consonant-le letters and vowel sounds and (for short vowel words) the immediate application of the vowel's colour. My rationale for implementing immediate and reliable (versus delayed or intermittent) changes was that (by the considerations I outlined under DE4), immediate and simultaneous changes better supported my objective of helping students acquire a solid foundational understanding of the relations between the doubled consonant and vowel sound or vowel identity and vowel sound. Consistent with the immediate and reliable changes, the word sets I used in my study contained no exceptions to the rules.

### ***Scheme 1: particular***

I designed the particular scheme to support children in discriminating short vowel sounds. Children with dyslexia struggle to acoustically segment sounds from words in continuous speech (2.3.1.3). My literature review suggested that presenting children with visual representations of the auditory distinctions might help (2.3.1.3, 2.4.1.1).

#### **Particular Vowel**

The relevant information (DE1) for short vowel discrimination are the phonemic contrasts between different particular vowel sounds. Lindamood approaches represented phonemic with colour contrasts. On the expectation that my sample might suffer phonological deficits which would impair their phoneme segmentation, I adopted the Lindamood approach, assigning to each vowel sound a unique whole colour (DE2). Phonemic contrasts were therefore represented by contrasts of different whole colours, which are easier to notice than differences in patterns of colours. In addressing DE4, how to map colours to information, I applied Wrembel's work on innate colour-vowel associations. I suspected that children would have an easier time associating both colours to sounds and then sounds to letters if I assigned the vowels colours to which children (might) intuitively associate vowels.

I focused on short vowel sounds because these tend to be the most confusable. Based on Wrembel's data, short "a" was red, "i" was yellow, "e" was green, "u" was cyan and "o" was blue. For methodological reasons which I describe in section 3.3.3, the letters of consonant-le units (in the consonant-le gemination activities) were magenta.

### ***Scheme 2: categorical***

I designed the categorical scheme to support children with dyslexia in mastering the consonant-le gemination rule. One objective was helping children see how the consonant-le gemination rule builds upon two facts they already possess about syllable types and syllable division: one, consonant-le is always its own syllable; two, vowels in closed syllables sound short; those in open syllables sound long. Appending consonant-le to a closed syllable results in a geminated consonant (cud+dle="cuddle"). Appending consonant-le to an open syllable does not ("bu"+"gle"="bugle"). The relevant information

(DE1) are the sound category (short or long) of the vowel in the dictated word, the presence/absence of the consonant-le syllable and the presence/absence of an extra consonant between the vowel and the consonant-le syllable. In line with guided learning theory (Mayer, 2004), I supposed that noticing the presence of the consonant-le syllable would activate children's memory of the rule about dividing words around consonant-le, while noticing the additional consonant was supposed to explicate the type of the remaining syllable (open, no additional consonant; closed, an additional consonant), and activate children's memory of the relations between syllable type and vowel sounds.

To help children notice the relevant information, I applied the following scheme (DE2, DE4):

Short versus long vowels had unique whole colours: short vowels were yellow. Long vowels were red. The letters of consonant-le units were coloured differently from letters that were not part of a consonant-le unit, but, excepting the silent-e, each letter of the unit was coloured the same (magenta). The silent-e had the magenta hue but was 50% darker than the consonant and "l". I did not assign a unique colour to the additional consonant. Like other "free" consonants, it was white. My justifications follow:

### **Short and Long Vowels**

My vowel colours were warm (red and yellow). Identifying a syllable type presupposes identifying the syllable. Tutors teach children to identify syllables by looking for vowels. Vowels therefore warranted a visual representation that was attentionally primary. With regard to DE2, I decided to map unique whole colours to short versus long vowels (versus an alternative, colouring consonants, leaving vowels white, and allowing short vowel to equate the pattern white-colour and open to equate the pattern colour-white), because differences in single colours (e.g., red versus yellow) are easier to notice than differences in colour patterns (colour-white versus white-colour). I choose warm colours because the majority of foveal photoreceptors are sensitive to long and medium wavelengths (Smith & Pokorny, 1975). I therefore expected that warm colours would be easier to focus to than cool ones. In addition, warm vowel colours represented how vowels are the "hot" acoustic energy peaks.

Although the use of warm colours was consistent with a colour-vowel association that tutors sustained (consistent with traditional OG approaches, vowel cards were salmon and consonant cards were white), the tutors did not consistently represent short versus long vowels with a specific colour contrast. I therefore sought an independent, third association by which to map vowel sound category to warm colours (DE4). The warm colours red and yellow correspond to “long” and “medium” wavelengths. I learned that the children who were prospected for participation in my study knew about the electromagnetic spectrum, and how yellow corresponds to a shorter wavelength than red. Wagner and Torgesen’s point (Wagner & Torgesen, 1981) about how mnemonics that are difficult to remember can increase working memory load encouraged me to choose colour-sound assignments that “made sense” in the context of the children’s pre-existing knowledge. I suspected that assigning red (long wavelength) to long vowels, and yellow (shorter wavelength) to short vowels would be easy for the children to remember.

### **Consonant-Le Units**

I coloured the letters of consonant-le syllables the same, and their colour was unique. Consonant-le gemination requires children to recognize and “segment away” the consonant-le syllable, and recognize the extra consonant that causes the short vowel sound. Berninger and Hines helped children recognize and group multi-letter units by colour-highlighting them. I therefore suspected that uniformly colouring consonant le letters would help children recognize and group them.

With respect to RQ1, I had assumed that appreciating the *presence* (versus the definition) of the consonant-le syllable was relevant, and so the colour-scheme should highlight the presence but not distinguish the specific roles of each letter of the consonant-le unit. My subsequent consultations with the tutors somewhat contradicted my assumption and led to the final design, in which the silent-e is darker. The tutors explained that, although the students had a partial grasp of the consonant-le syllable, they sometimes erred in neglecting the role of silent-e in producing the “vowel” sound (/uh/), leading to such misspelling as “stabul” or “stabl”. My discussions with the tutors suggested that- while the role of the initial consonant and l was well-established, the role of silent-e constituted an additional piece of *relevant information*. The tutors and I agreed

upon two aims- that the scheme should remind children of the presence and role of silent-e, (distinguishing silent-e as unique), whilst helping children group it with the consonant and l. Depending on the child, learning a rule may involve several different pieces of relevant information; by my framework (DE1), designers should attempt to highlight all relevant information. To highlight both pieces of relevant information, I assigned silent-e an equivalent *hue*, but a darker shade. Because perceptual grouping-by-hue can occur despite differences in illumination (Ware, 2012), darkening the “e” should not prevent children from appreciating its membership in the consonant-le unit. In addition, I thought the darker shade might bring to mind a “shadow” or “spectre” (substantively absent), and that either of these might naturally associate to a “silent” (phonetically absent) letter<sup>6</sup>. Magenta was chosen for methodological reasons that I describe in section 3.3.3.

### **Presence (versus absence) of Additional Consonant:**

My framework states that designers should minimize the number of unique colours. In this case, a unique colour would have been warranted if a) the presence/absence of an additional consonant entailed no salient visual differences b) there was reason to distinguish the additional from other (the initial) consonant, and c) no reason to *group* the additional and initial consonant.

Taking these into consideration, I decided that colour-highlighting the additional consonant was not only unnecessary, it was potentially contrary to my colour-schemes' overall goals. First, because consonants were white, but vowels and consonant-le units were coloured, the presence of the additional consonant already entailed a salient visual difference (presence/absence of a white element between the coloured vowel and consonant-le unit).

Second, one of my scheme's objectives was helping children activate previously learned knowledge about syllable types (i.e., recognizing that doubling correlates short vowel *because doubling produces a closed syllable*). The definition of the closed syllable to which my students were accustomed represents closed syllables as “CVC”. This

---

<sup>6</sup> The tutors suggested colouring the silent-e as a “ghost” (pale white, with a perturbed outline), but I thought this might represent too great an obstacle to perceptual grouping.

representation visually equates between the initial and final consonant (both are represented by “C”). Colouring the final consonant differently than the initial might suggest that the sub-word- initial consonant, vowel and doubled consonant- is somehow different from other closed syllables, which is the opposite of what I intended: that children should recognize it as another instance of the closed syllable category.



**Figure 3-3.** How the consonant-le (top) and vowel discrimination (bottom) words appeared in the categorical (right) and particular (left) colour-coding schemes.

### **3.2.4. Tutor Design Sessions**

#### ***Methodology***

To validate and refine the design of my colour codes and study, I conducted two interviews with four expert tutors at Kenneth Gordon Multisensory School. All tutors were female, and had a median of 20 years' experience teaching OG-style multisensory curricula children with dyslexia. Each session lasted about one hour and began after the regular tutoring hours, at about 3:30PM.

The sessions were informal and aimed to elicit the tutors' expertise in a) the literacy rules that are challenging to children with dyslexia b) the kinds of attentional supports and activities tutors use to help.

During the first session I asked the tutors a) if the children under their tutelage had difficulty with short vowel discrimination and consonant gemination and b) how they supported children. I asked the tutors to provide me example worksheets, on which I

based the PhonoBlocks learning activities. Between the first and second session I developed preliminary versions of the activities in PhonoBlocks. During the second session I showed tutors a play-through of the activities and asked them whether it matched their everyday practice.

## ***Results and Revisions***

### **Consonant-LE Geminatio**

The tutors confirmed that children under their tutelage struggled with consonant gemination in consonant-le words. The tutors validated my assumption that: children understand the meaning of the long and short vowel categories and that these categories are part of the children's metalinguistic curriculum. They confirmed that children's difficulty mastering consonant-le spellings involves consonant gemination.

To support children spelling consonant-le words, the tutors explained that they provide children with a mnemonic that involves vowel sound category: "the consonant brings his brother when the vowel sounds short". The tutors showed me an enactive technique in which they draw a box around the consonant-le stable syllable to reveal the type of the preceding syllable, either closed (vowel-consonant) or open (vowel). They provided me with a sample worksheet with which children practice consonant-le spellings. The worksheet consisted of initial consonant-vowel or vowel letters, which the child completes by adding the consonant-le unit and extra consonant, if necessary. Tutors read the word to children, so answering correctly requires children to attend to vowel sound category and remember and apply the gemination rule (e.g., for initial letters "bu", the tutor could say "spell *bugle*" or "spell *bubble*"). Because the worksheets provide children the initial letters of the first syllable, they focus children on the spelling decisions to which the rule pertains.

### **Short Vowel Discrimination**

The tutors validated my assumption that their students had difficulty discriminating short vowels, despite being introduced to them early in the curriculum. They specified that short "e" and "a", "i" and "e", and "o" and "u" were confusable.

The tutors confirmed that children's difficulty correctly spelling short-vowel words was identifying the vowel in continuous speech. Some children also struggled to recall sound-letter correspondences. The tutors showed me a multisensory support for reading that involved sequentially enunciating and "touching" each letter in a word, and one for spelling that involved enunciating each letter's sound as they wrote it. For example, they would say "spell *went*. *W- eh-n-t*". The objective was helping children segment and match sounds and letters. The tutors provided me with worksheets with which children practice short vowel discrimination. They were similar to the consonant gemination worksheets. They provide children with partially completed initial words. Children complete the word by adding a vowel. Tutors indicate the vowel by reading the words to children. For example, for the initial letters "p n", the tutor could say "spell *pin*" or "spell *pen*". The worksheets focus children on the spelling decision to which vowel discrimination pertains.

### **3.3. User Study at Kenneth Gordon Multisensory School**

#### **3.3.1. Study Design**

This section overviews my experimental procedure. My study had two variables: colour-coding scheme and activity type. Colour-coding scheme had two levels (categorical and particular) and was compared between-groups. Activity type had two levels (consonant-le and vowel discrimination) and was compared within-groups.

The study occurred over four weeks. Three times each week, every child used PhonoBlocks to complete four spelling activities. Two of these activities involved short vowel discrimination; two involved consonant-le consonant gemination.

Each child experienced only one colour coding scheme. Each scheme matched and mismatched one activity (i.e., the particulars scheme matched vowel discrimination but not consonant gemination; the categorical scheme matched consonant gemination but not vowel discrimination).

Before and after the experimental sessions, children completed an assessment of their abilities to spell consonant-le and short vowel words using PhonoBlocks, with and without the colours of the child's scheme.

I predicted an interaction between children's colour coding scheme and activity with respect to their pre-post test improvement. I expected children with the particulars scheme to improve more for short vowel discrimination than consonant gemination but for children with the categorical scheme to improve more for consonant gemination than short vowel discrimination.

Although I used an experimental paradigm, the small number of participants (10) warranted a supplementary, individualized cases approach. In addition to the between groups comparisons, I conducted qualitative observations of each child, interviews with the children's tutors and profiles of each child. I analyzed these to better understand how individual characteristics affected children's responses to the colour coding schemes.

### **3.3.2. Participants**

I recruited my participants through the tutors. The tutors recommended students who had difficulties with both vowel discrimination and consonant gemination. I administered consent forms to the parents of the children that the tutors recommended. The consent forms explained the rationale for the study and informed parents that, should they consent for their child to participate, their child would spend roughly 1 hour per week of their tutoring time using PhonoBlocks. Children whose parents consented for them to participate were administered a verbal assent script, read by their tutor.

Children who verbally assented to participate in the study formed the final group of 10 children. Five were male; five were female. 9 children had normal colour perception as measured by the Ishihara colour-blindness task<sup>7</sup>. One child presented red-green colour blindness (protanopia). This child was assigned to the categorical scheme, but I adjusted the colours to compensate for his protanopia. For this child, short vowels were blue, long vowels were yellow, consonant-le units were green.

---

<sup>7</sup> <http://www.colour-blindness.com/colour-blindness-tests/ishihara-colour-test-plates/>.

I randomly assigned the remaining children to one of the two colour-coding schemes. Originally, I had eight children and assigned four to each scheme. The 9<sup>th</sup> child was the protanope, and I assigned him to the categorical scheme. The 10<sup>th</sup> child began the study late. At the time, a child from the categorical group suggested she would terminate the study. In expectation of her attrition, I assigned the 10<sup>th</sup> child to the categorical scheme. The other child subsequently decided to continue the study. By this time, the 10<sup>th</sup> child had undergone pre-assessment and a session with the categorical scheme, and I decided that it was inappropriate to transfer her to the particulars scheme. I stand by my decisions although they unbalanced the design (six children, to the categorical scheme; four to the particulars scheme).

### ***Child Profiles***

To understand what factors might influence children's successes with PhonoBlocks, I compiled profiles of each child that included demographic variables (gender, age, ethnicity and the parents' levels of education), the child's academic history, dyslexia profile and other challenges or skills that might pertain to the colour codes. I asked about the child's academic history, proficiency and interest because I suspected that these would affect children's enthusiasm for reading and spelling and for using PhonoBlocks. I asked whether the child enjoyed artistic pursuits or demonstrated proficiency in other areas of visuo-spatial reasoning, i.e., playing with blocks or KNex, assembling models, geometry, because I suspected that the usefulness of the colour-which involves ventral temporal visual areas- might depend on the child's visuo-spatial or artistic abilities. I asked whether children enjoyed or struggled with mathematics or music because these disciplines involve mental symbol manipulation, and I suspected that one use of the colour-codes might be relieving children of the need to maintain acoustic information (vowel sound type, syllable type) in working memory. Finally, I asked whether the child had attentional or other behavioural challenges that might affect their motivation or ability to use PhonoBlocks.

### 3.3.3. Activities, Colours, PhonoBlocks

This sub-section describes how I implemented the tutors' worksheets (3.2.4.2) in the software and the methodological rationale for two additional system features, consonant-le colours and flashing colours.

#### ***Worksheet Implementation***

I based my activities off the worksheets the tutors provided. The worksheets were a fitting implementation because a) they had the mechanic of transforming one word into another by adding or changing letters, which enabled me to assess the impact of sound-based colour changes, and b) they were similar for consonant-le and vowel discrimination, in that, although they differed in the number of letters that children needed to add, both activities required children to spell only what pertained to the rule (either a vowel, or the end of an onset and consonant-le syllable). In consequence, the activities and my colour codes differed by the same variable: the requisite grain of information for attention.

First, the screen displays the initial, partially completed word. The letters are black. The system issues a voice command to “place the letters you see on the platform”. The initial letters for consonant-le words were the first consonants and vowel of the first syllable (e.g., “bu” for “bubble”; “sta” for “stable”). The initial letters for the vowel discrimination words were all letters excepting the vowel (e.g. “p\_sh” for “push”, “st p” for “stop”). Like the tutors' worksheets, this enabled the children to focus attention to the spelling decisions that were relevant to the activities' learning goals (consonant gemination or short vowel discrimination).

Placing the tangible letter in the corresponding slot causes the software to colour the onscreen and tangible letters. Placing the last initial letter causes the software to say which word to spell, for example: “make the word, *stable*”.

The main activity consisted of the child spelling a word using the tangible letters. The system required children to submit an entire word before receiving any feedback. This was intended to prevent children from approximating the answer through trial and

error. If the child's word was incorrect, the system said "That's not quite it... would you like a hint?"

After an activity was finished, the system added the completed word to a vertical "Word History" widget, which simulated the pedagogic device of paper and pencil multisensory curricula. The words appeared in their colours. The system then asked children to remove all the platform letters. Letters could be removed in any order. Removing the last letter began the next activity.

### ***Consonant-LE colours:***

I colour-highlighted consonant-le units in the categorical scheme (3.2.3.2). The consonant-le task involved two sub tasks: a) identify the syllable unit, consonant-le b) remember and apply the relationship between closed and open syllables and vowel sounds. Colour-highlighting consonant-le supports (a). I expected the categorical scheme's vowel colours to support (b). Children with the particulars scheme also had to complete consonant-le activities. My prediction was that colour codes would work by drawing attention to the relevant grain, but as it stood, the categorical scheme had two advantages (a and b) over the particulars scheme for consonant gemination. If the categorical scheme was superior for consonant gemination I would be unclear whether this attributed how it colour-highlighted the consonant unit (which Berninger and Hines' work suggested would be helpful) or how it coded the vowels (which no one, to my knowledge, has assessed). To focus my comparison on how the schemes coded vowel sound, I colour-highlighted consonant-le units in both of the schemes.

The consonant-le units were magenta because a) the colour had to be similar between the particular and categorical schemes b) the particulars scheme assigned the other colours to vowels. As my framework (DE4) recommends, I investigated any potential outstanding associations my sample might have between consonant-le units and a particular colour. The tutors denied using any particular colour to represent consonant-le. Lacking any precedent for consonant-le units to be coloured a different way, I consider my colour-assignment acceptable.

### ***Letter flashes***

One mechanism by which I suspected colour might help was focusing children's attention to relevant linguistic information, i.e., the letters that change. Another mechanism by which I suspected colour would help was by showing the relevant grain of information. As it stood, the schemes differed in both the encoded informational grain and number of visual changes. In consonant-le activities and the categorical scheme, the vowel colour changed as syllables changed between open and closed, but because the vowel did not change, the colours would not change in the particulars scheme. In vowel discrimination activities and the particular scheme, the vowel colour changed if children changed the vowel, but because all vowels were short, the colours would not change in the categorical scheme. If the categorical scheme were superior for consonant-le activities or the particular for vowel discrimination, it would be unclear whether this attributed differences in the coded grain of information or the number of visual changes. To focus my comparison on the difference in coded information grain, I equated the number of visual changes between the coding schemes. Each time the sound at a position changes (either because the letter sound or letter changes) the new letter flashes on and off.

#### **3.3.4. Procedure**

Each child experienced 12 experimental sessions: three sessions per week for a period of four weeks. A session consisted of four activities (3.3.3.1), two per activity type, blocked by activity type, with alternating presentation order.

Each activity involves one word. I selected words for each pair of activities (within a particular activity type) that highlighted the activity's challenging contrast. For consonant-le activities, one word involved a long vowel (no consonant gemination); the other involved a short vowel (consonant gemination): for example, "rifle" and "ruffle". For vowel discrimination activities, the two words involved two confusable vowels. Paired words involved one of: short e and short a, short e and short i, short o and short u. Appendix B contains the set of words used in the PhonoBlocks practice activities.

I selected vowel discrimination words from an online repository<sup>8</sup> of the top 60,000 most frequently occurring English words; all words had frequency ranks greater than or equal to 6000. All vowel discrimination words had four characters and were monosyllabic. Because each word contained exactly one vowel, each word contained a consonant digraph or blend, appearing in the initial or final position. I excluded words with final consonants “r”, “l” or “w” because their vowels do not sound short. I selected frequent monosyllabic words because these are what the tutors used. The tutors approved my final set of vowel discrimination words.

I selected consonant-le words from a list that the tutors provided. I excluded short vowel consonant-le words involving blends or digraphs because they were too long to spell with the system. Three long vowel consonant-le words involved blends (cradle, staple and trifle). My consonant-le syllables used six different consonants (“ple”, “gle”, “dle”, “ble”, “fle” and “tle”). When possible, to better focus children’s attention to the correlated vowel sound and doubled/single consonant, I matched the consonants of the session’s two words. I had sufficient words to match consonants for the first six sessions. The consonants of the words for the remaining sessions did not match, but I equated the overall frequencies of each consonant.

Each child experienced the same order of word-*pairs*, (e.g., the first session involved “rubble” and “table”) but the order of words *within* the pair was randomly determined (e.g., P1 might experience “table” then “rubble” but P3 might experience “rubble” then “table”). A predictable order would have enabled children to memorize and apply a response sequence (i.e., single then double consonant; e then a) without understanding.

Sessions occurred during the child’s tutoring blocks. With two exceptions, the child’s regular tutor was present throughout the sessions. One tutor was on holiday during the study; two substitutes replaced her. The tutor was responsible for two children, one each from the categorical and particular schemes. For these children, a substitute was present for the first, practice session, but the substitutes did not attend the remaining sessions. The substitutes believed that the children would concentrate

---

<sup>8</sup> <http://www.wordfrequency.info/>

better if they were absent. Similarly, tutor T1 absented herself for the latter five sessions. She was responsible for three children, each from the categorical scheme. Like the substitutes, T1 determined that her children focused better when she was absent.

I used the first session to show the children how to complete the activities and what the child's colour scheme signified. The first session consisted of a regular activity that the tutor and I completed with commentary that drew children's attention to the colours, colour changes, what these signified and how the children could exploit the colours as learning supports. We followed a script that I wrote. Appendix A contains the script.

For subsequent sessions, the tutor and I provided children feedback for their mistakes. We provided feedback only after the child submitted a word. If the child erred, tutors provided the same hints that they would during regular tutoring. The content depended on the child's error. If the error concerned the activity's main learning content (consonant gemination or an incorrect short vowel) I used the system to supplement the tutor's hint. Children were then allowed to change and re-submit the word. As before, we provided feedback after the child submitted the word, never while they spelled it. Feedback changed only if the child's error changed (e.g., for the target "stubble", the first submission was "stubel"; the second "stuble"). I allowed children as many attempts as necessary to spell the word correctly.

For sessions where the tutor was absent, I repeated the feedback the tutor provided during the first session. The loss of one tutor to a holiday did not threaten the internal validity of my study because one each of her children had the categorical and particular schemes. The loss of T1 threatened internal validity because all of her children had the categorical scheme. On the other hand, if T1 was correct, then the heightened distractibility of her children would also threaten internal validity. I deferred to T1's opinion. To minimize the threat to internal validity, for sessions after T1 absented herself, I provided the tutor's feedback- again, repeating what the tutors said during the earlier sessions- to children whose tutors were physically present.

In consonant-le gemination activities, I de-selected the consonant-le unit and played the sound of the remaining syllable. I then asked the child whether the vowel had

the correct sound. I then re-selected the first consonant from the consonant-le syllable (turning the word into a closed syllable) and read it. If the child had the categorical colour scheme I reminded the child of the colour's significance. For example: "remember that long vowels are red and short vowels are yellow. What colour is your vowel?" In vowel discrimination activities, I de-selected all letters except the vowel and the following consonant, and played its short sound. Then I repeated the target word's short vowel sound and asked the child whether they sounded the same. If the child had the particulars scheme, I reminded them that each vowel had a different colour and the colour of the target vowel. For example: "remember that short e is green. What colour is your vowel?"

The hints the tutors provided differed superficially, but shared several core features. For consonant-le words, if the child erred in consonant gemination, tutors gestured syllable division by placing their hand between the first letter of consonant-le and the last letter of the first syllable. The tutor then asked the child whether the preceding syllable was open or closed. If the child erred in spelling the consonant-le syllable, tutors made a hand gesture that represented consonant-le, and emphasized the syllable sound (e.g., *tull*). For vowel discrimination words, children could only err in placing the wrong vowel. Tutors repeated the target word, enunciating each sound, and asked the child to repeat the vowel sound.

I allowed tutors to provide their regular hints because a) the system's projected use context was the classroom and b) the children were accustomed to these hints and I thought that forcing tutors to provide a standardized hint might confuse the children.

After the child submitted both words for a particular activity, I used the Word History to review the learning content. For the first six sessions, I reinforced the contents. I played each word's sound and drew children's attention to the relevant differences in the words' sounds and orthographies. When applicable, I reinforced the significance of the colour codes. For consonant-le words, I said (for example, "stable" and "rubble"):

*"In stable, the vowel sounds long. In rubble, the vowel sounds short. In stubble, we have doubled the consonant. When I remove consonant-le I*

*have a closed syllable. In stable, we haven't doubled the consonant. When I remove consonant-le, I have an open syllable."*

If the child had the categorical scheme I added:

*"And in stubble, the vowel is yellow because it sounds short. In stable, the vowel is red because it sounds long."*

For vowel discrimination activities I said, (for example, "push" and "stop"):

*"The uh sound corresponds to u. The aw sound corresponds to o."*

If the child had the particulars scheme I added:

*"And the u is cyan, and the o is blue."*

For the final six sessions, I asked the children to explain why their submission was correct, i.e., whether they had abstracted the rule that the colours conveyed. If the child provided no adequate explanation I reinforced the contents, as in the early sessions.

For all sessions, I placed the initial letters and removed all letters. Pilot testing revealed that the tangible letters broke easily. I therefore wished to minimize children's contact with the tangibles.

### **3.3.5. Data Collection**

#### ***Quantitative Measures***

#### **Pre and Post Assessments**

#### **Assessment Procedure**

I designed the pre and post assessments to explore my two research questions: are colour codes that highlight strictly relevant information more effective for learning a particular literacy rule than those which highlight less or more information? And will performance transfer to uninstructed or uncoloured letters?

I assessed the children at their school. Assessments required children to spell consonant-le and vowel discrimination words using PhonoBlocks. I assessed each child twice, once before and once after the four week study.

Each assessment was similar to a PhonoBlocks activity. The system displayed an initial word. As in the experimental sessions, I placed the initial letters. The system then asked children to spell the target word. Children assembled the word and submitted their response by pressing the on-screen “check” button. PhonoBlocks recorded whether the response was correct. The system provided no feedback and did not add the words to the history.

My assessment activities were similar to the “spelling-by-dictation” assessments that others have used to measure improvement in children’s spelling abilities (Bhattacharya & Ehri, 2004; Berninger, Vaughn, Abbott, Brooks, Begayis, Curtin & Grahm, 2000). My assessment differs in that children only had to spell the portion of the word that pertained to the rule it involved, versus the entire word. My intervention only focused on gemination and consonant-le syllable formation, and vowel discrimination. Because children’s ability to read and spell certain irrelevant letter-units (e.g., consonant blends) differs according to their position or identity (Bruck & Treiman, 1990) and possibly other ways that I could not predict, I worried that forcing children to spell entire words would increase statistical noise. For this initial exploration into the potential uses of colour codes, I wished to focus my assessments on the specific skills that my intervention addressed. That is why I did not make children spell the onset (for consonant-le words) or rest of the short vowel words.

### **Assessment Words**

Each assessment involved 32 words. 16 were coloured. 16 were white. On the post-test, 8 coloured and 8 uncoloured words were new; the rest had appeared in the pre-test and the experimental sessions. The assessment factors were test session (pre and post) familiarity (old or new word) and colour (coloured or uncoloured). Crossing them yielded six sets. Each set contained: two short and two long consonant-le words and one e/l, one o/u and two a/e short vowel contrasts. I obtained new consonant-le words from the tutors’ list. I obtained assessment short vowel words from the same

online repository as for session words, applying the same selection criteria. Although children with dyslexia might show weaker frequency effects than children without dyslexia, children are typically better at spelling words that they have seen before (Share, 1999). Although word frequency is only a proxy for the familiarity of words to particular children, I equated the average frequency of the short vowel words in each set.

### **Counterbalancing**

Half the children per condition experienced the coloured words first; half the uncoloured words first. Within these groups, half experienced the consonant-le words first; half experienced vowel discrimination first. The order the words within sets was randomly determined, but the same for each child. On the post-test, old and new words alternated, but the old word always appeared first. I expected that children would have an easier time with old than new words, and that a difficult initial word might decrease children's confidence and worsen performance on subsequent words.

### **Analysis**

I analyzed pre and post and vowel discrimination and consonant le performance separately. Each child spelled 16 vowel discrimination and 16 consonant-le words. I split the words by the two test factors: were the words coloured or uncoloured (word appearance) and were they old or new? (word familiarity). Crossing the factors yielded sets of four words. Because all pre-test words were new, the only assessment factor was word appearance (coloured or not) and sets had eight words. For each child, I tallied the proportions of words (out of four or eight) that they spelled correctly. My main dependent variable was the proportion of words that children spelled correctly.

### ***The Use of Non-Parametric Tests, Medians and Effect Sizes***

Preliminary Shapiro-Wilks tests of normality on the distributions of overall accuracy, consonant-le and gemination errors revealed significant departures from normality (all  $ps < .001$ ), which warranted the use of medians and non-parametric comparisons. My research questions involved interactions, i.e., between activity type and colour coding scheme, or assessment word appearance, testing period and

familiarity. There are few non-parametric tests for assessing interactions. Interactions were assessed via multiple independent pair-wise comparisons. I evaluate each at the criterion necessary to compensate for the corresponding increases in type one error.

I calculated effect sizes ( $r$ ) for the Wilcoxon between-groups tests by dividing the test statistic ( $Z$ ) by the square root of the sum of observations (Pallant, 2007), and for the Wilcoxon signed ranks tests by dividing the test statistic  $W$  (sum of the signed ranks) by the sample rank sum<sup>9</sup> (Kerby, 2014).

### ***Comparison A: the overall effect of colour***

To assess the overall effect of matched colour coding scheme (RQ1), I calculated and compared the median proportion correct for consonant-le and vowel discrimination words between the categorical and particular scheme.

### ***Comparison B: “simple” effects of colour (transference and mechanisms)***

To assess transference and the mechanisms of colours’ effects, I compared median proportion correct (within consonant-le and vowel-discrimination words, and within the particular and categorical scheme) for coloured versus uncoloured and old versus familiar words.

CH1 was that colour would benefit children by promoting deeper encoding and understanding of the rule. If children understand the rule then they should perform equally well at post-test on new and old and coloured and uncoloured words. Each child spelled two coloured and two uncoloured words of each type (vowel discrimination or consonant-le), one familiar and one unfamiliar, and two familiar words of each type, one coloured and one uncoloured. To assess CH1 I compared the pre and post assessment median accuracies of the two coloured versus two uncoloured words between the two groups, categorical and particular, separately for consonant-le and vowel discrimination words. I then performed the same comparison on the median accuracies of the two instructed versus two uninstructed words. CH1 predicts that *any improvement in pre and post spelling accuracies observed for words that match a child’s colour coding scheme* should be equally strong for coloured and uncoloured and instructed and uninstructed

---

<sup>9</sup> For a sample of size  $n$ , the sample rank sum equals  $(n*(n+1))/2$

words. In other words, if children with the categorical scheme yield greater pre/post improvements in accuracy for spelling consonant-le words, that improvement should exist for coloured and uncoloured and uninstructed and instructed words. Likewise, if children with the particular scheme yield greater pre/post improvements in accuracy for spelling vowel discrimination words, that improvement should exist for coloured and uncoloured and uninstructed and instructed words. Note that although effects of either word colour or familiarity would falsify CH1, failing to find effects of either word colour or familiarity does not *confirm* CH1; null effects are always attributable to low power. My qualitative measures were partly intended to explore to what extent a null result (failing to falsify CH1) resulted from children's comprehending the rule (consistent with CH1), versus methodological limitations.

CH2 was that colour would benefit children through transient attentional effects, i.e., by directing attention to a relevant unit or piece of information, or possibly by cueing a general memory of how a rule works. If colour works by immediately cueing attention or memory to relevant information, then children should perform worse on uncoloured than coloured words. Colour codes that involve learning, such as the unique vowel sound or category colours, predict that benefits for coloured words should only affect post-test accuracies (because at pre-test, children would not know the colour codes' meaning). Colour codes that involve lower level perceptual or other effects, such as grouping or feedback, might cause benefits for coloured words as pre as well as post-test. CH2 therefore involved the same comparisons as CH1, (albeit without the additional comparison of instructed and uninstructed words). CH2 predicts that *any improvement in pre and post spelling accuracies observed for words that match a child's colour coding scheme* should be weaker for uncoloured than coloured words. In other words, if children with the categorical scheme yield greater pre/post improvements in accuracy for spelling consonant-le words, that improvement should exist for coloured words only. Likewise, if children with the particular scheme yield greater pre/post improvements in accuracy for spelling vowel discrimination words, that improvement should exist for coloured words only.

CH3 was that colour would benefit children by cueing memories of words that they had seen before. In contrast to CH1 and CH2, CH3 leaves open the possibility that

a mismatched colour scheme might also provide children some mnemonic benefits, however, CH3 is specific to post-test accuracies. CH3 predicts that children learn words better when they are coloured. It is possible but uncertain whether matching the linguistic variables relevant to spelling a particular word (i.e., the colour scheme) matters. CH3 predicts that post-test spelling accuracies for either type of word should be better for coloured familiar words than any other type of word. If the colour scheme matters, then the categorical group should show this effect for consonant-le and not vowel discrimination words; children in the particular group should show the opposite pattern.

***Comparison C: factoring in “type” of spelling mistake for consonant-le words***

Although vowel discrimination words only allowed one error (mistaking the vowel); consonant-le words allowed two errors. These were:

1. **Consonant-Le Formation error:** Children could fail to create the consonant-le syllable (e.g., misspell the “ble” sound in “stable” as “stabl” or “stabul”).
2. **Consonant Gemination error:** children could fail to properly geminate (e.g., “stuble”, not “stubble”; or “stabble” not “stable”).

Both coding schemes had the feature (highlighting the consonant-le unit) that I designed to mitigate consonant-le formation errors. Conversely, only the categorical scheme had the feature (colour coding short versus long vowels) that I designed to mitigate gemination errors.

I thought it possible that the schemes would only differ in the reduction of the of gemination errors. I wrote a program that interpreted and recorded the type of any consonant-le word spelling mistake and used it to analyze and record the children’s errors. I then summed these errors per child per factor-defined set (i.e., across pre and post, word familiarity, word appearance, etc).

To assess whether the categorical scheme selectively improved children’s gemination errors, I performed comparisons (A) and (B) separately for consonant-le syllable formation and gemination error frequencies. Strictly speaking, the hypothesized differences between the particular and categorical groups with respect to any of the

colour hypotheses (CH1-CH3) should only characterize consonant-gemination errors, because the schemes only differed in how they coded the vowels. Patterns of transfer that characterize consonant-le formation errors should exist in both the particular and categorical groups.

### ***Supplementary Individual Case Analyses***

My literature review and discussions with the tutors suggested that individual children might vary widely in their responses to the system and colour-coding schemes. My child profiles (3.3.2.1) confirmed that my participants had widely varying attentional and executive and visual skills, which might affect children's abilities to use the colour-codes. Wide individual variability warrants individual versus group-level analysis.

To assess whether individual children benefitted from the colour-coding schemes, I performed comparisons (A though C) an additional time, separately for each child.

### **Software Event Logs**

I programmed PhonoBlocks to record each input event, along with a timestamp and some additional information about the event. I intended the software event logs to provide a rigorous quantitative grounding for some supplementary questions that I had about how children might use the colour codes and change their behaviours over the course of the intervention. To that end, I chose events that I thought might reflect how colour hindered or helped, and which corresponded to the various supplementary questions that I expected to guide my qualitative analyses. The events were: changing a letter, submitting a word, or committing an "unproductive error" (removing an initial letter). The additional information differed between events. For single letter changes and errors, I recorded the new letter (which was a "blank" if a letter was removed), the position of the change, and the word that was modified. For submissions, I recorded the expected (correct) and submitted word.

Because my event logs also recorded the session, assessment word and group of the child, I could use them to answer the various supplementary questions that I expected to guide my qualitative analyses. The supplemental questions pertain to *how*

(versus whether) children used the colour codes. Below, I state and explain my supplemental questions and how I intended to explore them with the software event log data.

Because some of the error feedback (3.3.4) involved reference to the colours, I thought colour might help children benefit from error feedback (more frequent correct second submissions). Because the colour changes were bright and salient, I thought matched colours might help children engage (more pre-submission placement/removals). Because matched colours indicated sound, I thought matched colours might enable children to pre-check their responses (more pre-submission corrections). On the other hand: because colour changes can distract, I thought matched colours might increase off-task behaviour (more unproductive errors or placements). I prospected a unique hindrance in the categorical scheme. The initial letters always formed an open syllable. The vowel was always long (red). When forming a long-vowel consonant-le syllable, placing the consonant would turn the vowel yellow. The vowel would not return to red until the child completed the consonant-le syllable. If children understood that the vowel in the target word was long, the apparent change to short might motivate them to seek ways to make the vowel long. If they had trouble understanding the relationship between consonant-le and vowel sounds, they might change the word in other ways than completing consonant-le (more unproductive placements following the addition of the first consonant in consonant-le words).

## **Analysis**

I wrote a program that calculated, for each child, for each session, the number of submissions required before spelling the word correctly, and the errors that were committed on each preceding submission. My program also identified cases where children corrected a misspelled word before submitting it (and receiving system feedback), and identified “unproductive” errors (removing an initial letter). I used the program to assess whether matched colour a) helped children use feedback (respond correctly the second time) b) served as feedback, enabling children to correct mistakes before submitting the word, and to check whether particular children committed unproductive errors. Instances of “being misled by colour” were easier to check by hand. An event counted as “misleading by colour” if a child in the categorical scheme a)

engineered the correct vowel sound via a different orthographic change than doubling or removing a medial consonant (e.g., spelling “maple” as “mayple”) or b) placed the first consonant in consonant-le, then removed it in order to maintain the vowel’s long colour. The events were time-stamped. I cross checked candidate events that I detected from my event log analysis with my recordings of the session.

### ***Qualitative Measures***

I anticipated variability in my aggregate results, and I anticipated that only certain cases of children might yield data suggestive of a colour effect. Although I designed my quantitative metrics to help me disambiguate the mechanisms of colours’ effects, I acknowledged that qualitative observations might provide a richer explanation of whether and how any children used colour and why most children did not use colour. I therefore supplemented the quantitative assessments with two qualitative assessments: tutor interviews and unstructured observations and recordings. Because I intended my qualitative analysis to supplement my quantitative analysis, I assessed my qualitative data with regard to questions that ensued from my quantitative analysis.

### **Tutor Interviews**

I conducted two interviews with each tutor, one in the first or second and one in the final week. I did not interview the substitute tutors because a) they had less knowledge of the children and b) because they did not attend any sessions past the practice, they had less knowledge of the system. The interview questions appear in Appendix C. The first two questions were directed towards my first research question, whether colour codes that are designed to highlight information that is relevant to a task are selectively beneficial for mastering that task. The 3<sup>rd</sup> question was directed towards my second research question, whether the effects of colour would produce transferrable gains. The final questions were directed towards optimizing the design of PhonoBlocks, and assessing any possible interactions between the effectiveness of the colour codes and tangibility.

## ***Analysis***

I analyzed the tutor interview data after completing my quantitative analysis. My goal was seeking answers in the tutors' responses to my questions that might explain unexpected findings of the quantitative data. In the results section I present my summaries of the tutors' responses to my questions and validate my summaries with direct quotes.

## **Unstructured Observations and Recordings**

I recorded informal observations of each child. Although I did not formally structure my observations, they were guided by the same questions as my event logs- i.e., I intended my observations to supplement my main research questions- both of which queried *whether* colour helped and *whether* it promoted transfer- by understanding *how* the children who used colour used it, *why* colour did (or did not) help, and why the children who did not use colour did not use it.

To that end, I attended to how the children used the matched colours. To see whether colour provided an alternate form of feedback, i.e., whether the colours helped by visualizing an otherwise invisible property (sound), I attended to children's reactions to the colour versus their tutor's feedback, asked children and recorded whether children remembered what the colours meant and whether they attended to the colours spontaneously and what prevented them from attending the colours.

I also used my observations to explore my second research question, whether the colours helped children develop transferrable knowledge. Although children might understand a rule without being able to articulate it, children's ability to explain a rule they have learned predicts their ability to transfer knowledge to new problems or representations of the problem (Goldin-Meadow, Alibali & Church, 1993). To supplement my quantitative assessments of transfer, I recorded children's explanations of their submissions, once towards the middle and once at the end.

I recorded my observations by typing them into my laptop and with my Apple iPhone's video function. The children were initially wary of being recorded with the iPhone, so I typed my observations for the first six sessions. By the sixth session they

were comfortable with being recorded. I recorded the last six sessions with my Apple iPhone.

### ***Analysis***

I focused my analysis of my observations around three questions: did children remember what the colour codes meant? Did children understand the rules? And how did children use the colour codes? I describe general trends and supplement with direct quotes from my recordings.

### **Aggregate Analysis and Case Studies**

I describe the event logs and observations with regard to groups of children. I identify deviants to general trends but not describe them in detail. Following my group analysis, I perform an individual case study of each student. I summarize their profiles and behaviour as indexed by the event logs and my observations. I also provide their pre/post assessment data. If they were a unique case (from the event log analysis) I provide a direct transcript of the episode from my session recordings.

I conducted the detailed case studies to explore the interacting individual and environmental factors that might predict a child's ability to benefit from colour-code based spelling instruction. I intended this to help me understand how to improve our system and articulate general principles for designing and implementing effective colour-codes.

## **Chapter 4. Results**

### **4.1. Quantitative Analysis**

In this section I describe my analysis of children's pre and post-test spelling performance and software event logs. I divide this section into two subsections. First, I describe the pre-post assessments. Second, I describe the software event logs.

Throughout my analysis I refer to children and tutors by the convention: upper-case letter and number. Each child's letter is "P"; their number corresponds to the order in which the children came to see me during the day. Each tutor's letter is "T"; their number corresponds to the order in which I interviewed them.

#### **4.1.1. Pre and Post Assessments**

I divide this subsection into three main sub-sub sections. First, I describe the ceiling effects for vowel discrimination, which resulted in my dropping vowel discrimination from the study. Second, I describe aggregate trends. My aggregate analysis compares median pre and post accuracies, numbers of consonant-gemination and syllable formation errors, between the colour schemes and the different assessment word types. Third, I describe individual deviations from my aggregate trends, with respect to overall and transfer performance. I conclude by stating the questions that I formed from my quantitative analysis and that I used to guide my qualitative analysis.

##### ***Ceiling Effects for Vowel Discrimination***

The tutors suggested that the children had difficulty discriminating short vowel sounds, and that they would have difficulty completing words by choosing a short vowel (i.e., from "w nt", make the word "went"). The pre-test revealed this was false.

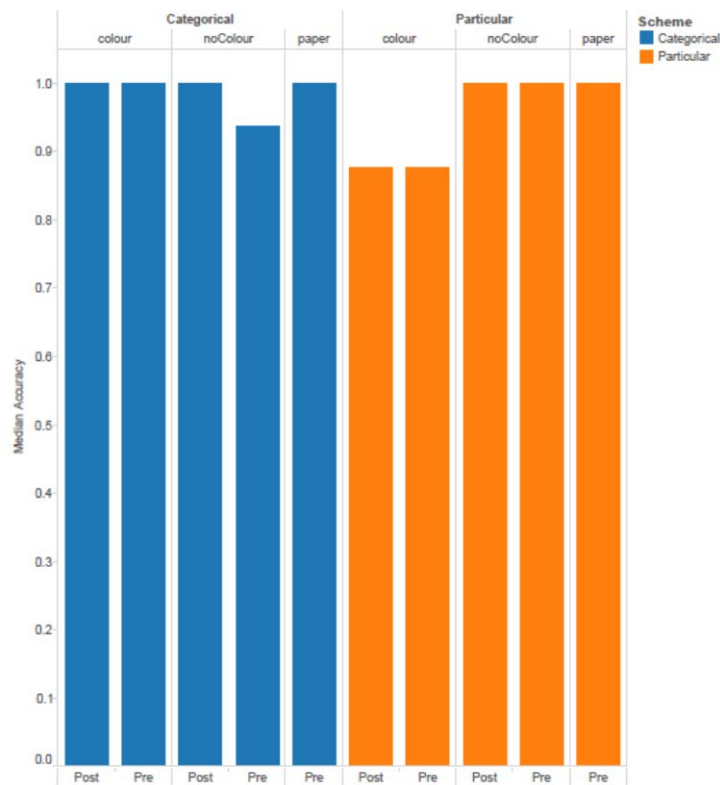
The median accuracy for short vowel discrimination was 94% (15/16 words). The lowest accuracy of any child was 92%. She misspelled two words, one coloured and one uncoloured. Overall, only five children scored below 100%. Four of these children had the categorical scheme; one had the particular scheme.

I was curious about the difference between the tutors' expectations of and children's spelling performance using PhonoBlocks. I thought it possible that children performed better with PhonoBlocks than they would with paper and pencil. To confirm that children were generally better short vowel discriminators and spellers than the tutors expected, I assessed children a second time with paper and pencil. Four paper and pencil assessment words were from the 'coloured' PhonoBlocks set. Four were from the uncoloured PhonoBlocks set. At the time of paper and pencil assessment, six children had already completed the PhonoBlocks assessment. I administered their paper and pencil assessments one or two days after the PhonoBlocks assessment. The remaining four children (two assigned to the categorical and two to the particular group) completed the paper and pencil assessment before the PhonoBlocks assessment.

To determine whether children's unexpectedly strong short-vowel word spelling indicated an effect of the system, I compared children's spelling accuracies for paper and pencil versus the coloured PhonoBlocks, separately for the two groups. The test yielded no evidence of a system-specific advantage (both  $ps > .9$ ,  $Md_{\text{categorical\_paper\_accuracy}} = 95\%$ ,  $Md_{\text{particular\_paper\_accuracy}} = 93\%$ ) and the effect sizes were negligible (both  $rs < .1$ ). There were no differences between children who spelled with paper and pencil before versus after with PhonoBlocks (both  $ps > .9$ ). The high accuracies were also consistent. Overall, only two children misspelled any word; both misspelled only one word.

Children's strong performance with short-vowel words continued into the study, with children virtually always spelling these words correctly on their first attempt. By the second week of the study I determined that having children spell the vowel discrimination words was not useful. I therefore conducted the post-assessment of vowel discrimination words at the end of the second week and thereafter only made children spell consonant-le words.

As expected, children's post-test vowel discrimination was good, with average accuracies of 91% and 93%, for the categorical and particular groups. There were no treatment effects for either group (both  $ps > .7$ ; both  $rs < .1$ ). There were no significant effects of word familiarity or appearance on children's pre or post-test accuracies, in either the categorical or particular group (all  $ps > .6$ , all  $rs < .1$ ). Figure 4-1 shows the pre and post assessment accuracies for short vowel discrimination, split by colour-coding scheme and word appearance:



**Figure 4-1. Pre and post accuracies for short vowel discrimination, split by scheme (categorical or particular) and assessment word appearance (coloured, uncoloured and paper and pencil).**

Vowel discrimination performance was therefore uninformative. This means that I am incapable of saying whether the particular scheme would be effective for learning short vowel discrimination. Replications with younger children who have yet to master vowel discrimination are needed to wholly assess my framework.

Fortunately for my study, children's spelling of consonant-le words was poor. The remainder of my analyses and subsequent discussion (Chapter 5) focus on consonant-le words.

### ***Performance and Transfer, by Group Medians***

#### **Pre and Post Performance**

In contrast to vowel discrimination, children's pre-test accuracies for spelling consonant-le words were low, with median accuracies of 21.9% (3.5/16) and 43% (7/16) in the categorical and particular groups. Children showed difficulty forming the consonant-le syllable and geminating the consonant. The median numbers of consonant-le formation errors were 9 in the categorical condition and 3 in the particular condition. In both conditions, the median number of consonant gemination errors was 8.

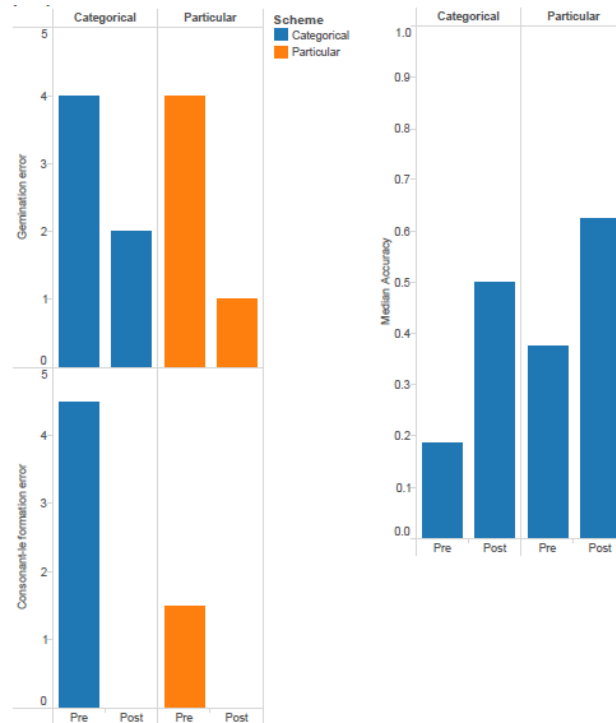
I first assessed RQ1, whether the categorical colour scheme would help children learn to spell consonant-le words. RQ1 predicted that children in the categorical group would yield superior gains for spelling consonant-le words than children in the particular group. My framework also predicted that the categorical group's relative gains would attribute decreases in consonant-gemination errors, because the categorical and particular schemes differed in how they colour-coded vowels.

A preliminary Wilcoxon-signed rank test on the grouped pre and post spelling accuracies approached but failed to achieve significance ( $W=29$ ;  $p<.07$ ), although the trend was for children's spelling to improve ( $Md_{pre}=31\%$ , about 9/32,  $Md_{post}=40\%$ , about 13/32;  $r=.64$ ).

To assess RQ1, whether improvement depended on colour-coding scheme, I calculated a set of "gain" metrics, equal to the difference of each child's pre and post scores. For example, each child's accuracy gain equaled their post accuracy minus their pre-accuracy.

A Wilcoxon two-groups rank-sum test on the gains for categorical versus particular schemes failed to support my hypothesis, finding no differences in the pre/post accuracy gains between the colour coding schemes ( $Md_{categorical}=14\%$ , roughly four

words.  $Md_{\text{particular}}=1.5\%$ , about “half” a word,  $Z(6,4)=.65$ ,  $p\sim.51$ ,  $r=.21$ ). Figure 4-2 (left) shows the median pre and post accuracies, split by group<sup>10</sup>:



**Figure 4-2.** Left, the median number of consonant-le and consonant gemination errors at pre and post-test, split by colour coding scheme. Right, median accuracies at pre and post-test, split by scheme.

Improvement could attribute decreases in consonant-le formation or gemination error. Either colour scheme might decrease consonant-le formation error, but only the categorical scheme was designed to decrease gemination errors. It seemed possible that a colour-scheme effect might present strictly for gemination error decreases, and be masked in the overall comparison. A Wilcoxon between-groups rank sum test on the gemination error decreases between the categorical and particular schemes failed to support my hypothesis, finding no differences between the groups' decreases of consonant-gemination errors ( $Md_{\text{categorical}}=1.5$  errors,  $Md_{\text{particular}}=0$ ;  $Z(6,4)=.59$ ,  $p\sim.52$ ,

<sup>10</sup> □ “Gains” appear greater on the graphs than what I reported for the statistical tests because they represent median pre and post accuracies, not median differences in each child’s pre and post accuracies. Because medians are calculated after re-ordering, but paired differences are calculated before re-ordering, these can differ. (Consider: median of [A] {2, 3, 4} = 3, median of [B] {0, 1, 0} = 0; median of [A] – median of [B] = 3, but median of [A-B] {2,2,4} = 2).

$r \sim .16$ ). To compare the relative decreases in consonant-le formation and gemination errors, I calculated effect sizes from the Wilcoxon signed-rank tests of the pre and post improvements in both types of errors, for both groups (all tests were non-significant). The effect size for the categorical group's improvement in consonant-le formation was .9 ( $W=19$ ,  $Md_{\text{gain}} = 4.5$  errors); the effect size for gemination was smaller ( $r = .6$ ;  $W=13$ ;  $Md_{\text{gain}} = 1.5$  errors). The particular group yielded a similar pattern. The effect size for the particular group's improvement in consonant-le formation was 1.0 ( $W=10$ ,  $Md_{\text{gain}} = 1.5$  errors); the effect size for gemination was smaller ( $r = .4$ ;  $W=4$ ;  $Md_{\text{gain}} = 0$  errors).

To summarize, the data provided little evidence that the vowel-sound colours affected performance. Both groups improved; for both groups, the improvement in consonant-le formation errors was more consistent and large than their improvement for gemination errors. Although the pre/post gains in overall accuracy and both types of errors appear larger in the categorical than particular group, such differences can be attributed to the spuriously better pre-test performance in the particular than categorical group (20% to 40%). By post-test, both groups performed similarly (about 40-60% accuracy), suggesting the existence of an improvement plateau. I discuss the nature of the plateau in the summary to this section.

## **Transfer**

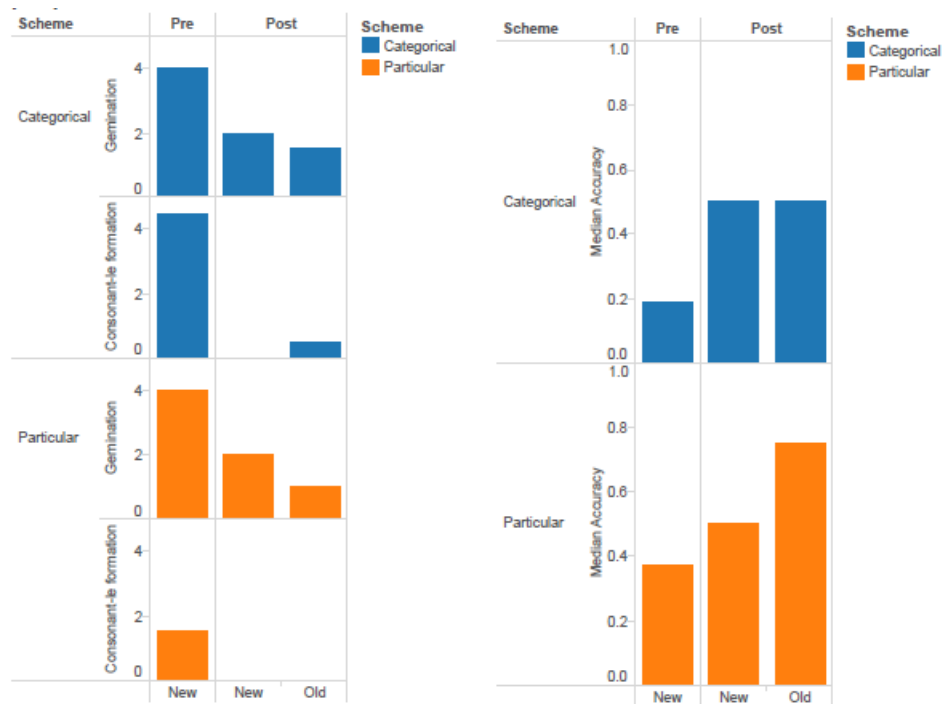
My second overarching research question concerned transfer. Although my overall assessment of children's pre and post accuracies yielded little support for a colour effect specific to the categorical scheme, two of my three proposed mechanisms of colour's effects (described in CH2 and CH3) predicted that colour effects would only present for certain types of words. If so, the overall comparison might mask such effects. In addition, both groups seemed to improve in consonant-le formation; both groups experienced colour-highlighted consonant-le units. CH1 could be therefore tested with respect to consonant-le formation errors.

Testing CH1-3 required separate comparisons of the pre and post accuracies for familiar versus unfamiliar and coloured versus uncoloured words. I describe these in turn.

## New versus Old Words

One question that pertained to CH1-3, and to the design of our system in general, was whether children's improvement was restricted to words they had encountered during the study. Improvement that was limited to words encountered in the study would suggest that children acquired no generalizable knowledge of spelling rules or conventions, but simply memorized some whole word-sound correspondences. Improvement (relative to pre-test) for post-test new words would suggest that children acquired some generalizable spelling knowledge, either of rules or of frequently occurring sub-word sound correspondences.

Neither group showed a significant advantage for old to new words (both  $ps > .4$ ), though the particular group tended towards a familiar word advantage ( $W_{\text{categorical}} = 5$ ,  $r = .23$ ,  $Md_{\text{new}} = 44\%$ ,  $Md_{\text{old}} = 48\%$ .  $W_{\text{particular}} = 6$ ,  $r = .60$ ,  $Md_{\text{new}} = 60\%$ ,  $Md_{\text{old}} = 69\%$ ). The post-test improvement for spelling new words approached but did not achieve significance (both  $ps \sim .06$ ,  $W_{\text{categorical}} = 14$ ,  $r = .66$ ,  $Md_{\text{pre\_new}} = 24\%$ ,  $Md_{\text{post\_new}} = 44\%$ .  $W_{\text{particular}} = 4$ ,  $r = .40$ ,  $Md_{\text{pre\_new}} = 44\%$ ,  $Md_{\text{post\_new}} = 60\%$ ), though the large effect sizes provide some grounds for concluding that children developed some transferrable spelling skills. Figure 4-3 (right) shows the pre and post accuracies for the two groups, broken down by word familiarity:



**Figure 4-3. (Right) Pre and Post Test Accuracies, split by word familiarity and scheme. (Left) pre and post consonant-le formation and gemination errors, split by word familiarity and scheme.**

Although non-significant, the modest improvement in children's post-test spellings of new words suggested that they learned and applied some transferrable strategies. To understand the nature of these strategies, I determined whether children's advantage for spelling new words at post-test characterized consonant-gemination or consonant-formation errors. If the strategies involved "dictionary rules", children should commit fewer gemination errors for new words at post than at pre-test. If the strategy involved memorizing frequently occurring sub-units, such as the six consonant-le syllables I exposed children to, children should commit fewer consonant-le formation errors for new words at post than at pre-test.

Two pairs of Wilcoxon signed-rank tests (evaluated at criterion .01 account for the four independent sources of type one error) comparing children's pre and post numbers of gemination ( $W_{\text{categorical}}=21$ ,  $p<.04$ ,  $r=1$ ,  $Md_{\text{pre}}=4$ ,  $Md_{\text{post}}=1.6$ ;  $W_{\text{particular}}=10$ ,  $p<.07$ ,  $r=1$ ,  $Md_{\text{pre}}=3.75$ ,  $Md_{\text{post}}=1.6$ ) and consonant-le formation errors ( $W_{\text{categorical}}=21$ ,  $p<.03$ ,  $r=1$ ,  $Md_{\text{pre}}=4.3$ ,  $Md_{\text{post}}=1$ ;  $W_{\text{particular}}=10$ ,  $p<.07$ ,  $r=1$ ,  $Md_{\text{pre}}=2$ ,  $Md_{\text{post}}=0$ ) approached but failed to achieve significance. The large effect sizes suggest that the

strategies that children transferred to unfamiliar words enabled them to commit fewer gemination and consonant-le formation errors, but the 40-50% median post-test accuracies cast doubt on their effectiveness.

The modest post-test old word advantage in the particular group suggested some effect of orthographic memory, that children encoded and applied memories of words encountered in instruction. The contents of children's orthographic memories could involve whole words or sub-units. The sub-units could be syllables (such as consonant-le) or smaller letter-sound units (the medial consonant). Whole word encoding predicts an old word advantage for consonant-le formation and gemination. Conversely, the findings that children have an easier time of segmenting words into syllables than letters (Fox & Routh, 1975), and of attending to initial or final than medial consonants (Battacharyna & Ehri, 2004), predicts an old word advantage for consonant-le formation errors only. To explore the contents of children's orthographic memories, I conducted and compared the effect sizes of two Wilcoxon signed rank tests on the particular group's children post-test old and new word consonant-le formation and gemination errors. Although neither test passed significance, ( $ps > .5$ ), the old word advantage for gemination errors yielded a moderate effect size ( $W_{\text{consonant-le\_formation}}=2$ ,  $r_{\text{consonant-le\_formation}}=.2$ ,  $Md_{\text{old}}=0$ ,  $Md_{\text{new}}=0$ ;  $W_{\text{gemination}}=6$ ,  $r_{\text{gemination}}=.6$ ,  $Md_{\text{old}}=.75$ ,  $Md_{\text{new}}=1.75$ ). The negligible effect for consonant-le formation errors seemingly attributes the particular group's generally strong consonant-le formation performance. The moderate advantage for old word gemination is consistent with the notion that children memorized the whole word, or that they focused especially on the medial consonants of words during instruction, perhaps because gemination caused them the most difficulty.

### **Coloured versus Uncoloured Words**

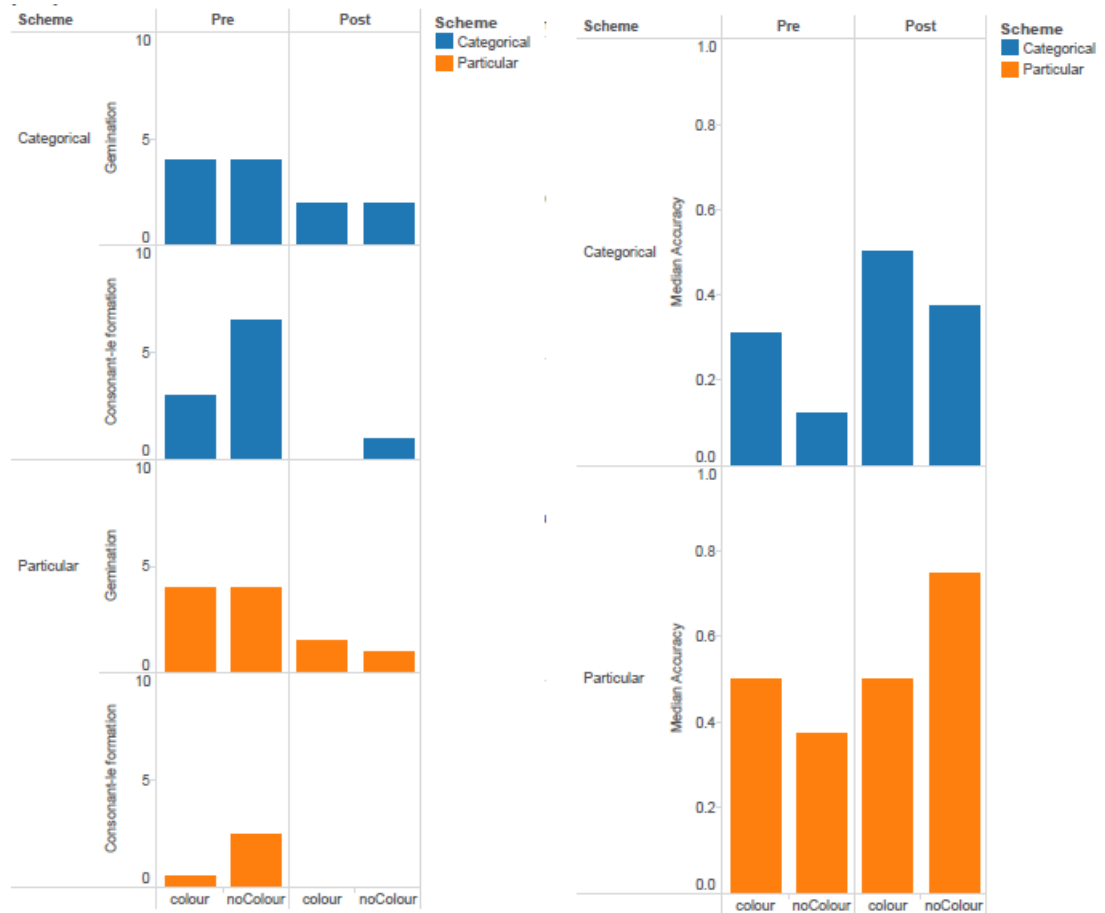
Fully addressing CH1-3, and the overarching question of how performance transferred, required an additional comparison of pre and post-test performance differences for coloured versus uncoloured words. In these comparisons I sought evidence for a weaker and transient impact of the categorical colour codes, which the overall comparison might have masked, or evidence of a transient attentional benefit of the consonant-le colouring, which was common to both schemes.

Transient benefits of the categorical codes should present as decreases in the number of consonant gemination errors at post-test for coloured but not uncoloured words, in the categorical but not particular group; transient benefits of the colour-highlighted consonant-le unit should present as decreases in the number of consonant-le formation errors, for both groups. Depending on the mechanism, such benefits might appear at post or pre-test.

As expected, although both groups tended towards somewhat better pre-test accuracies for coloured than uncoloured words, the advantage was not significant (both  $ps > .25$ ). The advantage was more consistent in the categorical group, as the greater signed rank sum ( $W$ ) statistic suggests ( $W_{\text{categorical}} = 10$ ,  $r_{\text{categorical}} = .51$ ,  $Med_{\text{pre\_colour}} = 31\%$ ,  $Med_{\text{pre\_nocolor}} = 12\%$ ;  $W_{\text{particular}} = 1$ ,  $r_{\text{particular}} = .1$ ,  $Med_{\text{pre\_colour}} = 50\%$ ,  $Med_{\text{pre\_nocolor}} = 37\%$ ). Post-test accuracies provided even less evidence of an overall colour-advantage (both  $ps > .8$ ;  $W_{\text{categorical}} = 2$ ,  $r_{\text{categorical}} = .09$ ,  $Med_{\text{post\_colour}} = 56\%$ ,  $Med_{\text{post\_nocolor}} = 37\%$ ;  $W_{\text{particular}} = 1$ ,  $r_{\text{particular}} = .1$ ,  $Med_{\text{post\_colour}} = 56\%$ ,  $Med_{\text{post\_nocolor}} = 68\%$ ). Figure 7 (right) shows the pre and post accuracies for the two groups, broken down by word colour.

Overall accuracies depend on consonant gemination and consonant-le formation errors. My framework predicted that the categorical group scheme would reduce gemination errors, but only at post-test, after children had learned the codes' meaning. Both schemes colour-highlighted the consonant-le unit, so both groups might show a colour advantage for consonant-le errors. If the mechanism involved low-level perceptual grouping or response reinforcement, an effect of colour-highlighting might appear at pre-test.

The lack of an effect of colour on overall accuracies suggested that colour failed to support at least one of gemination or consonant-le formation, but it could mask an advantage of colour to strictly one of these processes. To check for a colour-advantage specific to one type of error, I compared the pre and post-test quantities of gemination and consonant-le formation errors for coloured and uncoloured words, separately for the categorical and particular groups. Figure 4-4 (left) summarizes the results.



**Figure 4-4. (Right) Pre and Post Test Accuracies, split by word familiarity and scheme. (Left) pre and post consonant-le formation and gemination errors, split by word familiarity and scheme.**

As expected, neither group showed a colour-advantage for consonant-gemination at pre-test (both  $ps > .3$ ,  $W_{\text{categorical}}=1$ ,  $r_{\text{categorical}}=.04$ ,  $Md_{\text{pre\_color}}=4$ ,  $Md_{\text{pre\_nocolor}}=4$ ,  $W_{\text{particular}}=0$ ,  $r_{\text{particular}}=0$ ,  $Md_{\text{pre\_color}}=4$ ,  $Md_{\text{pre\_nocolor}}=4$ ) and the particular group showed no colour-advantage at post-test ( $p > .5$ ,  $W=1$ ,  $r=.1$ ,  $Md_{\text{post\_color}}=3$ ,  $Md_{\text{post\_nocolor}}=2$ ). Consistent with the absence of an effect of colour-coding scheme, the categorical group also showed no colour-advantage at post-test ( $p > .4$ ,  $W=6$ ,  $r=.28$ ,  $Md_{\text{post\_color}}=3.5$ ,  $Md_{\text{post\_nocolor}}=3$ ).

My framework predicted that both groups would show a colour-advantage for consonant-le formation, possibly at pre or post-test. To assess this, I compared the quantities of consonant-le formation errors between coloured and uncoloured words at

pre and post-test, but pooled the categorical and particular data. To compensate for the increase in type I error<sup>11</sup>, I evaluated the tests at criterion .002.

Although it did not pass criterion, at pre-test, children tended to commit fewer consonant-le formation errors for coloured words ( $p \sim .04$ ,  $W=24$ ,  $r=.75$ ,  $Md_{\text{colour}}=2.5$ ,  $Md_{\text{no\_colour}}=3.5$ ). The advantage was weaker at post-test ( $p \sim .12$ ,  $W=10$ ,  $r=.22$ ,  $Md_{\text{colour}}=0$ ,  $Md_{\text{no\_colour}}=.5$ ), likely because half of the children had perfected consonant-le formation by post-test.

A colour advantage for consonant-le formation that manifested at pre-test must attribute mechanisms that do not depend on learning. One candidate mechanism was that colour-highlighting the units triggered perceptual grouping mechanisms, which subsequently triggered children's memories of their lessons on the consonant-le syllable. This explanation supposes that the colour advantage for consonant-le formation errors should appear after the first few words, and be stable. Likewise, any "practice effects" (performance advantages for the second set of words that children spell) should be greater for children whose first set were coloured than uncoloured words.

Neither prediction was supported. Although children's post-test data yielded the expected mid-assessment stable colour-advantage (first-half median *uncoloured* advantage of 1.5 errors; second-half median 2 error coloured advantage), their pre-test data did not: the colour advantage was stronger for the first 7 words (median colour advantage of 3 consonant-le formation errors); for the final 7 words, the groups performed better with uncoloured words (median uncoloured advantage of 2 consonant-le formation errors). Similarly, a Wilcoxon between-groups comparison on the second-set accuracies of children who spelled the coloured versus uncoloured words first failed to find an advantage for spelling the coloured words first ( $p>.9$ ,  $r<.08$ ). Such results cast doubt on the true existence of a colour advantage, but there is an alternative explanation, which I discuss in the summary of this section.

---

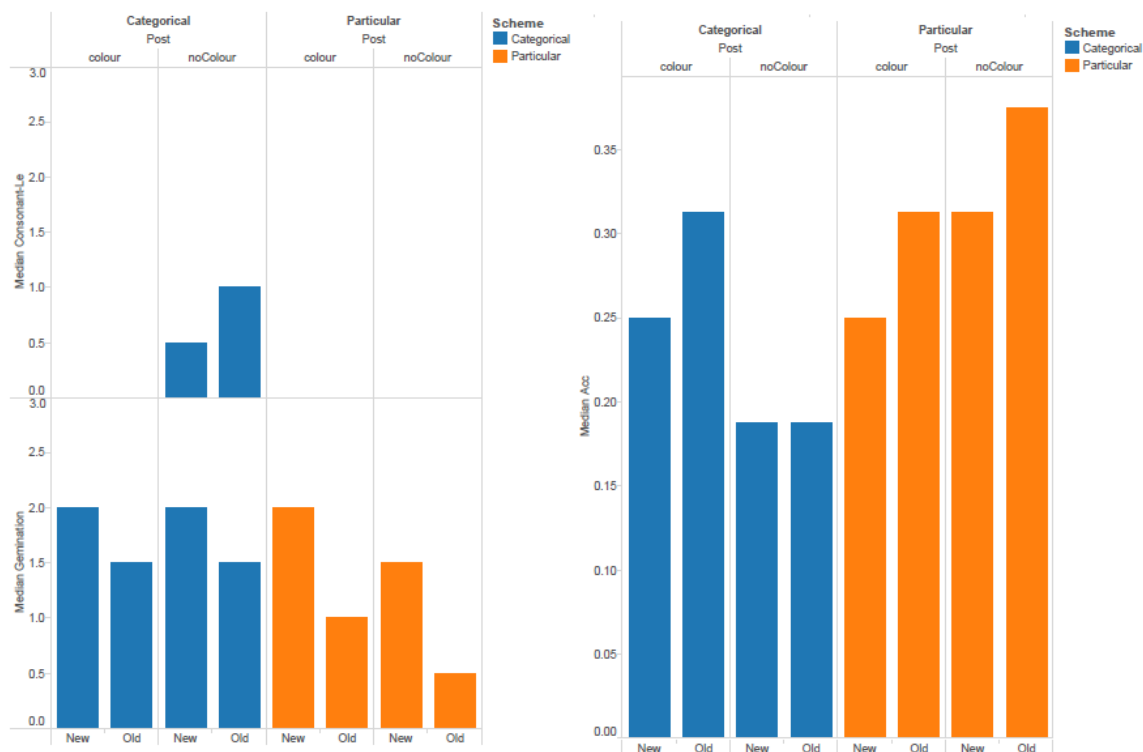
<sup>11</sup> Assuming 24 comparisons, the cross product of the three two level and one three level factor: colour-coding scheme, testing period, word appearance and metric- accuracy, consonant gemination error and consonant-formation error.

### Interactions between Word Appearance and Familiarity?

CH3 predicted that the advantage for coloured words would be greater for old than new words: in addition to whatever properties underlie the colour advantage for new words, colour might help trigger children's memories of previously seen words.

The only colour advantage I observed was for consonant-le formation errors. I did not observe the advantage at post-test, when children performed at ceiling for consonant-le formation. Still, the children exhibited some variability in their post-test colour advantages, and in their colour-advantages for consonant-gemination. To assess CH3, and to see whether colour advantages at post-test or for consonant gemination were masked by an interaction with word familiarity, I calculated a new metric, colour advantage (the difference between coloured and uncoloured scores), and compared the post-test colour advantages on each metric between old and new words. Consistent with the original comparisons, I assessed overall accuracy and consonant gemination within groups, and consonant-le formation across groups.

Neither group yielded a colour-advantage for either overall accuracy (both  $ps > .6$ ,  $W_{\text{categorical}}=8$ ,  $r=.17$ ,  $Md_{\text{colour\_adv\_old}}=.05$ ,  $Md_{\text{color\_adv\_new}}=.10$ ;  $W_{\text{particular}}=2$ ,  $r=.2$ ,  $Md_{\text{colour\_adv\_old}}=0$ ,  $Md_{\text{color\_adv\_new}}=0$ ) or consonant gemination that differed between old and new words (both  $ps > .4$ ,  $W_{\text{categorical}}=2$ ,  $r=.09$ ,  $Md_{\text{colour\_adv\_old}}=0$ ,  $Md_{\text{color\_adv\_new}}=.5$ ;  $W_{\text{particular}}=3$ ,  $r=.3$ ,  $Md_{\text{colour\_adv\_old}}=0$ ,  $Md_{\text{color\_adv\_new}}=0$ ). Likewise, there was no suggestion of a post-test colour advantage for consonant le formation that was restricted to old words ( $p \sim .25$ ,  $W=6$ ,  $r=.13$ ,  $Md_{\text{colour\_adv\_old}}=0$ ,  $Md_{\text{color\_adv\_new}}=0$ ). Although the categorical group yielded somewhat better performance for coloured old than coloured new or uncoloured old words, the advantage was inconsistent between children, as the small effect size suggests. The modest advantages for coloured and familiar words appeared independent and additive, not interactive. Figure 4-5 summarizes these results.



**Figure 4-5. (Right) post-test accuracies for familiar and unfamiliar words, split by word appearance and scheme (Left) consonant-le formation and gemination errors for familiar and unfamiliar words, split by word appearance and scheme.**

## Paper and Pencil

As with vowel discrimination, I assessed children's consonant-le spellings with paper and pencil, at pre and post-test. I did so for consistency with the vowel discrimination assessments and because children's ability to spell the same words on paper as with PhonoBlocks provided an additional measure of transfer. Because performance on the coloured PhonoBlocks letters could be impacted by colour scheme, or differ for gemination and consonant-le formation, I compared overall accuracies and quantities of gemination and consonant-le formation errors between paper and pencil and coloured PhonoBlocks at pre and post-test, for both groups. No test was significant (all  $ps > .8$ ), and the effects were negligible (all  $rs < .1$ ). As with vowel discrimination, there was no suggestion of an overall difference between children's consonant-le word spellings with PhonoBlocks versus paper and pencil.

### **Summary of Overall Effects:**

My aggregate data yielded little evidence for my framework. I designed the categorical scheme to target gemination errors and predicted greater improvement in gemination errors in the categorical than particular group. Neither group of children improved their gemination errors. Although most children improved by post-test, the improvement attributed a decrease in their numbers of consonant-formation errors.

The categorical group yielded a slightly larger overall pre/post gain than the particular group. This difference was attributable to the spuriously superior performance of the children in the particular group on consonant-le formation, and the restriction of improvement to consonant-le formation. By post-test, both groups were performing around chance- between 50 and 65% accuracy. 50% of the assessment words required a single, and 50% a double consonant. Children would perform at chance if they did not understand gemination (and geminated randomly, or consistently geminated or not), but understood consonant-le formation.

Both schemes colour-highlighted the consonant-le unit. My analysis of children's transfer performance suggested that the colour-highlighting might have played some role in reducing children's frequencies of consonant-le formation errors. Children committed slightly fewer consonant-le formation errors with coloured than uncoloured words. The effect was stronger at pre-test. At pre-test, I had not told children the meaning of the codes. The colour advantage must therefore attribute lower-level perceptual or general motivational mechanisms. One candidate mechanism, that colour-highlighting the unit helped children remember their lessons about consonant-le, is inconsistent with the observation that the colour advantage appeared for the very first word, and was stronger for the first than second half of assessment words. Another possibility is that children remembered certain whole words that included consonant-le, and that the unique colour change that occurred when they spelled them increased their confidence that their spelling was correct. This explanation presupposes that children had stronger whole-word orthographic memories of the first half of words than the second. Because I sampled all consonant-le words from a list that the tutors supplied, this explanation seems unlikely, although it is consistent with the observation that the period of advantage flipped between pre and post-test: if the pre-test first half colour advantage

attributed random word sampling error, then it would be unlikely to occur at post-test. At post-test, the first-half advantage had indeed disappeared.

Children committed fewer errors with old than new words. These data suggested that children a) remembered the pre-test words b) used the memories to spell the word. I surmised that colour might function partly by cueing children's memories of previously seen words. I consequently predicted that the difference between coloured and uncoloured words might be greater for old than for new words. Although the categorical group yielded somewhat greater median colour advantages for old than new words, the effect was inconsistent and weak.

Finally, children performed better with new words at post than at pre-test. The pre/post new word advantage seemed greater for consonant-le than gemination errors. Because children performed poorly on germination in general, these data do not imply that consonant-le formation is easier to generalize than gemination, though they suggest that children's improvement with consonant-le formation is generalizable.

My sample was small. The tutors and I observed that each child varied considerably in their attentional and linguistic challenges, in their behaviour during the sessions, and how they used the codes. Although my group analyses confirmed that my implementation of the categorical colour codes had no overall impact on children's spelling performance, I had observed some cases of children appearing to use the colours. To search for exceptions to these aggregate trends and thereby frame my qualitative, individual-cases analyses, I repeated the analysis in terms of individual children.

### ***Performance and Transfer, by Individuals***

#### **Pre and Post Performance:**

Four out of six children in the categorical group improved at post-test; two (P1 and P7) performed worse at post-test. The range of post-test accuracies was 7 to 87.5% (1 to 14 out of 16 words correct). One child (P1) who performed worse at post-test committed more consonant-le formation and gemination errors. The other (P7) committed fewer consonant-le formation errors, but more gemination errors. Two

additional children from the categorical group (P9 and P8) and one from the particular (P4) committed more gemination errors at post-test than pre-test. One child in the particular group performed worse at post-test, spelling only 32% of words correctly. The rest improved. The range of post-test accuracies was 32 to 100% (5 to 16 out of 16 words correct).

Five children (P2, P5, P10, P3 and P6) were exceptions to the general lack of a treatment effect for gemination errors. All children committed fewer gemination errors at post than at pre-test. P10 yielded the greatest improvement in gemination, achieving 0 errors at post-test but committing 9 at pre-test. P2 and P5 improved by 5 errors. P2 committed 7 and 2 gemination errors at pre and post-test, respectively, and P5 committed 9 and 4. P3 and P6 improved by less (3 and 1 error, respectively), down to 5 from 8 and 6. P5 and P2 had the categorical scheme; P10, P3 and P6 had the particular scheme.

## **Transfer**

### **Word Familiarity**

Four out of six children with the categorical scheme performed better with old than new words. P5 performed worse with old than new words. P7 performed equally with new and old words. Of the particular scheme, P10 achieved perfect performance for new and old words. The remaining children performed better with old than new words. All children sustained the trend of committing fewer consonant-le and gemination errors with old than new words. All children committed fewer gemination and consonant-le formation errors with new words at post-test than pre-test. All children in the particular group, and four children in the categorical group, sustained the trend for improvements in consonant-le formation errors on new words to substantially outstrip improvements on gemination errors. Two children in the categorical group (P5 and P2) improved substantially and almost equally in consonant-le formation and gemination (P5 committed 9 gemination errors at pre-test and 2 at post-test. P2 committed 7 at pre-test and 1 at post-test. P5 committed 9 consonant-le formation errors at pre-test and 0 at post-test. P2 committed 1 consonant-le formation error at pre-test and 0 at post-test). One child in the particular group (P10) also deviated from the aggregate trend. She

committed 9 and 8 consonant-le formation and gemination errors on pre-test new words; at post-test, she committed no errors. These children appeared to learn some transferrable strategy for choosing when to geminate as well as forming the consonant-le unit.

### **Assessment Word Colour**

Children varied in the magnitude of the effect of assessment word colour. Of the categorical group, two children performed worse (albeit by only two words) with coloured than uncoloured words. Two performed better with coloured words, but the differences were small (2 and 1 word). The remaining two showed a somewhat larger colour effect. P5 spelled 6 words correctly with coloured letters; 3 with uncoloured letters. P2 spelled 9 words correctly with coloured letters, 6 with uncoloured letters. One child in the particular scheme (P4) yielded colour effects as P2 and P5, with a 3 word advantage for coloured words. Two children yielded no difference. The last yielded a 2 word advantage for coloured words.

By my framework, a colour advantage in the particular group should only involve consonant-le formation. By contrast, a colour advantage in the categorical group should involve gemination and consonant-le formation. Because the colour-sound mappings for short and long vowels had to be learned, gemination colour advantages should only manifest at post-test. Although I failed to observe these effects in the aggregate, I thought they might characterize the responses of the participants- P2, P5 and P4- who showed the strongest colour advantages. I focus my analysis of how the colour advantage affected the two types of errors and interacted with assessment period around P2, P5 and P4, but compare their response patterns to those of their cohorts.

P5' behaviour was consistent with my predictions. His colour advantage for gemination only manifested at post-test. At pre-test, he committed 5 gemination errors with coloured and 4 with uncoloured words. At post-test, he committed 0 gemination errors with coloured and 2 with uncoloured words.

P2's behaviour was inconsistent with my predictions. P2's colour advantage for gemination was not only smaller than P5's (a one error difference), it was equal at pre and post-test.

P4's behaviour was also inconsistent with my predictions. P4's behaviour was similar to P5'. She showed no pre-test gemination advantage for coloured words, but showed a 2 error colour advantage at post-test.

P4 had the particular scheme. Her colours did not distinguish long and short vowel, so her colour advantage must attribute other factors. These could be general motivational or attentional effects of colour, or a more specific benefit of colour-highlighting the consonant-le syllable. Colour-highlighting the consonant-le syllable generated a colour difference between geminated and non-geminated words. It is plausible that this- without the different colours for short and long vowel- benefitted gemination as well as consonant-le formation.

Such benefits must be small. No other child in the particular group showed a colour advantage for gemination. P3 performed worse with coloured than uncoloured words at post-test; similarly at pre-test; P6 performed similarly at pre and post-test. Both children showed a strong colour advantage at pre and post-test for consonant-le formation. The remaining children of the categorical group provided little evidence for any effect of the short and long vowel colours. Two children committed equal gemination errors for coloured and uncoloured words, at pre and post-test. One committed more post-test gemination errors for coloured than uncolored words. The last (P8) committed one fewer post-test gemination error for coloured than uncoloured words. The children showed more consistent post-test consonant-le formation colour advantages. Two children performed better with coloured than uncoloured words, though the advantage was small (one error). One child (P8) committed no consonant-le formation errors, for either coloured or uncoloured words. The last child (P7) committed three consonant-le errors for coloured and uncoloured words.

I uncovered no striking individual deviations to the non-interactivity of word colour and familiarity effects.

## **Paper and Pencil**

In general, children who improved in consonant-le formation with coloured or uncoloured tangible letters transferred those gains to paper. Of the children who improved substantially in consonant gemination (P5, P2 and P10), only P2 and P10's improvement transferred to paper and pencil. Both children committed 0 gemination errors at post-test, down from 3 and 4 at pre-test. P10 committed one consonant-le formation error at post-test, but was still down from 4 at pre-test. P2 committed 0 consonant-le formation errors at pre and post-test. P5' improvement did not transfer to paper and pencil. With paper and pencil, he committed 3 gemination errors at pre-test; 4 at post-test.

## **Summary**

My individual-cases analyses sought exceptions to the aggregate behaviour that I described in section 4.1.2. In particular, I sought evidence of individual cases of my framework's predictions: children in the categorical scheme who improved in gemination as well as consonant-le formation, and whose improvement seemed attributable to the categorical codes. Although I uncovered two children (P2 and P5) who improved dramatically in gemination, only P5's transference pattern suggested an effect of colour. In addition, I observed this pattern in another child, P4, who had the particular scheme, and there was another child (P10) who had the particular scheme but improved more than either P2 or P5.

To summarize: my pre and post assessment analysis yielded no firm evidence of a colour effect, at either an aggregate or an individual level. My pre and post assessment data do not communicate how children in either scheme improved. Various strategies were available. Children might use their tutors' feedback. Children might learn to exploit the system or predictable word types. Children might use the colour codes, but do so unsuccessfully, unconsciously, or without understanding the underlying rules. The presence of a colour assessment word advantage suggests that colour did not help children understand the underlying rules. If it had, performance should transfer to uncoloured words.

I collected various additional metrics to help me explore these issues. I aimed to understand: why the colour codes did not work in the aggregate, what other supports enabled children to perform better at post-test, whether P5 and P2 yielded any evidence of using the colour codes, how they used them, whether other children attempted to use the codes, whether and why they were successful or not, what individual factors contributed to children's use or misuse of the colour codes and, most importantly, how might designers refine the codes or their implementation context to better support children in using them to acquire and transfer linguistic rules?

The first additional metrics were the software event logs. The remainder were my qualitative metrics.

#### **4.1.2. Software Event Logs**

My analysis of the pre and post spelling accuracies yielded several unexpected results. These were: the aggregate "plateaus" at roughly 50% accuracy, the aggregate absence of a colour effect and the possible existence of two children who could use the colours, albeit in different ways. I analyzed my event logs to explore the questions I posed in section 3.3.5.7. I expected my analysis of the first and fourth question to supplement my explanations of the aggregate plateaus and the absence of an aggregate matched-colour benefit, and my analysis of the second and third questions to clarify how P2 and P5 might have used colour, i.e., did P2 and P5' event logs yield cases of correcting erroneous first submissions?

##### ***Correct Second Submissions***

I hypothesized that the categorical colour codes might help children benefit from feedback, i.e., I predicted that more children in the matched than mismatched scheme would (following an incorrect first submission) subsequently submit the correct word. In other words, I predicted an effect of colour-coding scheme on the number of submissions required to submit the correct word.

Each session, children spelled two consonant-le words. Because children were expected to respond in fewer submissions on their second word, I used the maximum

number of submissions (between the two words, per session) as the dependent metric. A preliminary Shapiro-Wilkes test confirmed that the metric was not normally distributed ( $p < .001$ ), but positively skewed (skewness=1.2). The interquartile range was one to two submissions (on average), but four children required between three and five submissions (on average). In respect of the non-normal data, I used non-parametric comparisons.

A Wilcoxon between-groups comparison failed to find an effect of colour-coding scheme on the number of submissions required before responding correctly ( $p > .3$ ,  $Z = 1.1$ ,  $r = .34$ ). In both groups, the median number of submissions required was two, and the interquartile range was one to two submissions. The categorical group presented a somewhat greater total range, with more outlying children requiring five, four and three submissions.

Collapsing across the sessions, then, most children typically required two submissions before responding correctly. Still, it seemed possible that scheme might affect children's *trajectories*, for example, whether children converged on a minimum two submissions (and stayed there), on what session this occurred, or whether children's performance fluctuated over the sessions. My framework predicted that children in the categorical group would stabilize at two or one submissions, i.e., there should be an inverse relation between session and number of submissions, whereas children in the particular group should present less of a relationship. To assess this, I compared the Spearman rank coefficients ( $\rho$ ) relating session to number of submissions between the categorical and particular groups.

The overall relationships between session and number of submissions were non-significant and small ( $\rho_{\text{categorical}} = -.12$ ,  $\rho_{\text{particular}} = .13$ ), suggesting a greater degree of individual variability in the across-session responses than the aggregate data suggested. To better understand the nature of children's particular response tendencies, I analyzed the relationships between session and number of submissions at the level of individual children. I assessed each child's modal number of maximum responses required, the stability of the mode, and the session at which the child stabilized.

With three exceptions (P2, P5 and P10), who most frequently submitted the correct word the first time (spelling a minimum of 6 out of 12 words correctly the first time, at least 4 of which were consecutive), children in both groups most frequently (at least 8/12 sessions) submitted the correct word upon their second try.

P2, P5 and P10 achieved their first correct first submission (after which the majority of first responses were correct) by the second, eighth and third sessions, respectively, after which time they responded correctly the first time for the majority of sessions (i.e., they stabilized). P2 and P5 had the categorical scheme; P10 had the particular scheme. Of the remaining children, two achieved a correct second submission by session three and one by each of the fifth, fourth and third sessions. Thereafter, three children had an occasion of submitting two incorrect words, and four had occasions of submitting a correct initial word. P7 required a modal 3 submissions; she achieved this by session 6, and vacillated between two and three submissions for the remaining sessions.

The seven children who did not achieve modal first correct responses also failed to consistently achieve a first correct response for any period spanning more than one session, i.e., the times they responded correctly the first time were flukes. Failure to provide a correct first response suggests that children did not understand all spelling concepts involved in the words, though their improvement (from more than two to two requisite submissions) suggests they learned at least one concept. To understand which, I analyzed the trajectories of consonant-le formation and germination errors.

Two Shapiro-Wilkes tests confirmed that the distributions of consonant-le formation and gemination errors were non normal (both  $ps < .001$ ). To better understand the nature of children's improvement, and why so many stabilized at two submissions, I conducted two Spearman rank correlations between session and the quantities of consonant-le formation and gemination errors. Because children in the categorical and particular groups appeared similar, I computed correlations on their pooled data.

Neither relationship was significant (both  $ps > .1$ ), or substantial (both  $\beta s < .1$ ), suggesting again that the overall analyses were complicated by individual variability. To compensate, I again analyzed the data at the level of individual children.

Consonant gemination errors were more frequent than consonant formation errors, and consonant gemination errors were more resistant to treatment than consonant formation errors. Following session five, the majority of children (9/10) required a maximum of 2 submissions to spell a word correctly. By the fifth session, 9 out of 10 children committed more consonant gemination than consonant-le formation errors. Of the children who stabilized at requiring two submissions, 6 out of 7 would commit no more than three consonant-le errors for the remainder of the study.

In general, then, the children's first submissions erred in consonant-gemination, but they were also generally capable of fixing their error following a first incorrect submission. Because consonant gemination has only two options, this means that most children did not learn how to determine- from the sound of the vowel- how many consonants a word required. Rather, it is probable that children learned that a) when they erred, the error concerned the medial consonants and b) if one consonant was wrong, two was correct (and vice versa). This strategy would enable children to respond correctly the second time, but it predicts an aggregate ceiling accuracy of 50% (half of the consonant-le words required one and half two consonants) on the post-test assessment, where children had only one chance to spell the words.

### ***Correcting Erroneous First Submissions***

I hypothesized that children with the matched colour scheme might be capable of "reading" the sounds of the on-screen words from the colours, such that they could determine if a candidate word (e.g., "stuble" for "stubble") was correct. I consequently predicted more instances of children correcting a word *before* submitting it in the categorical than the particular scheme, i.e., those children would use the *colours* as feedback in place of the feedback their tutor and I provided following an incorrect first submission.

I observed only three instances of this behaviour, all by two children (P2 and P5) in the categorical scheme. I detail these instances in section 4.2.4. P2's instance occurred early in the study (session 2). After about session 2, P2 typically responded correctly the first time (P2 committed two mistakes following session two). P5's

instances occurred later (sessions 6 and 8). Although P5 erred on session 7 (on consonant gemination), he answered correctly the first time from session 8 onwards.

### ***Unproductive Errors***

I observed no unproductive errors.

### ***Misled by Colour***

I observed one case of a child (P9) being misled by the colours. The child had the categorical scheme. The case matched my hypothesis, which was that a temporary colour change (i.e., a long red vowel becomes yellow as children build a long-vowel consonant-le word) might deter children from correctly completing the word. I detail this instance in section 4.2.4. Children otherwise ignored or used the colours successfully.

## **4.2. Qualitative Analyses**

In this section I describe my analysis of my qualitative metrics. My objectives were understanding: whether and why some children used the colour codes, how they used them, and what design features might support children who failed to use them or did so unsuccessfully. I divide this section into five main subsections. Excepting the first, which describes a hardware failure that I suspect impacted children's use of the colour codes, each sub-section involves a different metric or level of analysis. I analyze each metric or level with respect to my objectives. The next two sub-sections involve metrics or descriptions at the aggregate level. I summarize general trends from the tutor interviews and my recordings and observations. The fourth section describes my individual case studies. I summarize each child's profile and behaviour during the experimental sessions and assessments, with reference to the tutors' remarks and my observations. I relate these observations to the child's pre and post-performance and software event logs. In the final section I briefly summarize my observations and design lessons that I draw.

#### **4.2.1. Hardware Failures**

On the second day of the study, the circuitry that communicated the letter identity to the software malfunctioned. As a result, I had to remove it. The circuitry that connected the software to the LED strips was not compromised. I simulated the letter circuit by communicating children's letters to PhonoBlocks using a wireless keyboard.

The software failure could have affected children's use of the codes. One reason I thought colour would help was by causing an instantaneous colour change, coupled to the child's changing of the letters. Although I became quite adept at entering the letters as the child placed them, I occasionally erred and caused noticeable delays between children's placement of the physical letters and their appearance on the screen and change in colour. By the time the letters changed, children were often focusing on a different part of the word, or some other irrelevant object or thought. Although I did not explicitly record the occurrence of these errors, I estimate that they occurred on about 20% of the sessions, distributed randomly across the children, but concentrated towards the earlier sessions. By about session four I had largely eliminated such errors.

In addition, several of the letters or platform pins malfunctioned. This meant that certain letters appeared in the wrong colour, or certain slots could not support certain colours. Fixing such problems required me to hold the letters in place, such as to sustain the connections between the pins and letter-base, which partially obstructed children's views of the letters, or may have distracted them from the letters. Failure to hold the letters would cause, for example, a short vowel to appear red, not yellow, or a member of the consonant-le unit to appear blue, not magenta.

These errors occurred more frequently, throughout the course of the study. I estimate that there was at least one LED colour error, involving a temporary change in the letter colour, or my having to adjust the letter because the child did not produce adequate pin contact, on 90% of the sessions, distributed randomly across the children.

#### 4.2.2. Tutor Interviews

Tutors underwent one interview at the end of the first or second week. At the end of the study I asked tutors if their opinions had changed. No tutor reported a change in opinion, so I did not administer the questionnaire a second time.

I analyzed my tutor interviews after completing the quantitative assessments, with an eye to explaining unexpected results in the data. In particular, I wished to supplement the quantitative assessment of RQ1, and attempt to understand a) *how* colour codes might help (in a different setting, context or implementation) and b) *why* my implementation of the colour codes did not largely help. I also planned to use the tutor's insights to explain or otherwise supplement my remaining qualitative analyses.

##### ***T1***

T1 did not believe that the colour codes helped. She observed no differences in her children's spelling behaviour using PhonoBlocks versus paper and pencil, except that they seemed "a little distracted" when using PhonoBlocks. She did not think that her children used the colours as feedback; she believed that the colours confused them. T1 pointed out that the colour changes were more salient on the screen, but that children paid little attention to the screen:

"I don't think they get the colour codes. I think- they're not always watching the screen. They look confused about what the colour is. And I even tell them- "red and green'- they're like a deer in headlights."

Because T1 did not believe that her children understood the colour codes, she thought that PhonoBlocks would be equally effective without them. She ventured that the colour codes might be more effective "at the start of the year" or with younger children.

T1 pointed out that the older children had already learned specific terms, mnemonics and visual codes for representing the consonant-le syllable, syllable division, and long and short vowel. The colour codes imposed a new way of representing these concepts. T1 related children's difficulty learning and applying the colour codes to a difficulty in integrating the colour codes with their established representational system:

“...if you use a different term, they don’t know what you mean... and that’s an awakener. We need to teach it to them in different ways.”

## **T2**

T2 believed that the colours and tangibles helped. She believed that her children were “more focused” when using PhonoBlocks than paper and pencil, and that her children enjoyed spelling with PhonoBlocks better than with paper and pencil. T2 believed that PhonoBlocks would be less effective without the colour codes.

I asked T2 to explain why she thought the codes would help. T2’s explanation referenced a description of how the categorical codes would apply to forming vowel-consonant-e words, (e.g., FADE versus FAD), which I did not actually test, but which my team and I had used during our initial demos of the system to the tutors:

“...but when I saw the vowel change color with what you had done on there... I instantly saw the rule in action visually. Words were not needed to explain the rule. I saw the change and I knew.”

T2 believed that the “reflection period” with the colour-coded letters contributed to children’s greater focus, and pointed out that “that’s something we don’t do”.

## **T3**

T3 believed that the colour codes helped two (P10 and P2, but not P1) of her three children. She observed that her children “understood the rule better” since using PhonoBlocks, and thanked me for “teaching them a difficult concept”.

Like T2, T3 believed that the colour changes helped children visualize concepts that were “difficult” to grasp and express in words. In T3 case, these concepts were consonant doubling and the formation of the consonant-le syllable, which she thought were supported by the “different colours for the long and short vowels” and the “consonant-le syllable”. T3 thought PhonoBlocks would be less effective without the colour codes.

### 4.2.3. Observations

#### *Children's Memory of the Colour Codes*

T1 believed that children did not understand the meaning of the colour codes. She attributed their inability to use the colour codes to their lack of understanding of their meaning. My team and I assumed that children using our system would require only immediate instruction in the meaning of the codes to effectively use them. In light of the general absence of a colour effect in the quantitative data, I considered it important to assess children's comprehension of the colours more directly, and relate this to other indices of their use of the codes.

About one third through the study, I asked all of the children whether they remembered what the colour codes meant. My question had two parts. First, could children recite what each colour represented? Second, could children explain why a colour changed? (Or, for children in the particular scheme, why a letter flashed).

4 out of 5 children in the categorical scheme remembered that red meant long vowel and yellow meant short vowel. By contrast, only one child (P2) could explain why the vowels changed colour (when, for example, they added an "e" to a consonant and "i"), or why the colours differed between words that differed by the number of consonants. Likewise, while all children in the particular scheme recognized that the vowel letter flashing meant that its sound had changed, only one (P10) could explain why it would change following the completion of consonant-le (in words with one consonant), or why the number of consonants affected the sound.

Explaining why the vowel changed colour required children to explain: a) that consonant le became a syllable, and that the remaining letters became a syllable and that b) the type of the remaining syllable and what sound vowels have in that type. P2 from the categorical group and P10 from the particular group provided satisfactory explanations:

P2: "because this [P2 places his hand on top of the letters 'ble'] is a syllable, and this [P2 places his hand above the letters "bi"] is a syllable" [I ask P2 the type of the syllable "bi". He replies that it is open].

P10: [places hand between the syllables “bu” and “gle”] “bu.. gle” [P10 enunciates the syllables. P10 was shy, so I prompted her to explain her response: “what type of syllable is ‘bu’?” “open”. [“What sound does the vowel have in an open syllable?”] “long”.

Three of the remaining children in the categorical scheme provided no response or said that they did not know. One child (P1) supplied:

“Because it's short”

He provided no response to my follow-up question: “But why would it sound short here, when there are two ts, but sound long when there is one t?”. None of the remaining children in the particular scheme supplied a response.

### ***Children’s Comprehension of the Rules***

Successful use of the colour codes required children to understand the relationship between vowel sound category and colour, and consonant gemination, colour and vowel sound category. Consonant gemination is an action. It involves a different modality (motor response) than colour change, which is visual.

As I discussed in section 4.2.2.1, by session five most children spelled the word correctly on their second submission. Because the children’s errors involved consonant gemination, their first and second submissions differed by the presence/absence of a single consonant. Although these children could not explain why the vowels of their first and second submissions had different colours (4.2.3.1) I thought it possible that they might be capable of explaining why they had doubled or removed a consonant, i.e., why their *new* submission was correct.

Probing children’s ability to articulate the reason for gemination (or not) also assessed whether children had attained a deeper understanding of the relationships between the abstract concepts, syllable division, types and long and short vowels, that the rule involved. Goldin-Meadow et al found that children’s ability to benefit from an alternate representation of mathematical “balance” problems depended on their ability to articulate the rules (Goldin-Meadow, Alibali, & Church, 1993). My colours were an alternate representation of spelling concepts. I had designed the colour codes to help

children understand how the gemination rule follows from syllable division and types, which should support transfer.

To better understand how transfer, children's use of the colour codes, and their capacity to articulate the gemination rule interrelated, I asked children to explain why they changed their (incorrect) first word by doubling or removing a consonant. A correct response required children to explain how the geminated consonant affected the types of the syllables into which the word split, and how these affected the vowel sound. I asked this question twice, once around session 5 and again around session 8.

At session 5, Only P2 could explain the reason for the difference in the number of consonants. P2 responded correctly the first time for both of his words. Consequently, I asked P2 to explain why his first word ("noble") had one "b", while his second word ("settle"), had two "ts":

EC: "...so for the first word, you only had one 'b'. But for this word, you have two 'ts'. How come you have two "ts" in this word?"

P2: "Because, the vowels..."

EC: "The vowels?"

P2: "The vowel sounds... short?"

EC: "Yes. So what is it about the extra t that makes the vowel short?"

P2: "...because this is a syllable and this is a syllable?"

Of the remaining children, three did not respond. Two said that they forgot ("Don't remember", ~P1; "I forget. Don't worry, I have a two minute memory", ~P9). One child (P4, from the particular scheme) was shy and refused to respond. P4's refusal to respond brings up an important limitation in the use of explicit probes to assess children's understanding. Children might understand a rule without having the vocabulary or willingness to articulate it. This is why I cross-checked children's explanations against their logged data, i.e., the number of submissions they required to submit the correct word, and which type of error they committed. Accordingly, P4's

failure to provide correct first responses on subsequent sessions, and her tendency towards gemination errors, suggests that she did not comprehend the rule.

Four children provided disorganized explanations. They produced terms (long and short vowel, open closed or syllable), phrases (“big brother”) or gestures (segmenting the word into syllable) that their tutors supplied, or correctly pronounced the misspelled words, but could not identify how the concepts or pronunciations related. The children could not answer my follow-up questions, which aimed to clarify whether they understood what the terms or gestures meant:

EC: “So how come you added the extra *p*?”

P5: “Big brother”

EC: “What does that mean?”

P5: “Don’t know”.

-----

EC: “So how come we have two *ps* in this word? [*supple*]”

P3: “Because it would say *soopull*”

EC: “Why would it sound like *soopull* if we only had one *p*?”

P3: “Because there wasn’t an extra *p* to like, um... we’re on the right track... so we could make, uh, the word.

-----

EC: “why do we have two *ds* for this one? [*fiddle*]”

P6: “uh... long vowel.”

EC: “Long vowel? Or, uh- what sound does the vowel have in *fiddle*?”

P6: “eye”.

EC: “In *fiddle*? Is that long or short?”

P6: "Ah, short."

EC: "Okay, so you said that- long and short vowel is the reason we have two ds. So do you remember what that has to do with syllable division?"

P6: "Not really".

Two children revealed that they had learned that a) (by that point of the study) when they erred, the error typically concerned the number of consonants b) the words required either one or two consonants:

EC: "How come you removed the second 'd'?"

P10: "Because I did it that way the last time."

-----

EC: "Why did you remove the second p?"

P8: "I just thought it would be a good thing to do."

EC: "Why?"

P8: "If it doesn't have two it has one."

I queried children's comprehension of the rules a second time on the 8<sup>th</sup> session. With the exception of P10, who became capable of explaining the rule, no child changed their response category. P8 had developed a seemingly minimal-effort strategy of a) always submitting a word with one consonant first, but b) readying the second tangible consonant, should the first submission be incorrect.

### ***Children's Use of the Colour Codes***

For the first five sessions, most children (9/10, excepting P2, who rarely committed and could correct his errors past session 2) were incapable of fixing errors in their initial submissions without feedback. For the remaining sessions, children's errors involved gemination and they repaired them without feedback, though given they could

not explain the reason for their changes (4.2.3.2), it is likely that they learned that words required either one or two consonants.

A key rationale and design goal for the colour-coding schemes was the exploitation of colour changes as an inherently salient attentional cue. A key observational question was whether the colours attracted children's attention. My observations yielded little support for this idea. In general, unless I explicitly drew children's attention to them, children did not look at the colours as they formed the letters, and they did not look to the colours as a source of feedback or hints when their submissions were incorrect.

I observed only four cases of children explicitly using the vowel sound colour codes to spell the words. These cases were distinguished by their spontaneity: they occurred without my prompting. Two of these children were successful (P5 and P2), in the sense of using the colours to spell the words correctly during the session. One child (P7) was unsuccessful. Another (P9) seemed misled by colour into submitting an incorrect spelling. In addition, the successful children developed different skills: P5 could not transfer his superior performance with coloured to uncoloured words; P2 could. P5 could not explain why or why not he doubled the consonant; P2 could. Contrasting these four cases can yield insights about how and when colour codes work. I detail these cases in section 4.2.4.

Despite the across-condition differences in the post-test frequencies of consonant-le formation errors for coloured and uncoloured words, I observed no clear evidence of children explicitly using the common consonant-le colours.

#### **4.2.4. Individual Cases: *why might have colour worked or not?***

In this section I provide a more detailed analysis of each child. If they appeared to use the colour codes, I describe how they appeared to use them. I integrate my observations of their pattern of post-assessment transfer and responses to my questions about the rules and colour code meaning. If children did not use the colour codes, I attempt to explain why with reference to my and tutors' observations of the child. I begin each case study with a summary of the child's profile.

My individual analyses supplemented my aggregate quantitative assessment of my research questions. My quantitative analyses indicated the absence of an aggregate colour effect. My system event logs suggested that children learned to exploit the predictable pattern of consonant-le words and thereby achieved ~ 50% accuracy and that children did not attend to the colours unless I drew their attention to them. T1 suggested that children's difficulty integrating multiple representations prevented them from using the colour codes.

Despite this, the children exhibited considerable variability and some seemed to use the codes. My individual cases analyses helped me to understand the specific individual factors that operated- singly or in concert with the environmental factors, such as hardware failure, integrative difficulties or the availability of familiar tutor feedback, to prevent or support children in using the colours.

### ***A Note about the Profiles***

Although I intended to ask tutors about children's diagnostic and academic histories, the director of the school informed me that this information was confidential. Despite this, two tutors provided me some informal diagnostic and academic historical information. The other two tutors did not. I was allowed to ask children informally about their favorite school subjects and extra-curricular activities. My profiles for (P1, P2, P8, P3, P10) are limited to these data and what I observed of their personality. Table 4-1 summarizes the children's profiles:

**Table 4-1. Summary of Child Profiles**

ID	Grade	Attitude Towards and Engagement with PhonoBlocks	Favorite School Subjects and Activities
P1	4	Compliant /neutral	PE/art
P2	5	Compliant/neutral	Math/Science; building
P3	7	Compliant/mildly interested	Art/Socio-Emotional Learning/Cooking
P4	6	Compliant/Shy/Interested	Art/Fantasy
P5	3	Interested/physically engaged	Science/Art/Building
P6	5	Interested/Easily distracted	"All of them"/cars
P7	7	Resentful/Disengaged	Art/Music

P8	5	Resentful/Exploited system	PE/Sports
P9	7	Dismissive/Sociable	Art/Dancing
P10	5	Compliant/Shy	Art

***Children Who Spontaneously Attempted to Use the Colour Codes: P2, P5, P9 and P7***

Four children (P2, P5, P9 and P7) attempted to use the colour codes. Apart from the colours, neither P2 nor P5 required additional support to spell the words, but only P2 transferred performance to uncoloured words. P7 corrected an error in her word, but required additional support to complete the word. P9 was misled by the colours.

**P2**

P2 was bright. His favourite school subjects were mathematics and science. In his spare time, P2 enjoyed constructive activities such as Lego and KNex. P2 was quiet and compliant. He appeared neutral about using PhonoBlocks, expressing no obvious interest or resentment.

P2 performed better than average at pre-test, spelling 8 out of 16 consonant-le words correctly. He committed seven gemination errors and one consonant-le formation error. P2 achieved the second best overall post-assessment score, committing only two gemination errors; both with the uncoloured words.

By the second session, P2 had become capable of generally spelling the word correctly the first time- i.e., without explicit feedback from his tutor or me (two exceptions were sessions 4 and 8, on which P2 required two submissions).

As mentioned (4.2.3.3) P2 explicitly used the categorical colour codes. I observed this during P2's third session. P2 had to spell the word "wiggle". P2 initially created the word, "wigle". He did not submit the word. Suddenly, he removed the letters "g", "l" and "e", and shifted them one space to the right. He inserted an extra "g". I asked P2 why he changed the word.

P2: "I remembered it was supposed to be yellow when it sounds short."

Accordingly, P2 was capable of reciting the meaning of the colours (red is long, yellow is short vowel). P2 was also unique in the categorical group because he could explain why the colours changed and why he doubled or did not double a consonant in terms of syllable division and types (4.2.3.2). Although P2 showed a minor benefit for coloured words at post-test (both of his post test errors, both of which were gemination, occurred on uncoloured words), he committed no errors on the paper and pencil post-assessment.

**P5:**

P5's tutor described him as "extremely bright", but challenged in executive functioning and attentional management. P5's favourite school subjects were science and art. He shared P5's interest in constructive toys, such as Legos and K'Nex. P5's tutor reported that he was a very talkative child. He enjoyed learning and readily memorized information. He enjoyed talking about what he had learned; his favourite subjects were cities and animals.

P5 sometimes had difficulty focusing on PhonoBlocks. He was distracted by the tangible letters. He frequently retrieved letters that were irrelevant to the current activity (e.g., the "x", "z" and "y") and assembled them into structures. He sometimes brought other tangible objects into the assessment room (e.g., a stress ball, a plastic finger puppet) and played with them. If the room contained tangible objects he would retrieve them and be unable to focus on PhonoBlocks.

P5 expressed interest in how PhonoBlocks worked. During his first session P5 attempted to open the tangible letters to see "how they light up". P5 appeared interested in the pogo pins on the platform slots. He touched them (dislodging some) and asked "what [these things] are for". P5 seemed to become more engaged with PhonoBlocks as the sessions progressed. On P5's sixth session, before the software had loaded and presented his word, P5 used the letters to spell "dad". P5 also expressed curiosity about the colour codes. During the pre-assessment, at which time I withheld the meaning of the colours, P5 asked why the colours changed.

P5 performed typically at pre-test, spelling 2 out of 16 consonant-le words correctly. He committed nine gemination errors and nine consonant-le formation error. P5 achieved perfect post-test accuracy for coloured words. He performed typically (four gemination errors; one consonant-le formation error; four out of eight words spelled correctly) for uncoloured words.

P5 yielded variable performance during the experimental sessions. Despite some early instances of correct first submissions, for the middle sessions (2 through 8) P5 required a modal three submissions before spelling the word correctly. P5 achieved correct second responses on sessions 6 and 7. P5 achieved a correct first response on session 8, and he sustained this for the remaining sessions.

Session 6 occurred on May 7<sup>th</sup>. On May 7<sup>th</sup>, I observed two cases that suggested that the changes in the colour codes helped P5 to notice the change in vowel sound category, and to remember the necessary change in orthography.

First, P5 had to spell the word “noble”. He first spelled the word “nobble” and submitted it to PhonoBlocks. PhonoBlocks indicated that his word was incorrect. Before I or his tutor provided feedback, P5 removed the first “b” and shifted “no” to the right. I attempted to explain to P5 that he had to shift the “ble” left, because the words had to begin at the 0<sup>th</sup> index, but P5 interrupted me and said (pointing to the “o”): “But this isn’t red”. (The “o” was white when P5 shifted it to the slot previously occupied by the “b” because I had not entered the new letter into PhonoBlocks). I asked P5 why he wanted the vowel to be red. He replied: “Because then it would say noble.”

Second, P5 had to spell the word “settle”. He first created the word “seetl” (both “es” were red). P5 removed the second “e” and moved the “t” against the first “e”. The “e” changed to yellow. When the “e” changed to yellow, P5 exclaimed:

“Oh, *set!*”

P5 then shifted the “l” right, inserted a “t”, and added “e”.

P5' use paralleled P2's. Like P2, P5 seemed to use the colours as a means of "checking" his response for correctness *before* submitting it to PhonoBlocks. By session 8, P5 could seemingly use the colours to achieve correct initial spellings. P5' behaviour on session 8 was consistent with the assumption that he could "read" the sounds of the vowels from the colours. P5 had to spell the word "boggle". He placed the first "g", and "l" and then added an "e". PhonoBlocks recognized "ble" as a syllable and coloured the "o" red.

P5: "Wait no- two gs."

P5 then removed the "g","l" and "e", shifted them right, and inserted an extra "g".

EC: "How come you remembered that?"

P5: "Because it went red."

Like P2, P5 could explain what the colours signified. He could "read" words by reading the colours, and determine whether his submission was correct. Unlike P2, P5 could not explain why doubling the consonant changed the vowel's sound:

EC: "So for the first word, cattle. Why did we need two ts?"

P5: "I don't know".

EC: "Come on [*anonymized*]. How about a guess?"

P5: "Because the person made it that way."

Although P5 seemed to use the colours primarily as a means of "checking" his responses, after session 8, it was not only P5's first *submissions* that were correct, but P5's first *complete words*. For sessions 8 through 11 (the first session is indexed 0) P5 spelled the words correctly with little deliberation and no changes. From this behaviour I had predicted that P5's performance would transfer, because it did not appear as though he was "guessing" and "checking". Despite his inability to explain why short and long vowel required two or one consonant, he seemed capable of judging how many

consonants were needed *before* even checking the colours. This behaviour is difficult to reconcile with his poor performance with the uncoloured transfer words, which suggests that he could not judge- from the vowel sound- how many consonants a word needed.

## **P7**

P7 had a hearing impairment and seemed to have difficulty segmenting phonemes and sequencing letters. Her tutor had provided a strategy of enunciating each sound. P7 used this strategy during her sessions. Her enunciations frequently omitted or reversed sounds (e.g., enunciating “cuddle”: *Kuh-uh-el-d*”, and providing an initial submission of “culd”). P7’s tutor noted that P7 required time to become comfortable around strangers, but that she formed strong emotional attachments to those she trusted, which helped her to perform academically. I observed that P7 and her tutor had a close and congenial relationship. P7’s tutor remarked on her “wonderful sense of humor”. I observed many episodes of P7 and her tutor sharing in-jokes or references to previous shared experiences. P7’s favourite school subject was art.

P7 expressed reluctance to use PhonoBlocks, particularly during the early sessions. For a time, we thought that P7 might withdraw from the study. Fortunately, P7 decided to stay. As the sessions continued, P7 became more comfortable with me. She volunteered jokes and information about her life and hobbies. By the fifth session, P7 completed the activities without much complaint.

P7 performed below average at pre-test and post-test, spelling two words correctly at pre-test and one word correctly at post-test. At pre-test, P7 committed 13 consonant-le formation and 7 consonant gemination errors. At post-test, P7 committed 11 consonant-le formation and 6 consonant gemination errors.

P7 yielded variable performance during the experimental sessions. She typically required at least two submissions before spelling a word correctly, but on four sessions she required three submissions, and on two she required four. P7’s performance did not stabilize, though between the fourth and final sessions she vacillated between 1 and 2 incorrect initial submissions.

P7's memory of the colour codes' meaning seemed unstable. On the third session, she was capable of stating that yellow meant short and red meant long vowel. P7 did not appear to use the colours. By the eighth session, P7 could not state what the colours signified, suggesting "when it's yellow, it's a consonant." On the eighth session I reminded her that red meant long and that yellow meant short. As usual, (section 3.3.4) my feedback on her incorrect responses referred to the colours. I emphasized that red meant long, yellow short. I showed her the colours in the contexts of the syllables "se" and "set", portions of the word (settle) that she was supposed to create.

P7's behaviour on the subsequent session suggested that she remembered what the colour codes signified, and that she could use the colours to determine whether her submission was correct. P7 had to spell the word "boggle". She initially spelled the word "boleg". (PhonoBlocks performs "ro/bin" division, so the vowel in "boleg" was long, and the o was red). After submitting "boleg" (incorrect) I provided P7 feedback directed to the missing consonant-le syllable. P7 rearranged the final three letters to spell the unit "gle". Her next submission was "bogle" (in which the vowel was also red). "Bogle" was also incorrect. Before I provided feedback, P7 said:

"We have to make it yellow."

P7 removed the "l" and "e", producing the word "bog". The "o" in "bog" was yellow. P7 was uncertain how to proceed. I attempted to guide her by asking her to reflect on how she created the "gle" sound in previous words. P7 did not respond. I explained the concept of syllable division and the stable consonant-le syllable. P7 looked at the table during my explanation. At one point, P7 interrupted me:

P7: "Sorry. I was in my imagination. I'd just rather be there than here."

On previous and subsequent sessions P7 was unable to explain the reason for why we doubled the consonant in some but not other words.

#### **P9:**

P9 was outgoing and friendly. She was also relatively old (grade 7). She enjoyed talking to me about her weekend plans, after school activities, personal life, friends,

boyfriends, etc., during the sessions. P9 sometimes approached me in the hallways or at lunch time to chat. P9's favorite subjects were PE and art. She enjoyed dancing.

P9 presented some difficulties focusing on the PhonoBlocks tasks. When she was uncertain how to spell a word she tended to initiate a discussion about her life or some other topic unrelated to the task. She sometimes behaved as though she were physically tired or sleepy. P9's sessions occurred in the final block of the day, so it is possible that she was genuinely tired. She made frequent references to her desire for a "red bull". She occasionally made self-deprecating remarks about her poor memory for the linguistic concepts or colours that PhonoBlocks involved.

On some occasions, P9 became distracted by what she perceived as humorous software events. For example, when she had to spell the word "title", the first placement ("t") formed the word "tit". P9's reactions suggested some awareness of her relative age:

P9: *laughs*.

P9: "Did anyone else laugh at that?"

EC: "Nope."

P9: "They're too young."

P9 was also amused by PhonoBlocks' hardware failures. When PhonoBlocks showed the wrong colour or I entered the wrong letter, P9 typically noticed and laughed. P9 commented that her favorite part about PhonoBlocks was "when it broke".

P9 performed below average at pre-test, spelling one out of 16 words correctly. She committed 8 gemination and 15 consonant-le formation errors. P9 performed better at post-test, spelling 10 words correctly. Much of her improvement involved consonant-le formation. At post-test she committed 2 consonant-le formation and 5 gemination errors.

P9's experimental performance was stable. By session five P9 had virtually eliminated her consonant-le formation errors. P9's gemination errors were resistant to

intervention. With two exceptions (session 5 and 10) P9 geminated incorrectly, and required 2 submissions to spell the word correctly.

Like P8, P9 seemed to learn and exploit the fact that all words required either one or two consonants, although P9 seemed to put more effort than P8 into understanding the rules, at least in the sense of trying- more so than P8- to spell the word correctly the first time.

P9 remembered what the colours signified, but she could not explain the rule. On one occasion, P9 presented some awareness of the association between colour and sound, but the colours misled her into an incorrect submission. The case occurred towards the middle of the study (around session 5). The child had to spell the word “maple”. Her first placements were “p”, and “p”, (creating “mapp”), as though she intended to spell the word “mapple”. Although the vowel changed from red to yellow after she placed the first “p”, she only remarked on the change after the second “p”. She said: “oh, it should be yellow. I have to make it yellow.” She removed both “ps” and inserted a “y” after the “a” (presumably because the digraph “ay” sounds like long a). She then replaced the two “ps”, and added an “l”. Her first (incorrect) submission was “mayppl”.

### ***Children who did not spontaneously use colour***

I observed no cases of any other child from the categorical scheme explicitly using the colours. I observed no explicit cases of children in the particular or categorical scheme using the common consonant-le colours. This section summarizes the observations that I collected for each child, with respect to their profiles (interests, unique challenges), general behaviour throughout the sessions and assessments and their responses to my questions. The remaining three children from the categorical scheme were P1, P8 and P9. The children in the particular scheme were P4, P6, P10 and P4.

#### **P1:**

P1 was quiet and shy. P1 had mildly dysfluent (slow) speech. P1’s tutor described him as a “good speller”, but said that he had difficulties in decoding and

applying rules. P1's favourite subjects were PE and art. P1 expressed neither interest nor resentment of PhonoBlocks. P1 was attentive and compliant.

P1 performed typically at pre-test, spelling 4 out of 16 words correctly, with 9 consonant-le formation and 8 gemination errors. P1 performed worse at post-test. He spelled only one word correctly, and committed 13 consonant-le formation and 8 gemination errors.

P1's performance during the sessions was stable. P1 typically required 2 submissions before spelling the word correctly (with the exception of one session, where he required three submissions, he required only two after session 4). P1 committed fewer consonant-le errors as the sessions progressed (the modal number for the latter half of the sessions being 0). P1's gemination errors remained frequent. The modal number of gemination errors for the first half of sessions was 1; for the second, it was 2.

P1 remembered the meaning of the colour codes, but he did not use them without prompting. When P1 erred in gemination, his tutor and I provided feedback. P1 did not fix the error following his tutors' feedback. P1 sometimes fixed his gemination errors after I drew his attention to the colours:

EC: "What colour is the vowel?" (*the vowel was red, because P1 had misspelt the word "supple" as "suple"*).

P1: "Two ps."

P1 then fixed the word. Although P1's response suggests that he associated the colour with a relevant linguistic property, most probably orthography, I observed no cases of P1 spontaneously using the colours as P2 or P5 did, i.e., to check the pronunciations of his words.

I was uncertain whether P1 understood how short versus long vowels sounded. On one occasion, I attempted to guide P1 to realize that he had to add a consonant by de-selecting the letters "tle" in his misspelling of "settle" ("setle"). I prompted P1 to read "se". P1 responded "set".

**P8:**

P8 suffered red-green colour blindness. I consequently changed P8's colour codes (section 3.3.2). P8's behaviour suggested that he was aware and sensitive of his colour-blindness. During our first session, I showed P8 the two different vowel colours I had chosen and asked him if they appeared different. P8 refused to answer:

P8: "Is this about my colour blindness?"

EC: "Yes."

P8: "I'd rather not answer that."

Although I cannot confirm that P8 discriminated the colours, P8's behaviour on subsequent sessions suggested that he could (when a colour changed, I asked him "did you see what happened?" P8 responded "The vowel changed colour").

P8 appeared clever and observant. By the third session, P8 had more or less eliminated his consonant-le formation errors, so all errors involved gemination. P8 was the first child to explicitly exploit and mention that all of PhonoBlocks' words required either one or two consonants. He seemed to enjoy- and take pride in (he asked me if "any other kids figured it out") exploiting this pattern:

EC: "Well, how about you try- answering correctly the first time?"

P8: [*Laughs*]

Although P8 was compliant and good humored, he did not enjoy PhonoBlocks. He remarked that he "hated its voice", and began each session by asking me "how many words we [had] to do". P8's favorite school subject was PE. His interests were sports.

P8 performed typically at pre-test, spelling 5 out of 16 words correctly, with 5 consonant-le formation and 8 gemination errors. He performed below-average at post-test, spelling 7 out of 16 words correct. All of P8's post-test errors involved gemination. P8 could explain what the colour codes meant, but, as might be expected by his reliance on a strategy that required minimal comprehension of the rule, P8 could not explain why

some words had two and some had one consonant. P8's comments during the post-test (which omitted feedback, randomized the order of short and long vowel words, and permitted one response only) belied his reliance on the system feedback:

P8: "It isn't working."

EC: "What do you mean?"

P8: "It's not telling me if I'm right."

EC: "No, you only get one chance for these ones."

P8: "That's not fair!"

**P6:**

P6 had the particular scheme. P6 was outgoing and friendly. P6's tutor said that his core challenges were speed (for spelling and reading), automatization and attentional management. P6 had received a diagnosis of ADHD. Like P5, P6 was easily distracted into task-irrelevant behaviour. Unlike P5, who was distracted by physical objects, P6 became distracted by his ideas or thoughts. P6's tutor provided some examples. He sometimes took her computer keyboard and attempted to show her a webpage of cars, P6's favorite topic. P6's favorite school subjects were "all of them", (though he admitted an especial fondness for PE and tutoring). P6 loved cars and car-related activities (HotWheels, racing videos).

P6's tutor believed that P6 was more focused using PhonoBlocks than paper and pencil to spell. My observations corroborated hers. Despite his ADHD, P6 seemed focused on the task. He appeared to enjoy handling the letters and observing their colours, although P6 sometimes appeared distracted when I or his tutor supplied feedback for mistakes. For example, he might look away, or look at his hands, or play with the tangibles.

P6 performed above average at pre-test, spelling 8 out of 16 words correctly. He committed 2 consonant-le formation and 8 gemination errors. At post-test, P6's

performance had apparently improved. He spelled 9 words correctly, but committed more consonant gemination errors (5) and the same number of consonant-le formation errors (2). This suggests that P6's pre-test errors were distributed among more words; P6 spelled more post-test words without any errors, but there some words on which he committed both errors. P6's post-test errors were evenly split between new and old words (he spelled 1 old and 2 new words incorrectly).

P6's assessment performance was stable, and consistent with his post-test performance. P6 never spelled the word correctly the first time, requiring between 2 and 7 submissions. His initial submissions typically involved both consonant gemination and consonant-formation errors.

#### **P4**

P4 was quiet and shy. Her tutor explained that P4 had refused to speak upon arriving at KGMS. P4 was reluctant to communicate directly to me; this did not change as the sessions progressed. P4's tutor described her as a "good decoder", but a poor speller. She reported that P4 had difficulty applying spelling rules. P4's favorite subject was art. She loved fantasy (fairies, playing make-believe, My Little Pony) and drawing. Like P7, P4 appeared to have a good relationship with her tutor. Although P4 would not look to me or answer my questions she could sometimes be coaxed to respond to her tutor. P4s tutor believed that P4 enjoyed spelling with PhonoBlocks and that P4 found the bright and varied colours (P4 saw more colours than those in the categorical scheme, because she had the particular scheme) and tangible letters made spelling more enjoyable for her. P4 seemed to enjoy handling the tangibles. She frequently ran her hands over the letters while and after she spelled the words.

Although P4 did not appear to attend to the colour feedback (she frequently looked away, handled irrelevant letters), she attended to her tutor. She made eye contact with her tutor and nodded as she or I explained the gemination and consonant-le concepts. P4 then seemed to attempt to apply the tutors' strategies. For example, during post-assessment, I observed P4 copy her tutors' hand gesture for "splitting" the word into syllables (dividing before consonant-le) on the tangible letters, and murmur her tutors' verbal mnemonics ("big brother", "consonant-le").

P4 performed above average at pre-test, spelling 7 words correctly, with 8 gemination and 4 consonant-le errors. At post-test, P4 spelled 10 words correctly, and committed no consonant-le errors. P4 improved less upon consonant gemination, committing 6 gemination errors at post-test.

P4's experimental session performance was consistent with her post-test performance. She typically required 2 submissions before spelling the word correctly, indicating an ability to use her tutors' and my feedback. Her errors involved gemination. She had virtually eliminated consonant-le formation errors by session 5.

P4 refused to answer my request for her to explain why we sometimes doubled the consonant. Her general inability to determine when to geminate suggests that she did not understand the rule.

### **P3:**

P3 was compliant and attentive. P3 expressed neither interest nor resentment in PhonoBlocks. P3's favorite subjects were art and "emotional-social learning" (a class about self-regulation). P3 enjoyed cooking and baking. P3 was friendly and communicative with me. Like P9, she sometimes volunteered casual information about her weekend or after school plans. Unlike P9, she volunteered these between activities, not during.

P3 performed above average at pre-test (spelling 7 words correctly), with 1 consonant-Le and 8 gemination errors. By post-test, she had eliminated all consonant-le formation errors, but still erred in gemination. She committed 5 gemination errors, and spelled 11 words correctly.

P3's performance during the sessions was somewhat variable. Unlike several of the other children, and in spite of her strong performance forming the consonant-le syllable at post-test, P3 committed consonant-le errors throughout all experimental sessions. Like the other children, her words almost always erred in gemination; she learned how to correct these errors for her second submission.

As with the other children, it is likely that P3 learned the effective response pattern of switching the number of consonants, rather than understanding the rule. P3 could not explain why we sometimes doubled the consonant.

#### **P10:**

P10 was quiet. She was compliant and attentive, and was focused on the task. P10 had a scheduled book club during the time of her experimental sessions. We conducted her sessions before book club and P10 seemed eager to finish her activities quickly, such as to maximize the time for book club. P10's favorite subject was art, although she said that she liked art "less this term because we don't get to draw as much".

P10 performed typically at pre-test. She spelled 6 words correctly, committing 8 gemination and 9 consonant-le formation errors. P10 achieved the best post-test performance. She spelled all words correctly.

P10's trajectory was stable. For the first 2 sessions P10 committed between two and one consonant-le and gemination errors. On the third session, P10 committed neither error. P10 sustained this performance for all subsequent sessions (with one exception, session 7, spelling "m", where she committed one consonant-le formation error-spelling "m" as "mm").

At two points during the study, I asked P10 to explain why she sometimes doubled the consonant. The first time (session 2), P10 did not explain the rule. She responded: "Because I did it that way the last time." The second time (session 6) P10 explained the rule correctly, in terms of syllable division, types and long and short vowel sound.

#### **4.2.5. Summary of Qualitative Analysis:**

Tutors believed that colour codes helped or would help, but one pointed out that children: a) paid little attention to the codes and b) had difficulty understanding the concepts when they were conveyed in different ways. The event logs showed that

children quickly eliminated their consonant-le errors and became capable of fixing their responses on the second submission.

Because I designed the study so that pairs of words required either one or two consonants, children could respond correctly on their second submission by changing the number of consonants, i.e., without understanding the rule. Children's inadequate responses to the question of why some words have one versus two consonants suggested that they did not understand the rule, and either consciously (P8, P9, P10) or unconsciously responded via exploiting this pattern.

Hardware failures caused a delay between the child's changes to the tangible letters and changes in the onscreen letters and their colours. This may have contributed to the lack of attention that most children paid to the colour changes. Somewhat contrary to T1's perceptions, children typically could recite what the colour codes signified, but children failed to identify why the colours changed with respect to the concepts-- syllable division and types-- that I intended the codes to teach.

I observed four cases of children spontaneously attending to and attempting to use the colour codes. All children seemed to remember the association between vowel sound and colour; they could therefore use the colours to tell if their vowel (and thus, submission), sounded correct. Two of these children additionally remembered the orthographic change (doubled or single consonant) that would change the vowel sound. The other two did not, and were either misled by the colours (P9) or required additional support (P7). Finally, I observed no evidence that colour- if it helped- helped children acquire a deeper and transferrable understanding of the rules. P2 of the categorical scheme appeared to use colour and could explain the rule and perform with uncoloured words. But colour was not necessary to achieve such comprehension, because P10 of the particular scheme also achieved it. P5 appeared to use colour, but could not explain the rule, and could not perform with uncoloured words. Colour was thus not sufficient to achieve such comprehension.

To summarize, my case studies provide two "existence proofs" that children can use colour to diagnose and correct spelling errors. Although my data do not permit me to identify the specific mechanism by which colour helped, certain explanations seem more

plausible than others. The idea that colour helped children learn and remember the *spelling rule* is inconsistent with the observation that P5 could not perform with uncoloured letters. The spelling rule has nothing to do with colour, and holds between the vowel sound category and number of consonants. Had P5 learned the rule, hearing the vowel of the assessment words should have enabled him to retrieve the correct number of consonants. A more plausible explanation is that P5 developed a lower-level association between colour, desired sound, and number of consonants. The curio is that P5's association seemed non-transitive: the association between colour and sound and colour and number of consonants did not engender a third association between sound and number of consonants. Because my prediction of transitivity was how I addressed the concern that learning through colour would not transfer, P5's non-transitive associations cast doubt on the viability of colour, at least as I implemented it. On the other hand, it is possible that P5 would have experienced transitivity- and transferred- had the instructional period been longer, or had I implemented the colours differently. In chapter 5, I discuss some ways I might revise the system to promote the development of transitive associations.

P5 and P2 differed in many respects, but shared an interest in science and constructive toys. Children who neither benefitted from colour, nor improved in gemination, shared some features with P2 or P10, but presented a variety of other behaviours or challenges, any of which might have impacted their use of the colours. All children had strong relationships with their tutors and were provided considerable feedback from their tutors following errors. Many children explicitly referred to this feedback during the experimental and assessment sessions. Children were accustomed to this feedback because they had experienced it many times previously. For many children, as T1 pointed out, familiar feedback may have been more accessible- and thus, more children used it- than the colour feedback, which was new, and possibly confusing.

From my qualitative analysis, I propose four factors that might explain children's poor use of the colour codes. These are:

1. Competition with Established Representational Systems
2. Availability of Alternate Response Strategies

3. Lack of Incentive for Correct first Responding or Disincentive for Incorrect First Responding
4. Hardware Failures and Design Limitations
5. Individual Factors (distractibility, sociability, emotionality, curiosity, mental flexibility)

Each factor impacts the child's *attention* to the colour codes. Lessening *attention* to the colour codes would prevent children from remembering and integrating them with orthographic properties or established representational systems.

In the final chapter of this thesis, I elaborate and ground these explanations in the qualitative and quantitative data. I describe how P2 and P5' individual factors may have interacted with [A-C] to enable them to benefit from the colour codes. Although my analysis is primarily applicable to colour-based feedback, I argue that most of my considerations apply to any kind of alternative, representational system. I use [A, B, C and E] to derive guidelines for designers implementing alternate representational systems. My guidelines focus around designing features that enhance the individual factors (cognitive flexibility, curiosity about/comfort with the system) that I think contributed to P2 and P5's success.

## Chapter 5. Discussion

The categorical colour codes were largely ineffective, but two children seemed to use them. Colour seemed to play some role in reducing children's consonant-le formation errors, but I uncovered no evidence of children consciously using the consonant-le colours. The detection of the advantage at pre-test suggests that the advantage did not attribute higher-level awareness of what the common colours represented. Colour-highlighting consonant-le units might have affected children similarly to how Berninger's colour-highlighted units affected children, via lower-level attentional or motivational mechanisms.

In this chapter I discuss what my results suggest about why P2 and P5 could use the categorical codes and why the other children could or did not use the categorical codes. I draw implications for the design of systems that seek to use alternate sensory supports, (including but not limited to colour-coding), to support children in acquiring transferrable conceptual knowledge. The unforeseeable flaws in my assessment words (my children performed at ceiling for vowel discrimination words) and the prototype (the hardware failures that likely reduced the effectiveness of the colour codes), combined with the small number and variability of participants complicate my assessment of my framework's general viability (RQ1 and RQ2). P2 and P5 provide some grounds for believing that sound based colour codes can help children spell, but additional studies are needed to confirm and elaborate the colour codes' role. Although I cannot determine which situational or individual factors affected children's use of the colour codes, I can suggest some candidates, which other researchers may wish to consider. In the final section I enumerate the implications of my observations for the design of future experiments.

## 5.1. Competition with Alternate Representational Systems

Applying dual-coding theory multimedia learning, Alty (Alty, 2002) found that providing learners textual and graphic or verbal representations resulted in poorer concept retention than a textual representation alone, presumably because having to integrate the textual and other representations incurred additional, non-germane processing costs. My participants may have experienced similar effects. All of my children had experienced several years of KGMS schooling. During KGMS schooling, tutors provide children various mnemonics, symbols and gestures that represent certain concepts, and describe or label concepts with consistent terms. For consonant-le gemination, tutors provide children a hand symbol that represents consonant-le, and they use a phrase (“big brother”) to refer to the additional consonant. They enact a procedure of splitting the word into syllables.

T1 observed that the colour codes appeared to confuse her children. She also pointed out that her children had difficulty understanding concepts that were expressed in atypical ways. During my study, I assumed that children would benefit from experiencing their customary feedback alongside my novel, colour feedback. This is why I allowed tutors to repeat their customary feedback to children before I supplemented the feedback with the colour codes. My assumption rested on my belief that children could and would integrate the two sources of feedback, for example, that children would see that their tutors’ hand gesture (“closing” their fingers for closed syllable) had the same meaning as the yellow vowel. The data suggested that my assumption was unwarranted.

The idea that children struggled to integrate their customary and the novel colour feedback is consistent with my observation that children who did not use the colours used their tutors’ feedback, e.g., gesturing syllable division and repeating the mnemonics (“big brother”), albeit unsuccessfully. Because the colours were new, and assuming that using the colours required integration with previously learned representational systems, focusing on the tutors’ strategies, versus attempting to learn my new strategies, may have required less cognitive effort.

The observation that most children in both groups showed an advantage for coloured over uncoloured assessment words in forming the consonant-le syllable is also

consistent with the idea that what impeded the use of the categorical codes was difficulty mentally integrating two different ways of representing the concepts. The advantage for colour-highlighted consonant-le syllables appeared at pre-test. At pre-test, children were not consciously aware of the colour codes' meaning. The children therefore had nothing to mentally integrate, so the colour advantage would not have required explicit mental integration. Accordingly, despite presenting the mild colour benefit, no child (excepting P2 and P10) explicitly articulated the stable syllable concept that the colour-highlighting conveyed; no child appeared to use or referred to the colour-highlighting as they used and referred to their tutor's representations.

## **5.2. Lack of Incentive for Correct first responding (which either learning the colour codes or the rules would enable) or Punishment for Incorrect first Responding**

Had children learned to use the colours as feedback, they might have improved their performance. They could have begun responding correctly the first time, as did P2 and eventually P5. Although I encouraged children to try responding correctly the first time, I offered no incentive to do so. Similarly, I imposed no penalties for incorrect first responses. Rewards for using or punishments for not using the colours might have encouraged children to integrate the customary and new colour feedback, but my study included neither incentive nor punishment. If anything, I may have unintentionally reinforced incorrect first responding. The tutor and I reacted to incorrect first responses by providing the child feedback and various assurances to prevent them from becoming discouraged. Such reactions may have seemed "rewarding" to children like P9 (who enjoyed talking, and may have viewed the feedback as an occasion to initiate conversation), P7 (whose positive relationship with her tutor and apparent dislike of the PhonoBlocks study context, may have driven her to seek her tutors' assurance) or P8 (who seemed to seek recognition of his exploitation of the system).

### 5.3. Availability of Alternate Strategies

The assessment words were predictable: for each pair, one required an extra consonant; one did not; both required a consonant-le syllable. This meant that neither learning to use the colour codes (i.e., determine from the vowel colour whether the spelling was correct) nor understand the rules (i.e., determine whether a word required one or two consonants) were necessary to respond correctly on at least the second submission, though either were necessary and sufficient to respond correctly the first time. Had the words been less predictable (i.e., each pair of words involved one that required one and one that required two consonants), so that understanding neither the codes nor rules resulted in longer sessions, children may have been more motivated to exploit the colours or learn the concepts.

### 5.4. Hardware Failures and Design Limitations

The system failures caused noticeable delays between children's letter placements and the colour changes. Tutor T1 pointed out that the colours were more salient on the screen but that her children did not attend to the screen. She believed that children enjoyed the system, but thought that their focus was more on the plastic letters than the letters' colours. Because of the hardware failures, the letters' colours were frequently faint or incorrect. The mechanism by which I supposed children would discover and benefit from the categorical codes was by noticing the correlated changes in colour and orthography. Provided the children remembered the colours' associations to sounds, this would be equivalent to "seeing" the correlated changes in orthography and sound. Most children could report the association between colour and vowel sound, suggesting they remembered them. But with the exception of P2 and P5, children did not remember the additional association between vowel sound, or colour, and the requisite number or medial consonants. This is why most children did not respond to "incorrect" colours (which they supposedly knew implied the wrong vowel sound) by changing the number of medial consonants.

To associate the colours (and sound changes) to orthography, children must *notice the changes in colour* that correlate changes to letters. The temporal window for

forging low level associations between different sensory events is small, about 300ms (Powers, Hillock & Wallace, 2009). The delay- as well as the distraction that observing me entering the words into the system presented- likely prevented children from forming these associations.

Another reason that children may have failed to associate colour and orthography was the physical distance between the tangible vowel and consonant-le letters. A crucial orthographic change was the child's completion of the consonant-le syllable. When they completed the syllable in a word with one consonant, the vowel changed colour from yellow to red. This change was supposed to help children understand the relation between syllable division, type and vowel sound. The tangible "e" was roughly one foot away from the vowel. The tangibles engaged children. Children focused on the tangibles they manipulated, but no others. Accordingly, as they added the "e" to "l" and "e", they sometimes had their backs to the vowels, or generally failed to look at them again. Given that children with dyslexia have small attention spans (Bosse, Tainturier & Valdois, 2007), it is unsurprising that participants failed to recognize the changes in colour that correlated the completion of consonant-le.

## **5.5. Individual Factors**

The children exhibited many individual differences that may have exacerbated or mitigated the issues I described in sub-sections 5.1-5.4. Of the categorical scheme, P1 was compliant but seemed unengaged with PhonoBlocks. P7 seemed resentful of missing her tutoring sessions. She may have disengaged with PhonoBlocks. P7's hearing impediment caused her some difficulties in understanding the assessment words. I sometimes had to repeat the words for her. She occasionally became frustrated during these occasions and expressed wishes for the session to end. Her phonological sequencing problems may have exacerbated her working memory load and further prevented her from learning and integrating the colour codes with her accustomed mnemonics and representational systems. Indeed, having to negotiate two different representational systems might have exacerbated her phonological challenges. P8 was also resentful of PhonoBlocks, though to a lesser extent than P7. P8 frequently voiced his dislike of PhonoBlocks' voice (google text-to-speech American female). Like P7, he

began each session by asking “how many words do we have to do?”. The pleasure that P8 took in “gaming” the system meant there was little reason for him to learn to use the colour codes. Finally, although P9 was attentive and compliant, some of P9’s comments and behaviour (e.g., chuckling at the sub-word “tit”) suggested that her relative age prevented her from taking PhonoBlocks seriously. An additional factor that may have prevented P9 from engaging with PhonoBlocks was her sociability. P9 made efforts to befriend me. She seemed to relate to me as she might an older sister or friend. During the sessions, she frequently attempted to engage me in discussions about boys she was interested in and her and her friends’ activities. Presumably, PhonoBlocks could not compete with these attentional distractions.

By contrast, P2 and P5 demonstrated attention to the colour changes and to PhonoBlocks in general. P2 was quiet and attentive to the task. P2 volunteered virtually no information about himself, beyond responding “fine” to my customary greeting (“How are you?”) at the beginning of the sessions. Like P9, P5 was easily distracted, especially during the initial sessions. A potentially relevant difference between P5 and P9 was that P5 was distracted by physical objects, whereas P9 was distracted by thoughts or ideas. This meant that I could substitute P5’ extraneous distractions (a squishy green object, for example), with PhonoBlocks. By the 4<sup>th</sup> session, P5 had re-directed his physical attentions to PhonoBlocks. Although some of his behaviour with PhonoBlocks suggested a lack of attention to the colours or feedback (such as when he built a structure out of irrelevant letters instead of attending to the screen), the system interested him. The other children did not ask me how PhonoBlocks worked or attempt to disassemble it; P5 did both<sup>12</sup>. His curiosity extended to the colour codes.

I suspect that P5’s inherent curiosity about the system contributed to his apparent capacity to associate the colours to the requisite number of consonants as well as to the vowel sounds. Curiosity might cause a child to attend to and attempt to explain an otherwise inexplicable event. Assuming that the children did not comprehend why the colours changed (at least initially), the colour changes would be an inexplicable event.

---

<sup>12</sup> I believe that P5 was responsible for the hardware failures.

One way of explaining an event is finding other factors that correlate it. It is possible that P5's curiosity drove him to seek- and thus to form- associations between the colour changes, sound changes and specific changes in orthography. P5 did not transfer performance to uncoloured words nor could P5 articulate the deeper reason for the correlation in terms of the linguistic concepts, syllable types and division. P5's curiosity seemingly did not drive him to integrate the colour codes with other representations of the linguistic concepts. I suspect that P5's poor attention to the tutor's and my feedback, which involved considerable verbal delivery, prevented him from relating the correlated colour/sound and orthographic changes to the linguistic concepts.

P2's improvement occurred early (session 2). By session 2, P2's behaviour suggested that he had associated colours and vowel sounds and orthographic changes. P2 could also explain the underlying rule; P2 transferred performance to uncoloured letters. In contrast to P5, during his early sessions (when P2 still geminated incorrectly) P2 attended to his tutors' feedback. Although P2's first use of colour was similar to P5's, (he determined that his word sounded wrong *by looking at the colours*), and suggested that he associated sounds and consonants via the colours, his ability to perform with uncoloured words suggests that he learned to map the sounds of the vowels in the words directly to the number of consonants. P2 thus achieved a deeper understanding of the rule than P5.

My data do not allow me to conclude why or by which processes P2 achieved this deeper comprehension. P2's tutor claimed that the consonant-le concept had been difficult for him to understand. P2's performed at about chance on germination at pre-test. The tutor feedback that P2 had received for several years before the study was therefore insufficient for him to understand the rule.

P2's tutor believed that the colour changes helped. It is possible that the colour changes provided P2 a final piece of information, or triggered a kind of "insight", that was necessary for him to understand his tutor's feedback. I had envisioned the colour codes performing this role. In this case, children's ability to mentally integrate the colour changes with their tutors' feedback would predict their use of the colour codes.

The ease of integrating various pieces of information is related to cognitive flexibility (the ability to hold and switch between multiple pieces of information in working memory), which is in turn related to executive functioning- a class of interrelated processes known to be predictive of literacy and compromised in children with dyslexia. I did not formally measure children's cognitive flexibility. I can only conjecture about it based upon their interests and in-session behaviour. P2 expressed interest in science and mathematics. Both disciplines, particularly mathematics, require learners to reconcile multiple representations of concepts into a more general, abstract description (Lakoff & Nunez, 1986). An interest in mathematics might predict a high degree of cognitive flexibility or abstractive/integrative capacity. Based on T1's comments, and my observations of children's use of my and the tutors' feedback, I think it possible that P2's ability to abstract helped him to acquire and transfer the gemination rule.

That said, I have no evidence that the colour codes played an independent role in helping P2 abstract or integrate them with his pre-existing knowledge or his tutors' feedback. P10 of the particular group also achieved transferrable and articulable comprehension that she did not have at pre-test. The colours were therefore not necessary to catalyze a deeper comprehension of the tutors' feedback. As to what factors are necessary, P10 was similar to P2 in that she was compliant, un-talkative, and focused on the task, but yielded no data that suggests a higher than average abstractive ability or mental flexibility. Her interests (art/drawing) were shared by many children who did not improve. Studies that systematically measure children's cognitive flexibility are needed to determine if it plays a role in children's capacity to benefit from novel kinds of feedback (the colours or tangibility that both conditions shared) or intensive instruction.

It is also possible that different factors underlay P2 and P10's superior performance. P10 might have been more motivated than the other children to finish the sessions quickly, which would require her to submit correct first responses. P10's sessions preceded her book club. The faster P10 completed her sessions, the faster she could join book club. I mentioned that the lack of an incentive for the other children to answer correctly the first time (i.e., by exploiting the colour codes or by understanding the rule) might have prevented them from using them. If this is true, then the greater incentive that P10 had to answer correctly the first time (and thereby avoid the lengthy

feedback that followed an incorrect first submission) may have motivated her to learn the rule.

## **5.6. Implications for Design**

The two children who seemed to successfully use the colour-codes achieved different levels of conceptual understanding. P5 seemed to forge lower-level associations between colour and sound and orthographic change, but not higher-level conceptual knowledge. P2 likely engaged additional mechanisms to P5 that integrated the colour with his tutors' feedback and enabled him to achieve transferrable conceptual knowledge. The remaining children seemed to form half of P5's associations, in that they remembered which vowel sounds the colours represented, but not the second half- the association with one or two medial consonants.

Although it is probable that individual factors unique to P2 and P5 enabled them to benefit from the colour codes, there might be ways to modify the system such as to help other children- those less naturally inclined towards curiosity or integration- to benefit in the same way. In this section, I propose some recommendations for designers seeking to help children learn an abstract concept through novel sensory feedback. I divide my recommendations into two sections, each based around helping children achieve P5' or P2's levels of comprehension.

The first recommendations involve helping children forge the lower-level associations that enabled P5 to perform with coloured letters. These are:

- Increase Engagement with the System by a) reducing discomfort with the system and b) increasing perceived age-appropriateness

- Integrate Correlated Sensory Events

- Incentivize Effective Exploitation of the Novel Feedback

The second recommendations involve helping children integrate these associations with higher-level conceptual representations to forge the deeper understanding that enabled P2 to articulate the rules and perform with uncoloured tangible and paper and pencil letters. These were:

Encourage reflection on the connections between alternate representations

Incentivize Abstraction

Although my study and analysis focused on colour, I base my recommendations on general learning mechanisms. My recommendations likely apply to systems that use alternate modalities (auditory, haptic or other visual features) to communicate concepts to children.

### **5.6.1. Engage Low-Level Attention to the Novel Feedback**

P5 demonstrated an intrinsic curiosity about the system that may have caused him to attend to the changes in the colour, and to explain them by locating other visual events (i.e., changes in orthography) with which they correlated. Other children demonstrated less attention to the correlated changes in orthography and vowel colour; each child had behavioural tendencies that may have interacted with different aspects of the system to prevent them from attending. Changing the system's features could compensate for children's antagonistic behavioural tendencies. I identified three classes of compensatory changes: increase engagement with the system, integrate correlated sensory events, and incentivize effective exploitation of novel feedback.

#### ***Increase Engagement with the System***

P5 seemed naturally curious about physical objects and PhonoBlocks in general. Because of this, even though P5 was easily distracted, he was naturally interested in and thus engaged with PhonoBlocks. Other children were less engaged, but for different reasons. Each reason warrants a different design strategy:

#### **Reduce Discomfort with the System**

P7 seemed to resent her PhonoBlocks sessions. She disliked losing her tutoring time and presented aversive emotional reactions to the system and activities. Children sometimes mentally disassociate from situations that cause aversive emotions (Foa & Hearst-Ikeda, 1996). P7's comment about wanting to "be in her imagination", rather than with the system, which occurred as I was delivering feedback for an error, is consistent

with the idea that she mentally disassociated. Reduced attention is a consequence of mental disassociation. Although P7's aversive reaction was uncommonly strong, and most children seemed to enjoy missing tutor time, designers of systems like PhonoBlocks may need to contend with children who are resentful of using the systems, and provide features that acclimatize children to them. One strategy is increasing children's comfort with the system. Designers might increase children's comfort with alternate systems by providing features that mimic children's traditional learning environments, and avoiding features that seem "foreign", artificial, uncanny and overly technical. For example, one feature that children (notably P7 and P8, but also P9) frequently criticized was PhonoBlocks' voice. PhonoBlocks' had the google text-to-speech female voice. Although it mimicked some natural human cadence, it sounded artificial, and occasionally mispronounced words. How PhonoBlocks used the voice may have also provoked negative reactions. PhonoBlocks provided the same voiced feedback ("That's not quite it. Would you like a hint?") for every error. Children strongly disliked the error feedback. They cringed, or covered their ears, before and after receiving it. Children's dislike of PhonoBlocks' voice may have caused them to disengage.

Children's dislike of PhonoBlocks' voice might have attributed the "uncanny valley effect": the sense of discomfort that attaches to simulations that approximate actual humans with a high degree of fidelity, but err in slight and noticeable ways (Seyama & Nagayama, 2007). PhonoBlocks' sounded more "like a human" than iconic, 80's era computer speech, but its cadence and syntax were unnaturally static. Humans vary their cadence and the words by which they express propositions. PhonoBlocks' repeated the exact same sentences (instruction, hint, error and congratulations) over the assessments. If the "uncanny valley" effect contributes to children's discomfort with software systems, designers could lessen discomfort by a) substituting human voice feedback for non-human sounds, e.g., a "plonk" sound for errors, a "ding" for correct or b) employing human voice actors and recording multiple (syntactically distinct but semantically equivalent) versions of each instruction, hint and feedback.

### **Increase Perceived Age-Inappropriateness**

P9 may have disengaged with PhonoBlocks because she considered PhonoBlocks immature. PhonoBlocks' visual interface resembled an elementary school classroom, and PhonoBlocks' sole activity (word completion) was typical of lower-grade instruction. Older children may have felt patronized, particularly when assessment included the vowel discrimination words, which were very easy. Designers might increase the interest of students like P9 by modelling interfaces that appeal to older students (such as an iPhone screen or computer desktop), or by interleaving simple activities with conceptually equivalent but superficially more complicated varieties. For example, we might have presented the words for completion within a paragraph concerning an age-appropriate topic.

### **Increase Low-Level Salience of Onscreen Elements:**

T1 believed that children paid insufficient attention to the screen, and children would have understood the colours better had they attended to the screen. An alleged reason was that the colours in the tangible letters were comparatively less salient. A reviewer of this thesis posed a similar critique of the onscreen letters. The default colours of the Unity renderer had different subjective intensities, with red and blue appearing darker than green, yellow, magenta and cyan. This means that the intensity differences between the letters and the background were unequal. Because figure-ground luminance contrasts directly correlate salience, differences in salience from colour were confounded with differences in luminance contrast. The confounding luminance intensity differences could have increased the children's confusion via implying a non-existent linguistic difference between the letters. My choice of dark gray (versus black) for the background colour compounded the problem because it resulted in overall less-salient figure-background luminance contrasts. Future iterations of our prototype, particularly those emphasizing screen-based interaction, will equate and maximize the colours' luminance levels and background luminance contrasts.

### ***Integrate Correlated Sensory Events***

Another issue was children's apparent difficulty noticing the correlated changes in orthography and colour, which represented the change in orthography and sound.

Hardware failures that are easily fixed likely contributed, but another likely contributed was the spatial arrangement of the tangible letters and screen. Children focused on the letters they manipulated, and frequently missed the changes in the colours of the vowels. Similarly, because I designed the system such that hearing the letters' sounds required multi-touch grouping and tapping, children did not hear the change in the sound at the same time they changed the letters and (possibly) saw the changed colour.

I did not implement simultaneous auditory events because I assumed they would enable children to spell the words through (pre-submission) trial and error; such considerations are appropriate for practice or assessment contexts. Conversely, in instructional contexts, simultaneous auditory representations might help children learn the correlations between novel sensory, orthographic and phonological information.

Instruction is one of PhonoBlocks' intended use contexts. In our case, simultaneous auditory representations could solve two design problems: children's physical inability to see the changes in vowel colour when handling letters in distant parts of the word, and the need for an explicit cue (such as the tutor or I telling them) to direct children's attention to the colours. Children's failure to attend to the colour changes presumably resulted from their failure to understand why the colour changes mattered, which in turn attributed their failure to understand that the colour changes indicated changes in the vowel sound.

We could solve both problems by embedding each tangible letter with a small audio emitter. The audio emitter could behave like the colours. Each time a letter's sound changed, the letter would omit it. Auditory localization is good, and humans are naturally biased to re-focus attention to the location of auditory events (Blauert, 1997). Vowels that emitted their new sound at the same time that they changed colour would re-focus children's attention to them, even if the letters were outside the child's field of view, and they would leverage multisensory processing encode the correlations between colour and sound.

### ***Incentivize Effective Exploitation of Novel Feedback***

P5 may have been intrinsically motivated to discover what the colour changes correlated. Other children seemed less motivated. One way of increasing children's attention to the correlated changes in colour and sound might be increasing the incentives for learning the correlation. We could increase the incentive in three ways: a) show children how understanding the meaning of the colour feedback would enable them to solve the problems without extra feedback b) reward correct first responses c) avoid implicitly rewarding incorrect first responses.

I attempted (a) in my initial and all subsequent sessions, when I provided children feedback that pointed out how the colours reflected the vowel sound or consonant-le units. I did neither (b) nor (c). Reflecting on my observations, I would advise against (c). Explicitly punishing incorrect responses would probably decrease children's comfort and thus engagement with the system. (b), increasing motivation to respond correctly, would likely suffice.

In the school, I observed one feature that might satisfy (b). In the hallway, there was a poster board showing the number of evenings of that month that each child engaged in personal reading. This resembles the "leaderboard" concept from online multi-player games. Designers might implement similar features. In our case, at the beginning of each session, children could see a screen that would show the distributions of correct first, second, etc., responses of children from the *previous* session. Children who responded correctly the first time could be highlighted with a star or some other unique visual mark. Children's names could be anonymized (i.e., given user names) to avoid eliciting negative intra-cohort hostility. This type of strategy might work best with students like P8 or P9, both of whom seemed somewhat sensitive to the achievements of the peers (recall P8 asking "how many other kids" figured out the word response pattern, or P9 asking if other children "got" the humor behind the sub-word "tit").

### **5.6.2. Engage High-Level Mental Integration**

I would expect the strategies that I recommend in 5.6.1 to help children achieve a level of comprehension and performance similar to P5's. Children would associate vowel

colour to sound and to orthographic change, but this would not suffice for a deeper understanding of the underlying rule. Consequently, children would perform well with coloured words (with which they could check their responses), but not uncoloured words (with which children must rely on an integrated correspondence between vowel sound and orthography, versus independent associations between vowel colour and sound and orthography).

The next step is achieving P2 and P10's comprehension. To do so might require children to integrate the colour codes (or any novel feedback) with their pre-existing knowledge and the feedback their tutors supply. T1 remarked that her children struggled to integrate multiple ways of expressing a concept. P2 may have presented better-than-average cognitive flexibility, which enabled him to integrate and abstract over these multiple representations. Designers could implement features that would support children who are less naturally capable of mental integration or generalization.

### ***Encourage Reflection on the Connections Between Alternate Representations***

Abstracting or “generalizing” over multiple representations is an essential skill in science, mathematics and various other disciplines (Goldin-Meadow, Alibali, & Church, 1993). Abstraction is a struggle for all children (Uttal, O'Doherty, Newland, Hand, & DeLoache, 2009). So far, programs directed towards helping children abstract have focused on identifying a representational format that is most “natural” or sensible to children. For example, Lakoff exposed children to physical or “embodied” representations of mathematical concepts because he considered them fundamental (Lakoff & Nunez, 2000). In a recent short paper, I argued that a limitation of these programs is their failure to explicitly integrate the so-called natural physical with unnatural- but essential- symbolic representations (Cramer & Antle, 2015). I synthesized evidence from various other interventionist studies into two design recommendations: a) use alternate representations that share constraints with the abstract concept b) promote reflection on the connections between the alternate and traditional (symbolic) representations. I attempted (a) by choosing colours that were metaphorically connected to the sound category labels. The reflection period with the Word History was how I attempted (b), but I focused the reflection period on the connections between the

colours, the sounds and the orthography. I did not explicitly connect the colour codes to the tutors' ways of representing the same concepts.

T2 reported that they did not typically use a "Word History" or reflection period as a follow-up to Guided Discovery activities. This surprised me because I learned about the Word History and reflection period in a OG-styled multisensory teaching handbook (Birsch, 2011). T2 considered the reflection period "really useful". She thought it helped children integrate some concepts. Children did seem to learn the association between colour and vowel sound category, but less colour and orthography. I clicked on the words to play their sounds, but the Word History font size was small. Consequently, the Word History might have played a greater role in associating colour to sound than either to orthography: the colours and sounds were more salient.

If such interface limitations contributed to children's failure to integrate the colour and other representations, a design guideline is explicating- with equal salience- all representations a child must connect. In our case, we might expand the word history during the reflection period so that it occupies the majority of the screen, and the letters are equally clear as the colours, and include visual cues that show children the explicit mapping between, for example, the colours of the closed syllable (white-yellow-white) and the visual and gestural representations of closed syllable that tutors traditionally use. A picture of the hand symbol, as well as the text CVC, with the breve above the vowel, could appear beneath the coloured word. We could tap on one letter of the word (e.g., the vowel), and see that portion illuminate on each representation, whilst hearing the vowel sound. These cues might help children explicitly connect these various forms of feedback and acquire a more general, abstract comprehension.

A reviewer of this thesis proposed another way to explicate the abstract relational structure of the target spelling rules. As discussed in my design rationale, because my objective was teaching children the gemination rule in a "toy context", I opted for instantaneous changes between the letters of the consonant-le syllable and the initial vowel. The reviewer suggested that children might have an easier time comprehending the rule if I expressed it as a causal relationship (e.g., in words lacking the extra consonant, the child's completion of the consonant-le syllable *causes* the word to

syllabicate into an open syllable and consonant-le, which causes the vowel sound to change from short to long). He pointed out that instantaneous changes mask the causal structure, and that I could explicate the causal structure by imposing a delay between the change in colour of the consonant-le syllable and vowel.

I had not considered this in my initial exposition of my framework because I had not thought of linguistic rules as “causal”. My reasoning was the difficulty establishing “directionality” in the relation between orthographic and phonological elements, *in the abstract*, though I agree that the causal model makes sense in a concrete spelling context, wherein the child can play the causal agent that establishes directionality (i.e., the child's spelling decisions affect the sound of their word).

Future work could explore whether modelling the relationships as explicitly causal would help children understand the underlying justification of the doubling rule, i.e., consonant-le *causes* the word to divide in such a way that the vowel is left with or without an additional consonant, which *causes* its sound. Such predictions are consistent with the constructivist emphasis on the learner as an “active” (causal) agent (Mayer, 2004). Such work could leverage research into how information visualization designers have represented causality, deploying not only temporality, but animations, sounds and other effects.

### ***Incentivize Abstraction***

Finally, P5's poor performance on uncoloured post-assessment words reinforced the need to incentivize children to use alternate supports to *learn*, versus to substitute, generalizable knowledge of orthographic rules. We might have achieved this by interweaving practice sessions with uncoloured words. Assuming we retained the leader-board feature to motivate correct first responses, we would thereby motivate children to a) use the alternate supports to help them respond correctly the first time but b) to use them as a support for developing deeper conceptual knowledge. Part of what we would incentivize is attending to the integrated feedback provided during the reflection period. Just as few children attended to the colour changes, few children (including P5, who developed no transferrable knowledge) attended during the reflection period. Children

might pay greater attention if they believed that the reflection period contained information that would improve their leader-board standing.

### ***Assess the Intuitiveness of the Visual Codes***

In my review of Ehri's (Ehri, Deffner & Wilce, 1984) integrated-picture mnemonics, I emphasized the importance of representing information with visual properties for which the relation to the information is easy to process. "Processing" involves low and higher level operations. At the low level, the visual property must be easily discriminable and easy to maintain in visual working memory. At the higher level, the property and the information should relate somewhat intuitively, such that children do not need to memorize and apply an unwieldy chain of third associations. Colour satisfies the low level requirements, and on the basis of Wrembel's work (Wrembel, 2009), and the relative prevalence of colour-grapheme to other forms of pure and pseudo-synaesthesia (Colizoli, Murre & Rouw, 2012), I thought that colour might be an intuitive channel for linguistic information. In my framework I emphasized the need to balance low and high level considerations in the choice of colour codes. Reflecting on my specific design choices, it is possible that I over-emphasized low-level considerations. My rationale for the colour assignments was an association I expected the children to hold between the colours (red, yellow), and the words "long" and "short". Although children reported that "red" meant "long vowel" and yellow meant "short vowel", I do not believe that they associated the colours with the actual vowel sounds. My evidence is the fact that many children (when asked to articulate, for example, "long" or "red" i) produced the incorrect sound. My approach, therefore, may have been based on some false assumptions.

Future work will dedicate more research to understanding whether children intuitively associate *categories* (versus particular vowel sounds) of sound to particular or to sets of colors (e.g., warm and cool). Future work may also wish to more rigorously test my assumption that color is an ideal channel for conveying all varieties of linguistic information. My advocacy of colour rests more on its low level salience than an argument for its intuitive connection to linguistic information. Other visual properties (for example, visual texture), while less supportive of rapid low-level discrimination and encoding, might more intuitively correspond to differences in sound or in orthography.

Texture, like colour, is appropriately categorical. Future work will probe not only the question of whether “innate” vowel-colour associations extend to categories of phoneme, but also whether other visual dimensions might be more appropriate than color for coding specific kinds of information. One advantage of incorporating texture is that texture and colour (relative to colour and illumination) are independent (Ware, 2012), and could therefore code different properties simultaneously. Such capacities would relax the limitations for DE1, enabling designers to, for example, communicate overall presence of a unit (perhaps by colouring all elements the same), but jointly allow children to confirm the role of each letter (by assigning each letter a different visual texture).

## **5.7. Implications for Study Design**

On the basis of my study, I am uncertain whether my approach to designing colour codes is effective. I can suggest ways that future studies might improve upon mine to more rigorously assess my framework. My methodological recommendations are in addition to my design recommendations. They are:

- Pre-Screen children on their abilities to spell the target words

- Match children on age, favoured school subject and motivation to engage with the system

- Remove all feedback that enables trial and error responding

### **5.7.1. Pre-Screen Children on Their Ability to Spell the Target Words**

I did not pre-screen children on their ability to spell vowel discrimination words. In consequence, I was unaware that my participants performed at ceiling for vowel discrimination words. Lacking an alternative to the categorical scheme’s matched consonant-le activity, my study cannot say whether my general framework- that the grain of any rule should determine the grain of its corresponding code- is legitimate. Future studies will be conducted wherein children’s baseline performance on the target words is measured before the study commences. Another benefit of pre-screening children is the ability to match groups on children’s pre-test facility with the words. Another

methodological compromise was the spurious difference between the categorical and particular groups' pre-assessment facility with consonant-le formation, and thus overall accuracy. Coupled with a low plateau for spelling accuracies for certain kinds of words, such imbalances can lead to spurious differences in pre and post gains.

### **5.7.2. Match Children on Age, Favoured School Subject and Motivation to Engage with the System**

On the basis of my observations, I recommend matching groups of children on several factors that might impact their ability to attend to and use colour codes. I would recommend matching children on their age, favored school subject, (in particular, distinguishing between science and math versus art and physical education), and on their general motivation to engage with the system in question. The latter might be measured with a questionnaire.

Because I did not design my study to assess how these factors impact children's use of colour or other alternative representations, my recommendations are only conjectural. More research is needed, but I believe that my case studies provide some rationale for mounting the necessary investigations.

### **5.7.3. Remove All Feedback that Enables Trial-and-Error Responding**

I had not anticipated that children would learn and exploit the predictable nature of the paired per-session consonant-le words. Children's abilities to exploit the fact that each session contained one word with one consonant and one with two consonants enabled them to achieve correct second responses via "trial and error". Children's ability to respond by trial and error likely contributed the aggregate 50% accuracy "plateau".

The implication for design researchers is that systems must be cleansed of any possible cues or hints that enable trial and error responding. Antle and Wang reached a similar conclusion in their assessment of a collaborative, multi-touch puzzle game. There, the mere addition of a "snapping" sound indicating correct puzzle placement

enabled users to solve the puzzles via a trial-and-error strategy of “blindly” manipulating the pieces until hearing the “click” (Antle & Wang, 2013).

## Chapter 6. Conclusion

In this thesis, I articulated a rationale and a framework for designing colour codes that seek to draw children's attention to orthographic and phonological features that are relevant to understanding a given linguistic rule. I applied my framework to develop two colour schemes, each one tailored to a different literacy rule. Finally, I assessed my framework by implementing the colour codes in a tangible software system and deploying it in a four-week intervention at a school for children with dyslexia. Because my sample performed at ceiling at pre-test for vowel discrimination, my analysis was restricted to consonant gemination in consonant-le words.

I uncovered no aggregate effects of the categorical colour schemes, which I designed to benefit consonant gemination. I uncovered some suggestion of a benefit for colour-highlighting the consonant-le unit in reducing the number of consonant-le formation errors, but lacking either a control group who lacked the colours nor sufficient participants to power my inferential statistics I cannot conclude whether or what role colour might have played. That said, I observed two cases of children that used the categorical codes; only one developed knowledge that transferred to uncoloured words. These children provide an "existence proof" for the potential of colour-codes that are designed by my framework to support children in acquiring spelling rules, if only as a bridge between lower-level sensory associations- between a colour, a sound and a spelling change- and a generalizable comprehension of the underlying principles. I analyzed my and the tutors' qualitative observations of children to identify a) system factors that may have prevented children from using the colour codes, to develop either transferrable or non-transferrable skills and b) individual factors that exacerbated or mitigated the system factors. Although I cannot confirm that my findings would apply to other systems than PhonoBlocks, my observations have motivated various important design revisions. In particular, I observed a need to better explicate the significance of the colour codes, both in terms of their lower-level attentional salience, and their

connections to children's outstanding knowledge. I accordingly used (a) and (b) to develop design recommendations for future iterations of PhonoBlocks or for other software systems that attempt to use alternate representations of concepts to help children understand them. I identified *interest* in the system and *ease of integration* as the key factors affecting children's acquisition of colour-sound and orthographic connections and relating these to the knowledge they tutors supply.

My contributions were my identification and demonstration of colour as a *possible* means of helping children understand an abstract linguistic rule, and my identification of and design recommendations for mitigating factors that prevent the typical child with dyslexia from using colour feedback.

## References

- Alexander, A. W., & Slinger-Constant, A. M. (2004). Current status of treatments for dyslexia: Critical review. *Journal of child neurology*, 19(10), 744-758.
- Altemeier, L. E., Abbott, R. D., & Berninger, V. W. (2008). Executive functions for reading and writing in typical literacy development and dyslexia. *Journal of Clinical and Experimental Neuropsychology*, 30(5), 588-606.
- Alty, J. L. (2002). Dual Coding Theory and Computer Education: Some Media Experiments To Examine the Effects of Different Media on Learning.
- Antle, A. N., & Wang, S. (2013, February). Comparing motor-cognitive strategies for spatial problem solving with tangible and multi-touch interfaces. In *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction* (pp. 65-72). ACM.
- Antle, A. N., Fan, M., & Cramer, E. S. (2015, January). PhonoBlocks: A Tangible System for Supporting Dyslexic Children Learning to Read. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction* (pp. 533-538). ACM.
- Apel, K. (2011). What is orthographic knowledge?. *Language, Speech, and Hearing Services in Schools*, 42(4), 592-603.
- Archer, A. L., Gleason, M. M., & Vachon, V. L. (2003). Decoding and fluency: Foundation skills for struggling older readers. *Learning Disability Quarterly*, 26(2), 89-101.
- Ball, (1997) Phonological awareness implications for whole language and emergent literacy programs
- Banich, M. T., Milham, M. P., Atchley, R. A., Cohen, N. J., Webb, A., Wszalek, T., & Brown, C. (2000). Prefrontal regions play a predominant role in imposing an attentional 'set': evidence from fMRI. *Cognitive Brain Research*, 10(1), 1-9.
- Bender, R. L., & Bender, W. N. (1996). *Computer-Assisted Instruction for Students at Risk for ADHD, Mild Disabilities, or Academic Problems*. Allyn & Bacon Inc., 160 Gould St., Needham Heights, MA 02194.
- Berninger, V. W., Abbott, R. D., Brooksher, R., Lemos, Z., Ogier, S., Zook, D., & Mostafapour, E. (2000). A connectionist approach to making the predictability of English orthography explicit to at-risk beginning readers: Evidence for alternative, effective strategies. *Developmental neuropsychology*, 17(2), 241-271.

- Berninger, V. W., Yates, C., & Lester, K. (1991). Multiple orthographic codes in reading and writing acquisition. *Reading and Writing*, 3(2), 115-149.
- Berninger, V. W., Vaughan, K., Abbott, R. D., Brooks, A., Begayis, K., Curtin, G., ... & Graham, S. (2000). Language-based spelling instruction: Teaching children to make multiple connections between spoken and written words. *Learning Disability Quarterly*, 23(2), 117-1350
- Berninger, V. W., Abbott, R. D., Thomson, J., Wagner, R., Swanson, H. L., Wijsman, E. M., & Raskind, W. (2006). Modeling phonological core deficits within a working memory architecture in children and adults with developmental dyslexia. *Scientific Studies of Reading*, 10(2), 165-198.
- Bhattacharya, A., & Ehri, L. C. (2004). Graphosyllabic analysis helps adolescent struggling readers read and spell words. *Journal of Learning Disabilities*, 37(4), 331-348.
- Bhattacharya, A. (2006). Syllable representation in written spellings of sixth and eighth grade children. *Insights on Learning Disabilities*, 3(1), 43-61.
- Birsh, J. R. (2011). *Multisensory teaching of basic language skills*. Brookes Publishing Company. PO Box 10624, Baltimore, MD 21285.
- Blachman, B. A. (1991). Phonological awareness: Implications for pre-reading and early reading instruction. *Phonological processes in literacy: A tribute to Isabelle Y. Liberman*, 29-36.
- Blauert, J. (1997). *Spatial hearing: the psychophysics of human sound localization*. MIT press.
- Blok, H., Oostdam, R., Otter, M. E., & Overmaat, M. (2002). Computer-assisted instruction in support of beginning reading instruction: A review. *Review of educational research*, 72(1), 101-130.
- Bosse, M. L., Tainturier, M. J., & Valdois, S. (2007). Developmental dyslexia: The visual attention span deficit hypothesis. *Cognition*, 104(2), 198-230.
- Brosnan, M., Demetre, J., Hamill, S., Robson, K., Shepherd, H., & Cody, G. (2002). Executive functioning in adults and children with developmental dyslexia. *Neuropsychologia*, 40(12), 2144-2155.
- Bruck, Maggie, and Rebecca Treiman. "Phonological awareness and spelling in normal children and dyslexics: The case of initial consonant clusters." *Journal of experimental child psychology* 50.1 (1990): 156-178]
- Bruck, M. (1992). Persistence of dyslexics' phonological awareness deficits. *Developmental psychology*, 28(5), 874. .
- Casalis, S., Colé, P., & Sopo, D. (2004). Morphological awareness in developmental dyslexia. *Annals of dyslexia*, 54(1), 114-138.

- Casco, C., Tressoldi, P. E., & Dellantonio, A. (1998). Visual selective attention and reading efficiency are related in children. *Cortex*, 34(4), 531-546.
- Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition*, 47(2), 149-180.
- Cestnick, L., & Coltheart, M. (1999). The relationship between language-processing and visual-processing deficits in developmental dyslexia. *Cognition*, 71(3), 231-255.
- Christ, R. E. (1975). Review and analysis of color coding research for visual displays. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 17(6), 542-570.
- Colizoli, O., Murre, J. M., & Rouw, R. (2012). Pseudo-synesthesia through reading books with colored letters. *PLoS One*, 7(6), e39799.
- Cox, A. R. (1985). Alphabetic phonics: An organization and expansion of Orton-Gillingham. *Annals of Dyslexia*, 35(1), 187-198.
- Cox, A. R. (1974). *Structures and Techniques: Remedial Language Training: Multisensory Teaching for Alphabet Phonics*. Educators Pub. Service.
- Cunningham, A. E., Perry, K. E., Stanovich, K. E., & Share, D. L. (2002). Orthographic learning during reading: Examining the role of self-teaching. *Journal of Experimental Child Psychology*, 82(3), 185-199.
- Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational psychology review*, 3(3), 149-210.
- Cramer, E. S., & Antle, A. N. (2015, January). Button Matrix: How Tangible Interfaces can Structure Physical Experiences for Learning. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction* (pp. 301-304). ACM.
- Cunningham, P. M. (1998). The multisyllabic word dilemma: Helping students build meaning, spell, and read "big" words. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, 14(2), 189-218.
- Das, J. P., Mishra, R. K., & Kirby, J. R. (1994). Cognitive patterns of children with dyslexia: A comparison between groups with high and average nonverbal intelligence. *Journal of Learning Disabilities*.
- Dautrich, B. R. (1993). Visual perceptual differences in the dyslexic reader: Evidence of greater visual peripheral sensitivity to color and letter stimuli. *Perceptual and motor skills*, 76(3), 755-764.
- de Jong, P. F., & van der Leij, A. (2003). Developmental changes in the manifestation of a phonological deficit in dyslexic children learning to read a regular orthography. *Journal of Educational Psychology*, 95(1), 22.
- Dehaene, S. (2009). *Reading in the brain: The new science of how we read*. Penguin.

- Devonshire, V., Morris, P., & Fluck, M. (2013). Spelling and reading development: The effect of teaching children multiple levels of representation in their orthography. *Learning and Instruction, 25*, 85-94.
- Diehl, S. F. (1999). Listen and learn? A software review of Earobics®. *Language, Speech, and Hearing Services in Schools, 30*(1), 108-116.
- Ecalte, J., Magnan, A., & Calmus, C. (2009). Lasting effects on literacy skills with a computer-assisted learning using syllabic units in low-progress readers. *Computers & Education, 52*(3), 554-561.
- Ehri, L. C., Deffner, N. D., & Wilce, L. S. (1984). Pictorial mnemonics for phonics. *Journal of educational psychology, 76*(5), 880.
- Ehri, L. C., & McCormick, S. (1998). Phases of word learning: Implications for instruction with delayed and disabled readers. *Reading & Writing Quarterly: Overcoming Learning Difficulties, 14*(2), 135-163.
- Ehri, L. C. (2014). Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. *Scientific Studies of Reading, 18*(1), 5-21.
- Ehri, L. C., & Robbins, C. (1992). Beginners need some decoding skill to read words by analogy. *Reading Research Quarterly, 13*-26.
- Ehri, L. C., & Wilce, L. S. (1979). The mnemonic value of orthography among beginning readers. *Journal of Educational Psychology, 71*(1), 26.
- Ellis, N. C., & Hooper, A. (2001). Why learning to read is easier in Welsh than in English: Orthographic transparency effects evinced with frequency-matched tests. *Applied Psycholinguistics, 22*(04), 571-599.
- Facoetti, A., Lorusso, M. L., Paganoni, P., Cattaneo, C., Galli, R., Umiltà, C., & Mascetti, G. G. (2003). Auditory and visual automatic attention deficits in developmental dyslexia. *Cognitive brain research, 16*(2), 185-191.
- Facoetti, A., Paganoni, P., Turatto, M., Marzola, V., & Mascetti, G. G. (2000). Visual-spatial attention in developmental dyslexia. *Cortex, 36*(1), 109-123.
- Facoetti, A., Turatto, M., Lorusso, M. L., & Mascetti, G. G. (2001). Orienting of visual attention in dyslexia: evidence for asymmetric hemispheric control of attention. *Experimental Brain Research, 138*(1), 46-53.
- Facoetti, A., Trussardi, A. N., Ruffino, M., Lorusso, M. L., Cattaneo, C., Galli, R., ... & Zorzi, M. (2010). Multisensory spatial attention deficits are predictive of phonological decoding skills in developmental dyslexia. *Journal of cognitive neuroscience, 22*(5), 1011-1025.
- Foa, E. B., & Hearst-Ikeda, D. (1996). Emotional dissociation in response to trauma. In *Handbook of dissociation* (pp. 207-224). Springer US.

- Fox, B., & Routh, D. K. (1975). Analyzing spoken language into words, syllables, and phonemes: A developmental study. *Journal of Psycholinguistic Research*, 4(4), 331-342.
- Fredembach, B., de Boisferon, A. H., & Gentaz, E. (2009). Learning of arbitrary association between visual and auditory novel stimuli in adults: the “bond effect” of haptic exploration. *PloS one*, 4(3), e4844.
- Frederickson, N., & Jacobs, S. (2001). Controllability Attributions for Academic Performance and the Perceived Scholastic Competence, Global Self-Worth and Achievement of Children with Dyslexia. *School Psychology International*, 22(4), 401-416.
- Frith, U., Wimmer, H., & Landerl, K. (1998). Differences in phonological recoding in German- and English-speaking children. *Scientific Studies of Reading*, 2(1), 31-54.
- Galaburda, A. M., Menard, M. T., & Rosen, G. D. (1994). Evidence for aberrant auditory anatomy in developmental dyslexia. *Proceedings of the National Academy of Sciences*, 91(17), 8010-8013.
- Gattegno, C. (2010). *Teaching reading with words in color*. Educational Solutions World.
- Gillingham, A., & Stillman, B. W. (1946). *Remedial Training for Children with Specific Disability in Reading, Spelling and Penmanship...* The authors.
- Goodman, M. D., & Cundick, B. P. (1976). Learning rates with black and colored letters. *Journal of Learning disabilities*, 9(9), 600-602.
- Gellatly, A., Pilling, M., Cole, G., & Skarratt, P. (2006). What is being masked in object substitution masking?. *Journal of experimental psychology: human perception and performance*, 32(6), 1422
- Goldin-Meadow, S., Alibali, M. W., & Church, R. B. (1993). Transitions in concept acquisition: using the hand to read the mind. *Psychological review*, 100(2), 279.
- Gori, S., & Facoetti, A. (2015). How the visual aspects can be crucial in reading acquisition? The intriguing case of crowding and developmental dyslexia. *Journal of vision*, 15(1), 8.
- Goswami, U., Ziegler, J. C., Dalton, L., & Schneider, W. (2003). Nonword reading across orthographies: How flexible is the choice of reading units?. *Applied Psycholinguistics*, 24(02), 235-247.
- Green, B. F., & Anderson, L. K. (1956). Color coding in a visual search task. *Journal of experimental psychology*, 51(1), 19.
- Haggard, M. P., Summerfield, Q., & Roberts, M. (1981). Psychoacoustical and cultural determinants of phoneme boundaries: Evidence from trading F<sub>0</sub> cues in the voiced–voiceless distinction. *Journal of phonetics*.
- Hanna, A., & Remington, R. (1996). The representation of color and form in long-term memory. *Memory & Cognition*, 24(3), 322-330.

Hansen, P. C., Stein, J. F., Orde, S. R., Winter, J. L., & Talcott, J. B. (2001). Are dyslexics' visual deficits limited to measures of dorsal stream function?. *Neuroreport*, 12(7), 1527-1530.

Hari, R., Valta, M., & Uutela, K. (1999). Prolonged attentional dwell time in dyslexic adults. *Neuroscience letters*, 271(3), 202-204.

Hari, R., & Renvall, H. (2001). Impaired processing of rapid stimulus sequences in dyslexia. *Trends in cognitive sciences*, 5(12), 525-532.

Henry, M. K. (1998). Structured, sequential, multisensory teaching: The Orton legacy. *Annals of Dyslexia*, 48(1), 1-26.

Hines, S. J. (2009). The Effectiveness of a Color-Coded, Onset-Rime Decoding Intervention with First-Grade Students at Serious Risk for Reading Disabilities. *Learning Disabilities Research & Practice*, 24(1), 21-32.

Hook, P. E., Macaruso, P., & Jones, S. (2001). Efficacy of Fast ForWord training on facilitating acquisition of reading skills by children with reading difficulties—A longitudinal study. *Annals of Dyslexia*, 51(1), 73-96.

Hulme, C., Monk, A., & Ives, S. (1987). Some experimental studies of multi-sensory teaching: the effects of manual tracing on children's paired-associate learning. *British Journal of Developmental Psychology*, 5(4), 299-307.

Joffe, L. S. (1981). *School mathematics and dyslexia: Aspects of the interrelationship* (Doctoral dissertation, Aston University).

Jones, M. W., Branigan, H. P., & Kelly, M. L. (2008). Visual deficits in developmental dyslexia: relationships between non-linguistic visual tasks and their components of reading. *Dyslexia*, 14(2), 95-115.

Kast, Monika, et al. "Computer-based learning of spelling skills in children with and without dyslexia." *Annals of dyslexia* 61.2 (2011): 177-200.

Kast, Monika, et al. "Computer-based multisensory learning in children with developmental dyslexia." *Restorative Neurology and Neuroscience* 25.3 (2007): 355-370.

Keller, T., Gerjets, P., Scheiter, K., & Garsoffky, B. (2006). Information visualizations for knowledge acquisition: The impact of dimensionality and color coding. *Computers in Human Behavior*, 22(1), 43-65.

Kerby, Dave S. "The simple difference formula: an approach to teaching nonparametric correlation 1." *Innovative Teaching* 3.1 (2014): Article-1.

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

Kujala, T., Karma, K., Ceponiene, R., Belitz, S., Turkkila, P., Tervaniemi, M., & Näätänen, R. (2001). Plastic neural changes and reading improvement caused by audiovisual training in reading-impaired children. *Proceedings of the National Academy of Sciences*, 98(18), 10509-10514.

Lakoff, G., & Núñez, R. E. (2000). Where mathematics comes from: How the embodied mind brings mathematics into being. Basic books.

Lallier, M., Tainturier, M. J., Dering, B., Donnadieu, S., Valdois, S., & Thierry, G. (2010). Behavioral and ERP evidence for amodal sluggish attentional shifting in developmental dyslexia. *Neuropsychologia*, 48(14), 4125-4135.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of experimental child psychology*, 18(2), 201-212.

Lovett, M. W. (1987). A developmental approach to reading disability: Accuracy and speed criteria of normal and deficient reading skill. *Child Development*, 234-260.

Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A definition of dyslexia. *Annals of dyslexia*, 53(1), 1-14.

MacKay, D. G., & Ahmetzanov, M. V. (2005). Emotion, memory, and attention in the taboo stroop paradigm an experimental analogue of flashbulb memories. *Psychological Science*, 16(1), 25-32. MacKay, & Ahmetzanov, 2005).

Magnan, A., Ecalle, J., Veuillet, E., & Collet, L. (2004). The effects of an audio-visual training program in dyslexic children. *Dyslexia*, 10(2), 131-140.

Magnan, A., & Ecalle, J. (2006). Audio-visual training in children with reading disabilities. *Computers & Education*, 46(4), 407-425.

Mann, V. A. (1986). Phonological awareness: The role of reading experience. *Cognition*, 24(1), 65-92.

Manis, F. R., & Keating, P. (2005). Speech perception in dyslexic children with and without language impairments. *The connections between language and reading disabilities*, 77-99.

Manis, F. R., Seidenberg, M. S., Doi, L. M., McBride-Chang, C., & Petersen, A. (1996). On the bases of two subtypes of development dyslexia. *Cognition*, 58(2), 157-195.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning?. *American Psychologist*, 59(1), 14.

Miller, Paul, and Amirit Kupfermann. "The role of visual and phonological representations in the processing of written words by readers with diagnosed dyslexia: evidence from a working memory task." *Annals of dyslexia* 59.1 (2009): 12-33.

Mioduser, D., Tur-Kaspa, H., & Leitner, I. (2000). The learning value of computer-based instruction of early reading skills. *Journal of Computer Assisted Learning*, 16(1), 54-63.

Nunes, T., Bryant, P., & Bindman, M. (1997). Morphological spelling strategies: developmental stages and processes. *Developmental psychology*, 33(4), 637.

- Oakland, T., Black, J. L., Stanford, G., Nussbaum, N. L., & Balise, R. R. (1998). An Evaluation of the Dyslexia Training Program A Multisensory Method for Promoting Reading in Students with Reading Disabilities. *Journal of Learning Disabilities*, 31(2), 140-147.
- Ozcelik, E., Karakus, T., Kursun, E., & Cagiltay, K. (2009). An eye-tracking study of how color coding affects multimedia learning. *Computers & Education*, 53(2), 445-453.
- Pallant, Julie. *SPSS survival manual*. McGraw-Hill Education (UK), 2013.
- Pammer, K., & Vidyasagar, T. R. (2005). Integration of the visual and auditory networks in dyslexia: a theoretical perspective. *Journal of Research in Reading*, 28(3), 320-331.
- Pöder, Endel. "Effect of colour pop-out on the recognition of letters in crowding conditions." *Psychological Research* 71.6 (2007): 641-645.
- Pokorni, J. L., Worthington, C. K., & Jamison, P. J. (2004). Phonological awareness intervention: comparison of Fast ForWord, Earobics, and LiPS. *The Journal of Educational Research*, 97(3), 147-158.
- Poljac, E., Simon, S., Ringlever, L., Kalcik, D., Groen, W. B., Buitelaar, J. K., & Bekkering, H. (2010). Impaired task switching performance in children with dyslexia but not in children with autism. *The Quarterly Journal of Experimental Psychology*, 63(2), 401-416.
- Powers, A. R., Hillock, A. R., & Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *The Journal of Neuroscience*, 29(39), 12265-12274.
- Read, C., Yun-Fei, Z., Hong-Yin, N., & Bao-Qing, D. (1986). The ability to manipulate speech sounds depends on knowing alphabetic writing. *Cognition*, 24(1), 31-44.
- Reid, G. (2013). *Dyslexia: A practitioner's handbook*. John Wiley & Sons.
- Regtvoort, A. G., & van der Leij, A. (2007). Early intervention with children of dyslexic parents: Effects of computer-based reading instruction at home on literacy acquisition. *Learning and individual differences*, 17(1), 35-53.
- Richardson, U., Thomson, J. M., Scott, S. K., & Goswami, U. (2004). Auditory processing skills and phonological representation in dyslexic children. *Dyslexia*, 10(3), 215-233.
- Ritchey, K. D., & Speece, D. L. (2006). From letter names to word reading: The nascent role of sublexical fluency. *Contemporary Educational Psychology*, 31(3), 301-327.
- Ritchey, K. D., & Goeke, J. L. (2006). Orton-Gillingham and Orton-Gillingham—based reading instruction a review of the literature. *The Journal of Special Education*, 40(3), 171-183.
- Roach, N. W., & Hogben, J. H. (2007). Impaired filtering of behaviourally irrelevant visual information in dyslexia. *Brain*, 130(3), 771-785.

- Roberson, D., Davies, I., & Davidoff, J. (2000). Color categories are not universal: replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General*, 129(3), 369.
- Roschelle, J. M., Pea, R. D., Hoadley, C. M., Gordin, D. N., & Means, B. M. (2000). Changing how and what children learn in school with computer-based technologies. *The future of children*, 76-101.
- Rumseyt, J. M., Maisog, J. M., & Woods, R. P. (1996). Abnormal processing of visual motion in dyslexia revealed by functional brain imaging. *Nature*, 382, 4.
- Sadoski, M., & Paivio, A. (2004). A dual coding theoretical model of reading. *Theoretical models and processes of reading*, 5, 1329-1362.
- Schatschneider, C., & Torgesen, J. K. (2004). Using our current understanding of dyslexia to support early identification and intervention. *Journal of Child Neurology*, 19(10), 759-765.
- Serniclaes, W., Ventura, P., Morais, J., & Kolinsky, R. (2005). Categorical perception of speech sounds in illiterate adults. *Cognition*, 98(2), B35-B44.
- Seyama, J. I., & Nagayama, R. S. (2007). The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence: Teleoperators and Virtual Environments*, 16(4), 337-351.
- Seymour, P. H., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of psychology*, 94(2), 143-174.
- Share, D. L. (2004). Orthographic learning at a glance: On the time course and developmental onset of self-teaching. *Journal of experimental child psychology*, 87(4), 267-298.
- Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of experimental child psychology*, 72(2), 95-129.
- Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, 55(2), 151-218.
- Shmuelof, L., & Zohary, E. (2005). Dissociation between ventral and dorsal fMRI activation during object and action recognition. *Neuron*, 47(3), 457-470.
- Smith, V. C., & Pokorny, J. (1975). Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm. *Vision research*, 15(2), 161-171. Smith & Pokorny, 1975
- Snowling, M. J. (1981). Phonemic deficits in developmental dyslexia. *Psychological research*, 43(2), 219-234.
- Stein, J., & Walsh, V. (1997). To see but not to read; the magnocellular theory of dyslexia. *Trends in neurosciences*, 20(4), 147-152.

- Stoet, G., Markey, H., & López, B. (2007). Dyslexia and attentional shifting. *Neuroscience Letters*, 427(1), 61-65.
- Treiman, R., & Kessler, B. (2006). Spelling as statistical learning: Using consonantal context to spell vowels. *Journal of Educational Psychology*, 98(3), 642.
- Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance*, 8(2), 194.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, 12(1), 97-136.
- Uttal, D. H., O'Doherty, K., Newland, R., Hand, L. L., & DeLoache, J. (2009). Dual representation and the linking of concrete and symbolic representations. *Child Development Perspectives*, 3(3), 156—159.
- Valdois, S., Bosse, M. L., & Tainturier, M. J. (2004). The cognitive deficits responsible for developmental dyslexia: Review of evidence for a selective visual attentional disorder. *Dyslexia*, 10(4), 339-363.
- Van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron*, 43(2), 271-282.
- Vidyasagar, T. R., & Pammer, K. (2010). Dyslexia: a deficit in visuo-spatial attention, not in phonological processing. *Trends in cognitive sciences*, 14(2), 57-63.
- Vidyasagar, T. R., & Pammer, K. (1999). Impaired visual search in dyslexia relates to the role of the magnocellular pathway in attention. *Neuroreport*, 10(6), 1283-1287.
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 27(1), 92.
- Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological bulletin*, 101(2), 192.
- Ware, C. (2012). *Information visualization: perception for design*. Elsevier.
- Watson, M. R., Blair, M. R., Kozik, P., Akins, K. A., & Enns, J. T. (2012). Grapheme-color synaesthesia benefits rule-based Category learning. *Consciousness and cognition*, 21(3), 1533-1540.
- Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, 131(1), 48.
- Wrembel, M. (2009). On hearing colours—Cross-modal associations in vowel perception in a non-synaesthetic population. *Poznań Studies in Contemporary Linguistics*, 45(4), 595-612.

Wrembel, M. (2007, August). Still sounds like a rainbow-A proposal for a coloured vowel chart. In *Proceedings of the Phonetics Teaching and Learning Conference PTLC2007* (pp. 1-4).

Wright, C. M., Conlon, E. G., & Dyck, M. (2012). Visual search deficits are independent of magnocellular deficits in dyslexia. *Annals of dyslexia*, 62(1), 53-69.

Wolf, M., Miller, L., & Donnelly, K. (2000). Retrieval, Automaticity, Vocabulary Elaboration, Orthography (RAVE-O) A Comprehensive, Fluency-Based Reading Intervention Program. *Journal of learning disabilities*, 33(4), 375-386.

Wolf, M., Bally, H., & Morris, R. (1986). Automaticity, retrieval processes, and reading: A longitudinal study in average and impaired readers. *Child Development*, 988-1000.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it?. *Nature Reviews Neuroscience*, 5(6), 495-501.

Zelazo, P. D., Carlson, S. M., & Kesek, A. (2008). The development of executive function in childhood.

Zentall, S. S., Grskovic, J. A., Javorsky, J., & Hall, A. M. (2000). Effects of noninformational color on the reading test performance of students with and without attentional deficits. *Assessment for Effective Intervention*, 25(2), 129-146.

Ziegler, J. C., & Goswami, U. (2006). Becoming literate in different languages: similar problems, different solutions. *Developmental science*, 9(5), 429-436.

Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., ... & Blomert, L. (2010). Orthographic depth and its impact on universal predictors of reading a cross-language investigation. *Psychological Science*.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological bulletin*, 131(1), 3.

## **Appendix A.**

### **Initial Session Script**

The words for the first consonant-le session were “table” and “rubble”. The words for the first vowel discrimination activity were “chap” and “debt”. Tutors of children in each condition followed the same general script for reminding children of the corresponding words; the script differed between the conditions with respect to the kinds of visual changes that the tutors could draw children’s attention to. The scripts for the 2 activities follow:

#### **Consonant-Le:**

Placing the “b” after “ta” caused the “a” sound to change from long (in an open syllable) to short (in a closed syllable). In each condition, a different visual change correlated the change of sound. In the particular scheme, the letter flashed (but did not change colour). The tutor said:

“Did you notice that the letter flashed? A letter flashes when the sound changes. The “a” in “ta” was long, but not the “a” in “tab” is short.”

In the categorical scheme, the “a” changed colour from red to yellow. The tutor said:

“Did you notice that the letter changed colour? A letter changes colour when the sound changes. The “a” in “ta” was long. “Long” vowels are red. Now the “a” in “tab” is short. “Short” vowels are yellow.”

The tutor then created the stable unit, “ble”, by adding the letters “l” and “e”. In both schemes, the letters “b,l,e” all became magenta, reflecting their membership in the stable unit. The tutor said:

“Did you notice these letters all became the same colour? These letters form the stable consonant-le syllable, *b/le*. When a consonant and l and e appear at the end of the word, they make a stable syllable, and they will all become magenta.”

Simultaneously, because “ble” is a stable unit, the first 2 letters (“ta”) were re-interpreted as an open syllable in which the “a”, again, sounded long. The change in colour or the flash that corresponded to the “ble” unit therefore correlated a second visual change, in that the “a” also flashed (in the particulars scheme), or changed colour from yellow back to red. The tutors drew children’s attention to this change as well. In the particulars scheme, the tutor said:

“Did you notice that when I created the stable syllable, the “a” also flashed?”

In the categorical scheme, the tutor said:

“Did you notice that when I created the stable syllable, the “a” changed back to red?”

The tutor then prompted the child to conjecture why the “a” might have flashed or changed colour. My intention here was encouraging the children to directly correspond changes in visual to changes in phonological properties. The tutors asked:

“Why do you think the letter changed? How does the “a” sound now?”

The expected answer was something to the effect of, “the sound of the “a’ changed. It used to sound short, but now it sounds long.” I did not expect that children would be able to articulate an explanation in terms of the new way of dividing the word following the creation of the consonant-le unit. If the child responded incorrectly, the tutor would say:

“When a letter [flashes or changes colour] it means that the sound has changed. The “a” [flashed or changed from yellow to red] because its sound changed from short to long.”

In either case (whether the child did or did not identify the change in sound as the visual change) the tutor said:

“Now why do you think the sound changed? Why would it change after I created the consonant-le syllable?”

The expected response was something to the effect of: “we divide the word into an open syllable, “ta”, and the consonant-le syllable, “ble”. Before there was a consonant-le syllable the vowel appeared in a closed syllable. Vowels in closed syllables sound short, vowels in open syllables sound long.” This is what the tutors said if the child did not answer correctly. Following this, the tutor reinforced the implication of the sound change upon how the word was syllabified, using the system’s multi-touch functionality. Left-swiping a screen letter “de-activates” it. Right swiping a screen letter “re-activates” it. My algorithm discounts de-activated letters from its calculation of the sounds (and colours) that the letters have. The system updates the colours after each de/re-activation. I instructed the tutors to use this functionality to show children the effect of removing the stable syllable. Tutors left-swipe across the entire stable syllable, starting from the e. De-activating the “e” and the “l” causes the system to re-interpret the word as the simple closed syllable, “tab”. The colours of the screen and tangible letters instantly update to reflect this: the “a” again flashes in the particulars scheme; the “a” becomes yellow in the categorical scheme. (De-activated letters are black/turned off). The tutor does not rest until they have also de-activated the “b”. Their finger comes to rest to the right of the “a”. I designed this action- de-activating the grouped letters of the consonant-le syllable- to mimic the act of “drawing a box” around the stable syllable. Both activities are supposed to establish a visual boundary between the stable syllable and what remains such as to explicate the type of the syllable preceding the stable one. In this case, the type was open; as the tutor de-activates “b”, the “a” would flash a final time, or change from yellow back to red. My script instructed tutors to draw children’s attention to the change:

“We can figure out the kind of syllable that comes before the consonant-le syllable by taking the consonant-le syllable away. I have taken the consonant-le syllable

away. We can see that this syllable, which comes first, is open. The vowels in open syllables sound long.”

The next word (rubble) involved the opposite sound, the short vowel, and therefore required consonant gemination. The steps of this activity were identical to the first (except that the vowel changed from long to short and did not become long again); the main difference occurred when the tutor enacted syllable division. In this case, deactivating the stable syllable (“ble”), which was, in the particular and categorical scheme, visually distinguished by its magenta colour, left the closed syllable, “rub”. The tutor said:

“Here, the syllable that comes before is closed. Rub. The vowels in closed syllables sound short. If we hear a word that has a consonant-le syllable and that has a short vowel, we need to add an extra consonant, here. That way, what’s left after we take away the consonant-le syllable is a closed syllable, not an open one.”

After the tutor submitted the second consonant-le word to the history, the system imposed a short break. During this break, I instructed the tutors to take advantage of the words in the history to leverage the colour codes for Guided Discovery. After the 2 activities for a given type were completed, both words of the activity appeared in the history. The words always exemplified a key contrast; depending on the child’s condition, a colour contrast corresponded to it. The tutor stepped through the same “script” that is used to facilitate Guided Discovery more generally. The tutor prompted the child to notice the correlated differences in orthography and sound between the 2 words. For consonant-le words, the tutor said:

“What is different about the sounds of the words? What is different about how the vowel sounds?”

If the child responded correctly (“the vowel in the first word sounds short, the other sounds long”), the tutor progressed to the next question. Otherwise, the tutor would guide the child to attend to the vowel sounds (“Listen to the vowel sounds. What is different about them?”). If the child failed to answer correctly after the second prompt, the tutor provided the answer:

“The vowel in *table* sounds long; the vowel in *rubble* sound short.”

The second question probed the child’s knowledge of the orthographic difference. The tutor said:

“What’s different about the way the words look?”

In both the particular and the categorical scheme, the doubled consonant was made more salient because it corresponded to a local contrast of white and magenta that was absent in the word without the doubled consonant, wherein the magenta consonant-le syllable flanked the coloured vowel. In the categorical scheme, the vowel for the geminated word was yellow; the vowel for the non-geminated word was red. In the particulars scheme, if the vowels differed they had different colours, but children’s attention had not been drawn to the colours of the vowels in the particulars scheme because they never changed *within* an activity as they did in the categorical scheme. If the child failed to answer correctly, the tutor said:

“The consonant is doubled in this word, rubble, where the vowel sounds short. The consonant is not doubled in *table*, where the vowel sounds long.”

In the categorical scheme, the tutors reinforced the correspondence to the vowels colours:

“Short vowels are always yellow. Long vowels are always red. You can see that the vowel with the doubled consonant is yellow, and the vowel without the doubled consonant is red.”

### **Vowel Discrimination:**

Tutors stepped through the vowel discrimination activities in a similar way. In this case, the only spelling decision was which vowel to place. On the basis of my interviews with the tutors, I expected that children would likely experiment with several different “alternate spellings” of the vowel sounds, and likely submit a few incorrect answers,

which would also showcase the alternate vowel colours in the particulars scheme. I designed the tutor script to emulate this process.

The initial letters of the vowel discrimination words are all of the letters minus the vowel, leaving an empty space where children are supposed to add the vowel. The first word was “chap”. After tutors placed the initial letters of “chap” (“c”, “h”, and “p”), the first vowel they placed was an “e”. In each scheme, the “e” flashed once when it was placed. In the categorical scheme, the “e” was yellow; in the particulars scheme, the “e” was green. The tutors said:

“Let’s try submitting this word and see what happens.”

The word was incorrect, so the system said:

“That’s not quite it,” and offered the hint. This yielded an opportunity for the tutors to show children the hint system. The tutors said:

“PhonoBlocks gives us a hint if we answer incorrectly. See this button? We can press it to hear a hint.”

The tutors then pressed the button. The hint was as I described; the system repeated the instructions, including the word to create, and emphasized each sound. The tutors extended the auditory supporting by saying:

“Pay attention to the vowel. What vowel sound did you hear?”

The tutor then repeated the word and the enunciation of each sound. The child was supposed to identify the sound as “*ah*” (or short a). If the child failed to identify the sound, the tutors supplied it for them:

“The missing sound is *ah*. *Chap*. We have *eh*. *Chep*. Do you hear the difference? *ah*, *eh*. *Chap*, *chep*.”

The tutor then removed the “e” and replaced it with the “a”. In all schemes, as before, the new letter flashed to reflect the change in sound; however, only in the

particulars scheme did the “a” and “e” have different colours. In the particulars scheme, the tutors drew children’s attention to this fact:

“Did you notice that the *a* and the *e* have different colours? The *e* was green, but the *a* is red. Each vowel has a unique colour, just as each vowel has a unique sound.”

The tutors submitted the new word (*chap*), which was correct.

One objective of the vowel discrimination practice session was introducing children in the particulars scheme to the colours for each vowel. This required the tutors to substitute the remaining vowels (o, i, u) into the place in the second word, *debt*, that was reserved for the “e”. First, the tutor placed the initial letters (“d”, “b”, “t”) of “*debt*”. Then, in all schemes, tutors said:

“When we’re completing these words, we know that the missing letter is always a vowel. Every word needs a vowel, and these words only have consonants. We could place any vowel here, and it would make a pronounceable word.”

The tutor then successively placed each vowel into the available slot, finishing with the correct vowel, “e”. In the particulars scheme, the tutor additionally mentioned the unique colours of each vowel as they placed them:

“The “i” is yellow, the “o” is blue, the “u” is blue-green [cyan].”

Upon finishing on the “e”, the tutor said:

“But only one vowel sound is the right one. Each vowel sounds a little bit different, so we need to pay close attention to the vowel in the word. Right now, we need to make the word *debt*. *Debt*. The vowel sound is *eh*. So we need this e.”

As with the consonant-le activities, following the second activity, there were 2 words in the history that highlighted a challenging contrast (in this case, short e and short i). The tutors used the history to reinforce the orthographic and phonological differences:

“Here are the 2 words, shed and ship. Chap, debt. Chap, debt. Can you identify the vowel sounds in these 2 words?”

The expected response was “ah” and “eh”. If the child did not supply this response, the tutor provided it for them. Thereafter, the tutor drew children’s attention to the orthographic differences:

“The letter *a* represents *ah*, the letter *e* represents *eh*.”

In the particulars scheme, the tutors reinforced the differences’ correspondence with the colours:

“Each vowel letter and each short vowel sound corresponds to a different colour. *Eh* corresponds to green. *ah* corresponds to red.”

## Appendix B. Assessment and PhonoBlocks Session Words

Pre-Test:

Consonant-Le

Coloured:

"bugle"/"settle"

"maple"/"gobble"

"rifle"/"ruffle"

"table"/"topple"

Uncoloured:

"bible"/"cattle",

"ogle"/"cuddle",

"ladle"/"muffle",

"noble"/"rubble"

Vowel-Discrimination

Coloured

"that"/"left"

"twin"/"rent"

"trap"/"sled"

"knot"/"drug"

Uncoloured

"pass"/"bent"

"list"/"debt"

"pack"/"ness"

"posh"/"push"

Post-Test:

Consonant-Le

Coloured:

*Familiar*

"settle"/"bugle"

"gobble"/"maple"

*Unfamiliar*

"hubble"/"sidle"

"toddle"/"gable"

Uncoloured:

*Familiar*

"bible"/"cattle"

"cuddle"/"ogle"

*Unfamiliar*

"saddle"/"sable"

"waffle"/"idle"

Vowel-Discrimination

Coloured

*Familiar*

"posh"/"push"

"pack"/"ness"

*Unfamiliar*

"test"/"cash"

"loss"/"club"

Uncoloured

*Familiar*

"knot"/"drug"

"twin"/"rent"

*Unfamiliar*

"plan"/"rest"

"lost"/"plug"

Session Words:

Consonant-Le:

"rubble"/"table"

"ruffle"/"rifle"

"cuddle"/"cradle"

"wiggle"/"bugle"

"ripple"/"maple"

"cattle"/"title"

"settle"/"noble"

"supple"/"trifle"

"boggle"/"ladle"

"pebble"/"ogle"

"fiddle"/"staple"

"muffle"/"bible"

Vowel Discrimination

"chap"/"bent"

"list"/"sled"

"trap"/"push"

"posh"/"ness"

"pack"/"left"

"twin"/"drug"

"knot"/"rent"

"that"/"bred"

"lass"/"fled"

"grip"/"dreg"

"pass"/"bust"

## Appendix C: Tutor Interview Questions

1. Did you notice any positive or negative behavioural differences in children practicing their spelling with PhonoBlocks, relative to when they don't use PhonoBlocks?
  - Can you describe the changes for me?
  - Were any of these differences unique to cases where the child's colour coding scheme matched the activity? [if yes]:
  - Did children appear to use the colours as feedback in helping them decode or spell?
2. Do you think PhonoBlocks would be as effective without the colour-codes? Why or why not?
3. Do you think PhonoBlocks would be as effective without the plastic letters? Why or why not?
4. Do you have any other comments or thoughts?